

A cluster randomised controlled trial and evaluation and cost-effectiveness analysis of the Roots of Empathy schools-based programme for improving social and emotional well-being outcomes among 8- to 9-year-olds in Northern Ireland

Paul Connolly, Sarah Miller, Frank Kee, Seaneen Sloan, Aideen Gildea, Emma McIntosh, Nicole Boyer and Martin Bland



**National Institute for
Health Research**

A cluster randomised controlled trial and evaluation and cost-effectiveness analysis of the Roots of Empathy schools-based programme for improving social and emotional well-being outcomes among 8- to 9-year-olds in Northern Ireland

Paul Connolly,^{1*} Sarah Miller,¹ Frank Kee,²
Seaneen Sloan,¹ Aideen Gildea,¹ Emma McIntosh,³
Nicole Boyer³ and Martin Bland⁴

¹Centre for Evidence and Social Innovation, Queen's University Belfast, Belfast, UK

²Centre of Excellence for Public Health Research (Northern Ireland), Queen's University Belfast, Belfast, UK

³Institute of Health and Wellbeing, Public Health and Health Policy, University of Glasgow, Glasgow, UK

⁴Department of Health Sciences, University of York, York, UK

*Corresponding author

Declared competing interests of authors: Frank Kee reports that he chairs the National Institute for Health Research (NIHR) Public Health Research (PHR) programme Research Funding Board. Emma McIntosh reports that she is a member of the NIHR PHR programme Research Funding Board.

Published March 2018

DOI: 10.3310/phr06040

This report should be referenced as follows:

Connolly P, Miller S, Kee F, Sloan S, Gildea A, McIntosh E, *et al.* A cluster randomised controlled trial and evaluation and cost-effectiveness analysis of the Roots of Empathy schools-based programme for improving social and emotional well-being outcomes among 8- to 9-year-olds in Northern Ireland. *Public Health Res* 2018;**6**(4).

Public Health Research

ISSN 2050-4381 (Print)

ISSN 2050-439X (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full PHR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/phr. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Public Health Research* journal

Reports are published in *Public Health Research* (PHR) if (1) they have resulted from work for the PHR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Public Health Research* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

PHR programme

The Public Health Research (PHR) programme, part of the National Institute for Health Research (NIHR), evaluates public health interventions, providing new knowledge on the benefits, costs, acceptability and wider impacts of non-NHS interventions intended to improve the health of the public and reduce inequalities in health. The scope of the programme is multi-disciplinary and broad, covering a range of interventions that improve public health. The Public Health Research programme also complements the NIHR Health Technology Assessment programme which has a growing portfolio evaluating NHS public health interventions.

For more information about the PHR programme please visit the website: <http://www.nets.nihr.ac.uk/programmes/phr>

This report

The research reported in this issue of the journal was funded by the PHR programme as project number 10/3006/02. The contractual start date was in January 2012. The final report began editorial review in September 2016 and was accepted for publication in June 2017. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The PHR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the PHR programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the PHR programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2018. This work was produced by Connolly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Public Health Research Editor-in-Chief

Professor Martin White Director of Research and Programme Leader, UKCRC Centre for Diet and Activity Research (CEDAR), MRC Epidemiology Unit, Institute of Metabolic Science, School of Clinical Medicine, University of Cambridge; Visiting Professor, Newcastle University; and Director, NIHR Public Health Research Programme

NIHR Journals Library Editor-in-Chief

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the EME Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA and EME Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Director of the NIHR Dissemination Centre, University of Southampton, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Director, Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

A cluster randomised controlled trial and evaluation and cost-effectiveness analysis of the Roots of Empathy schools-based programme for improving social and emotional well-being outcomes among 8- to 9-year-olds in Northern Ireland

Paul Connolly,^{1*} Sarah Miller,¹ Frank Kee,² Seaneen Sloan,¹
Aideen Gildea,¹ Emma McIntosh,³ Nicole Boyer³ and Martin Bland⁴

¹Centre for Evidence and Social Innovation, Queen's University Belfast, Belfast, UK

²Centre of Excellence for Public Health Research (Northern Ireland), Queen's University Belfast, Belfast, UK

³Institute of Health and Wellbeing, Public Health and Health Policy, University of Glasgow, Glasgow, UK

⁴Department of Health Sciences, University of York, York, UK

*Corresponding author paul.connolly@qub.ac.uk

Background: There is growing consensus regarding the importance of attending to children's social and emotional well-being. There is now a substantial evidence base demonstrating the links between a child's early social and emotional development and a range of key longer-term education, social and health outcomes. Universal school-based interventions provide a significant opportunity for early intervention in this area and yet the existing evidence base, particularly in relation to their long-term effects, is limited.

Objectives and main outcomes: To determine the effectiveness and cost-effectiveness of Roots of Empathy (ROE), a universal school-based programme that, through attempting to enhance children's empathy, seeks to achieve the following two main outcomes: improvement in prosocial behaviour and reduction in difficult behaviour.

Design: A cluster randomised controlled trial and an economic evaluation. A total of 74 primary schools were randomly assigned to deliver ROE or to join a waiting list control group. Seven schools withdrew post randomisation and a further two withdrew before the immediate post-test time point. Children ($n = 1278$) were measured pre test and immediately post test, and then for 3 years following the end of the programme. Data were also collected from teachers and parents.

Setting and participants: The intervention schools delivered ROE to their Year 5 children (aged 8–9 years) as a whole class.

Intervention: ROE is delivered on a whole-class basis for one academic year (October–June). It consists of 27 lessons based around the monthly visit from a baby and parent who are usually recruited from the local community. Children learn about the baby's growth and development and are encouraged to generalise from this to develop empathy towards others.

Results: Although it was developed in Canada, the programme was very well received by schools, parents and children, and it was delivered effectively with high fidelity. ROE was also found to be effective in achieving small improvements in children's prosocial behaviour (Hedges' $g = 0.20$; $p = 0.045$) and

reductions in their difficult behaviour (Hedges' $g = -0.16$; $p = 0.060$) immediately post test. Although the gains in prosocial behaviour were not sustained after the immediately post-test time point, there was some tentative evidence that the effects associated with reductions in difficult behaviour may have remained up to 36 months from the end of the programme. These positive effects of ROE on children's behaviour were not found to be associated with improvements in empathy or other social and emotional skills (such as emotional recognition and emotional regulation), on which the trial found no evidence of ROE having an effect. The study also found that ROE was likely to be cost-effective in line with national guidelines.

Conclusions: These findings are consistent with those of other evaluations of ROE and suggest that it is an effective and cost-effective programme that can be delivered appropriately and effectively in regions such as Northern Ireland. A number of issues for further consideration are raised regarding opportunities to enhance the role of parents; how a time-limited programme such as ROE can form part of a wider and progressive curriculum in schools to build on and sustain children's social and emotional development; and the need to develop a better theory of change for how ROE works.

Trial registration: Current Controlled Trials ISRCTN07540423.

Funding: This project was funded by the National Institute for Health Research (NIHR) Public Health Research programme and will be published in full in *Public Health Research*; Vol. 6, No. 4. See the NIHR Journals Library website for further project information.

Contents

List of tables	xi
List of figures	xv
List of supplementary material	xvii
List of abbreviations	xix
Plain English summary	xxi
Scientific summary	xxiii
Chapter 1 Introduction	1
Rationale for current study	1
Scientific background	2
Existing impact evaluations of the Roots of Empathy programme	2
Objectives	4
Chapter 2 Methodology	5
Introduction	5
Trial design	5
<i>Deviations of the evaluation from the original protocol</i>	5
Participants	6
Intervention	7
Outcomes	8
<i>Data collection</i>	8
Sample size	12
Randomisation	12
Statistical methods	14
<i>Sensitivity analyses</i>	15
Qualitative process evaluation	15
<i>Selection of the sample</i>	15
<i>Ethics, consent and data analysis</i>	16
Cost-effectiveness analysis	17
<i>Methods overview</i>	18
<i>Methods</i>	19
<i>Quality-adjusted life-years</i>	20
<i>Missing data</i>	20
<i>Analysis</i>	21
Stakeholder engagement	23
<i>Partnership meetings</i>	23
<i>Stakeholder members of the Trial Steering Committee</i>	23
<i>Process evaluation</i>	23
<i>End-of-project consultation meetings</i>	23
<i>Dissemination events</i>	24

Chapter 3 Results from the trial and cost-effectiveness analysis	25
Introduction	25
Participant flow	25
Recruitment	25
Baseline data	26
Outcomes and estimation	29
Ancillary analyses	31
<i>Secondary outcomes</i>	31
<i>Exploratory subgroup analyses</i>	31
<i>Further exploratory analysis</i>	32
<i>Primary outcomes at 12-, 24- and 36-month follow-up</i>	33
<i>Secondary outcomes and exploratory analyses</i>	34
<i>Sensitivity analysis for missing data</i>	36
Cost-effectiveness analysis	38
<i>Missing data</i>	38
<i>Clustering in economic evaluation</i>	42
<i>Costs and quality-adjusted life-years</i>	42
<i>Cost-effectiveness</i>	43
<i>Discussion</i>	45
<i>Limitations</i>	47
Conclusions	47
Chapter 4 Process evaluation	49
Introduction	49
Implementation and fidelity	49
Programme content	51
Mentoring and support	53
Limitations of the programme	54
<i>Teachers' perspectives of the limitations of Roots of Empathy</i>	54
<i>School principals' perspectives of the limitations of Roots of Empathy</i>	55
<i>Instructors' perspectives of the limitations of Roots of Empathy</i>	56
Main challenges	57
Benefits of the programme	58
<i>Children's perspectives of the benefits of Roots of Empathy</i>	58
<i>Teachers' perspectives of the benefits of Roots of Empathy</i>	59
<i>School principals' perspectives of the benefits of Roots of Empathy</i>	63
<i>Instructors' perspectives of the benefits of Roots of Empathy</i>	64
<i>Parents' perspective of the benefits of Roots of Empathy</i>	65
Parental involvement in Roots of Empathy	68
<i>School personnel's perspectives on parental involvement</i>	68
<i>Parents' perspectives on parental involvement</i>	70
<i>Children's perspectives on parental involvement</i>	70
Conclusion	71
Chapter 5 Discussion and conclusions	73
Introduction	73
Key findings	73
<i>Effectiveness of Roots of Empathy</i>	73
<i>Cost-effectiveness of Roots of Empathy</i>	73
<i>Programme delivery and stakeholder perspectives</i>	74
Limitations	74
Generalisability	75
Interpretation	75

Acknowledgements	77
References	79
Appendix 1 Meta-analysis of existing evaluations of Roots of Empathy	85
Appendix 2 Statistical models for immediately post test (time 1), time 2, time 3 and time 4	95
Appendix 3 Sensitivity analysis to assess impact of not including three child-level covariates in the statistical models	105
Appendix 4 Resource use questionnaire	107

List of tables

TABLE 1 Summary of meta-analyses of previous evaluations ($n = 7$) of the ROE programme	3
TABLE 2 Description of the outcomes, measures and reliability pre and post test	10
TABLE 3 Number of intervention and control schools in each of the participating health and social care trusts at pre test and immediately post test	14
TABLE 4 Data from ROE trial collected for the economic evaluation	18
TABLE 5 List of sensitivity analyses	22
TABLE 6 The ICER for cost-effectiveness analyses on SDQ	23
TABLE 7 Schools in the sample, by type, compared with the Northern Ireland population in 2011/12	25
TABLE 8 Sample characteristics	26
TABLE 9 Comparison between control and intervention groups on demographic characteristics	27
TABLE 10 Comparison between control and intervention groups on primary and secondary outcome measures pre test	27
TABLE 11 Comparison of demographic characteristics between those who completed pre-test questionnaire only and those who completed pre- and post-test questionnaires	28
TABLE 12 Comparison of outcome measures pre test between those who completed pre-test questionnaire only and those who completed pre- and post-test questionnaires	28
TABLE 13 Summary of main effects immediately post test (T1)	29
TABLE 14 Comparison of effect sizes for the primary outcomes used in the main analysis (teacher-rated SDQ) with those measured by parent- and child-rated SDQ and teacher-rated CBS	30
TABLE 15 Statistical significance of coefficients for interaction effects added to statistical models for child gender, family socioeconomic status and number of siblings, respectively, immediately post test (T1)	31
TABLE 16 Comparison of ROE sample at baseline and post-test SDQ mean scores with UK, Ireland and US national averages	33
TABLE 17 Duration (minutes) of PDMU delivered in class per week by intervention and control teachers	33

TABLE 18 Summary of main effects at 12-month follow up (T2)	34
TABLE 19 Summary of main effects at 24-month follow-up (T3)	35
TABLE 20 Summary of main effects at 36-month follow-up (T4)	36
TABLE 21 Statistical significance of coefficients for interaction effects added to statistical models for child gender, family socioeconomic status and number of siblings, respectively, at 12-month follow-up (T2)	37
TABLE 22 Statistical significance of coefficients for interaction effects added to statistical models for child gender, family socioeconomic status and number of siblings, respectively, at 24-month follow-up (T3)	37
TABLE 23 Statistical significance of coefficients for interaction effects added to statistical models for gender, socioeconomic status and number of siblings, respectively, at 36-month follow-up (T4)	38
TABLE 24 Comparison of main effects estimated with the observed data only with the effects estimated using data sets with multiple imputation	39
TABLE 25 Variable descriptions and missing data percentages	40
TABLE 26 Outcome variables for cost	40
TABLE 27 Cost of the ROE intervention	42
TABLE 28 Unit costs of public sector service use	43
TABLE 29 Cost-effectiveness results	44
TABLE 30 Characteristics of previous quantitative evaluations included in the meta-analysis	86
TABLE 31 Risk-of-bias summary: authors' judgements about each risk-of-bias item for each included study	89
TABLE 32 Multilevel models fitted for the teacher-rated prosocial behaviour (SDQ) variable at each time point	95
TABLE 33 Multilevel models fitted for the teacher-rated difficult behaviour (SDQ) variable at each time point	96
TABLE 34 Multilevel models fitted for the child-rated 'reasons why a baby cries' variable at each time point	97
TABLE 35 Multilevel models fitted for the child-rated 'ways to help a crying baby' variable at each time point	98
TABLE 36 Multilevel models fitted for the child-rated emotional recognition variable at each time point	99

TABLE 37 Multilevel models fitted for the child-rated empathy variable at each time point	100
TABLE 38 Multilevel models fitted for the child-rated emotional regulation variable at each time point	101
TABLE 39 Multilevel models fitted for the child-rated bullying (victim) variable at each time point	102
TABLE 40 Multilevel models fitted for the child-rated bullying (bully) variable at each time point	103
TABLE 41 Multilevel models fitted for the child-rated quality-of-life variable at each time point	104
TABLE 42 Comparison of effect sizes immediately post test between the models used in the main analysis, these models extended with additional covariates and then analysed with multiple imputation of missing data	105

List of figures

FIGURE 1 Logic model for the ROE programme	9
FIGURE 2 Flow diagram of recruitment and testing of children	13
FIGURE 3 Pattern of missing data in (a) costs and (b) QALYs	41
FIGURE 4 Cost-effectiveness plane representing 1000 bootstrapped cost and QALY pairs	45
FIGURE 5 Cost-effectiveness acceptability curve showing probability of ROE being cost-effectiveness compared with the control	45
FIGURE 6 Forest plot of SMDs between intervention and control for prosocial behaviour	92
FIGURE 7 Forest plot of SMDs between intervention and control for aggressive behaviour	92
FIGURE 8 Forest plot of SMDs between intervention and control for child-reported empathy	94
FIGURE 9 Forest plot of SMDs between intervention and control for child-reported emotional regulation	94

List of supplementary material

Report Supplementary Material 1 Research instruments used in study

Supplementary material can be found on the NIHR Journals Library report project page (www.journalslibrary.nihr.ac.uk/programmes/phr/10300602/#/documentation).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

List of abbreviations

CAMS	Child Anger Management Scale	NIMDM	Northern Ireland Multiple Deprivation Measure
CBS	Child Behaviour Scale		
CHU9D	Child Health Utility – 9D	PDMU	Personal Development and Mutual Understanding
CI	confidence interval	QALY	quality-adjusted life-year
EQ-5D-Y	EuroQol-5 Dimensions youth version	ROE	Roots of Empathy
ES	effect size	SD	standard deviation
HSCT	Health and Social Care Trust	SDQ	Strengths and Difficulties Questionnaire
ICC	intracluster correlation coefficient	SEL	social and emotional learning
ICER	incremental cost-effectiveness ratio	SMD	standardised mean difference
ISRCTN	International Standard Randomised Controlled Trial Number	T	time
NICE	National Institute for Health and Care Excellence		

Plain English summary

Children's early social and emotional development remains a significant predictor of their future social, education and health outcomes. Roots of Empathy (ROE) is a school-based programme that is delivered on a whole-class basis for one academic year. It seeks to increase children's empathy, leading to positive behaviour change and preparing them better for later life. It consists of 27 lessons based around the monthly visit from an infant and parent, who are usually recruited from the local community. This study provides a robust evaluation of the effectiveness and cost-effectiveness (value for money) of ROE. ROE involved 74 primary schools in Northern Ireland randomly divided into two equal groups that either delivered ROE during 2011–12 or acted as a control group. The effectiveness of ROE was measured immediately at the end of the year and for the following 3 years (up to 2015). The study found that ROE was effective in increasing the children's prosocial behaviour and reducing their difficult behaviour immediately at the end of the programme. Although the effects on prosocial behaviour fell away after the first year, there was some possible evidence that the effects on reducing difficult behaviour may have been sustained for the following 3 years. Although originally developed in Canada, ROE was very well received by teachers, parents and children, and it was effectively delivered in schools in the context of Northern Ireland. The study also found that ROE was likely to be a cost-effective use of society's resources as a means of improving children's quality of life.

Scientific summary

Background

Children's early social and emotional development remains a significant predictor of future social, education and health outcomes, and there is substantial evidence linking early social and emotional development to later academic performance and a number of key health outcomes. The recent Marmot Review in England (Marmot M. *Fair Society, Healthy Lives: The Marmot Review. Strategic Review of Health Inequalities in England Post-2010*. Executive Summary. London: Department of Health; 2010) identified the policy objective of giving every child the best start in life as its 'highest policy recommendation' (p. 14), placing particular emphasis on reducing inequalities in the early development of physical, cognitive and non-cognitive skills. Among some of the key recommendations is the need to prioritise developing the capacity of schools to address and improve children's 'social and emotional development, physical and mental health and well-being' (p. 18).

A substantial body of evidence now exists that suggests that well-designed school-based prevention programmes can be effective in improving a variety of social, health and academic outcomes. Roots of Empathy (ROE) is a universal school-based social and emotional learning (SEL) programme that has been developed and implemented in Canada, and has only recently been introduced into the UK. It is delivered on a whole-class basis for one academic year and consists of 27 lessons, which are based around a monthly classroom visit from an infant and volunteer parent (typically the mother) who are usually recruited from the local community. Children learn about the baby's growth and development through interactions and observations with the baby during these monthly visits. ROE is a mentalisation-based programme that aims to develop empathy in children. The labelling of feelings and the exploration of the relationship between feelings and behaviour is achieved through the mother–infant interaction as observed by the children in the classroom.

Several evaluations of ROE have been conducted to date and this report synthesises the findings from these. Of seven eligible studies, only one was a (cluster) randomised controlled trial. The pooled data from these studies suggest that ROE is effective in leading to small improvements in prosocial behaviour [standardised mean difference (SMD) 0.13] and reductions in aggressive behaviour (SMD –0.18). There is no evidence to suggest that it is effective in improving other SEL outcomes among children, in this case empathy and emotional regulation. Only one evaluation studied the longer-term impact of the programme, suggesting that after 3 years the intervention group had poorer prosocial behaviour than the control group [SMD –0.12, 95% confidence interval (CI) –0.17 to –0.07]. With respect to aggressive behaviour 3 years post intervention, the intervention group was displaying only slightly less aggressive behaviour than the control group (SMD –0.06, 95% CI –0.09 to –0.03) and, although statistically significant, this effect was greatly reduced from that observed immediately post test (SMD –0.25).

Objectives

Given the limited existing evidence base for ROE, the aims of the current evaluation are to:

- evaluate the immediate and longer-term impacts of ROE on social and emotional well-being outcomes among 8- to 9-year-old pupils
- evaluate the cost-effectiveness of the programme.

The purpose of the research is to answer the following research questions.

1. What is the impact of the programme post test, and up to 3 years following the end of the programme, on a number of specific social and emotional well-being outcomes for participating children?
2. Does the programme have a differential impact on children depending on their gender, the number of siblings they have and their socioeconomic status and/or the socioeconomic profile of the school?
3. Does the impact of the programme differ significantly according to variations in implementation fidelity found?
4. What is the cost-effectiveness of the programme in reducing cases of aggressive behaviour and increasing prosocial behaviour among school-aged children?

Methods

This study consisted of a cluster randomised controlled trial, a qualitative process evaluation and a cost-effectiveness evaluation.

Sample

Seventy-four primary schools from four of the five trust areas in Northern Ireland were recruited to the trial between March and June 2011. All primary schools and their Year 5 cohort were eligible to take part in the study. Schools were randomly assigned to each of the intervention ($n = 37$) and control ($n = 37$) groups. The intervention schools received the ROE programme in their selected Year 5 class for one academic year (2011/12). The remaining schools in the waiting list control group continued with the regular curriculum and usual classroom activity.

Outcomes and measures

The primary child outcomes are increases in prosocial behaviour and decreases in difficult behaviour as measured by the teacher-rated version of the Strengths and Difficulties Questionnaire (SDQ). Additional data from alternative sources (parent- and child-rated SDQ) and alternative measures (teacher-rated Child Behaviour Scale) were collected in order to triangulate the data. Secondary outcomes included understanding of infant feelings (Infant Facial Expression of Emotions Scale), recognition of emotions (Emotion Recognition Questionnaire), empathy (Interpersonal Reactivity Index), emotional regulation (Child Anger Management Scale), bullying (Revised Olweus Bully/Victim Scale) and quality of life [Child Health Utility – 9D (CHU9D)]. The additional information was the parents' home postcode, the number/age of any siblings the child had, the parents' highest level of qualification and the parents' occupation.

Data collection

Initial pre-test data from the children, parents and teachers were collected in October 2011. The first post-test data were collected in June 2012 and data were collected again at 12, 24 and 36 months. Teachers were asked to complete a questionnaire for each participating child at each time point. Parents were contacted by post and asked to complete a questionnaire and return it to the research team in a Freepost envelope. Field workers administered questionnaires to the children on a whole-class basis.

Seven schools withdrew from the study before the start of the trial; however, retention rates were good overall, with 1182 pupils tested pre test and 902 (76.3%) remaining in the study at the final 3-year follow-up.

Cost-effectiveness analysis

The economic evaluation aimed to conduct:

1. a cost–utility analysis comparing the costs and utilities of the two groups over a 3.75-year period
2. a cost-effectiveness analysis comparing costs and effects between groups such as decreases in difficult behaviour and increases in prosocial behaviour as measured by the SDQ.

The base-case analysis compared the ROE intervention group with the usual classroom activities control group in terms of (1) costs incurred over the 3.75-year period and (2) quality-adjusted life-years (QALYs) gained over the 3.75-year period. Data were collected at five time points: pre test, post test, 12-month follow-up, 24-month follow-up and 36-month follow-up. A cost–utility analysis was undertaken, in which costs considered from a public sector perspective (2014 GBP) and health outcomes were measured by QALYs. Health utilities were measured using the CHU9D. All of the analyses were performed on individual patient-level data, taking clustering into account, and collected from the ROE trial.

Resource use was measured over the length of the trial and made up of the following data collection: (1) resource use resulting from the delivery of the intervention, (2) NHS resource use and (3) societal costs.

Process evaluation

A qualitative process evaluation was conducted alongside the trial to provide in-depth qualitative data on both the implementation and outcomes of the ROE programme. The delivery process of the programme was monitored and tracked across all schools, and a more detailed inquiry of underlying broad patterns outlined from across the schools was the focus of an in-depth case study approach conducted in six of the intervention schools. Interviews and focus groups were carried out with school personnel, local programme co-ordinators, volunteer mothers, children and parents. Observational classroom data were also collected.

Results

Immediately post test

After controlling for pre-test scores and clustering, children who participated in the ROE programme were rated by their teachers as more prosocial (effect size, $g = +0.20$; $p = 0.045$) and as exhibiting less difficult behaviour ($g = -0.16$; $p = 0.06$) than those in the control group.

With regard to the secondary outcomes, children who participated in the ROE programme were able to report a greater number of reasons why infants cry (effect size $g = +0.24$; $p = 0.01$). It is important to note, however, that part of the intervention involves explicitly teaching children about how infants communicate and why they cry, and it is conceivable, therefore, that this measure is biased in favour of the intervention group. Furthermore, the effect is small. No evidence of any differences between the groups was found in relation to the other secondary outcomes.

Prespecified subgroup analyses were undertaken to explore whether or not the programme worked better according to gender, socioeconomic background and number of siblings. No clear or consistent pattern emerged to suggest that there are underlying differential effects. The programme was found to have been uniformly delivered with high fidelity across all intervention schools. It was, therefore, not possible to assess the potential moderating effects of varying levels of fidelity on the outcomes achieved.

Effects at 12-, 24- and 36-month follow-up

The initially positive effects on prosocial behaviour were found to disappear at all subsequent time points. Moreover, there were no statistically significant differences in scores between those in the intervention group and those in the control group at any of the subsequent follow-up time points for any of the other outcome variables (at 12, 24 or 36 months post intervention). There also remained no clear or convincing pattern of any subgroup effects at any of these subsequent time points.

However, and in relation to total difficulties (as measured by the teacher-rated SDQ), the effect size immediately post test appears to have been maintained at the 12-month ($g = -0.14$), 24-month ($g = -0.13$) and 36-month ($g = -0.14$) follow-up time points. However, and because of the reduction in sample size owing to attrition, this effect is not statistically significant and so it needs to be treated with a degree of caution.

Sensitivity analysis

Multiple imputation was used to test whether or not attrition introduced any bias into the findings; however, the findings using the imputed data sets are broadly similar to those using just the observed data.

Cost-effectiveness analysis

Overall, it is estimated that the average cost of delivering ROE is £4057 per school and £175 per pupil. The incremental cost of delivering ROE was £153 (95% CI £14 to £292). The incremental QALY gain from ROE was 0.0160 (95% CI -0.0143 to 0.0462). Against generally accepted national guidelines, the findings of this present study suggest that ROE is a cost-effective intervention. In particular, the National Institute for Health and Care Excellence suggests that interventions costing the NHS < £20,000 per QALY gained are cost-effective. It also suggests that those costing between £20,000 and £30,000 per QALY gained may be cost-effective. For the present evaluation, the incremental cost-effectiveness ratio was £9571 per QALY gained (95% CI -£87,776 to £106,676). It was found that ROE had an 83.1% chance of being cost-effective at the £20,000 per QALY threshold and a 90.1% chance at the higher threshold of £30,000 per QALY.

Process evaluation

The ROE programme was delivered with high fidelity, with all lessons being delivered in all of the intervention schools. This was seen as being the result of the clearly defined structure of the programme and the strong training and ongoing support provided to ROE instructors in schools. The programme was also very well received overall and was felt to include good resources and be linked in closely with the Northern Ireland curriculum, particularly the element on personal development and mutual understanding. Five key issues emerged from the qualitative process evaluation:

1. A belief among some that it would be beneficial for ROE instructors to be teachers within the school to facilitate stronger communication and planning between the instructor and the class teacher.
2. A perception that the delivery of the programme in the first year may have been a little more challenging, especially for those schools where the ROE instructor was not a teacher within that school.
3. A concern regarding the resources required to deliver the ROE programme, especially if the ROE instructor is to be one of the teachers within the school, and whether or not it is sustainable in the longer term.
4. A concern that the ROE programme lasts for only 1 year and is not followed up in subsequent years. Additionally, and relatedly, there was a view among some that it would be worthwhile building the key knowledge and skills among children at an earlier age and before the ROE programme, with some mentioning the Seeds of Empathy programme.
5. The relative lack of involvement of or engagement with parents around the programme and how this may have been partly restricted because of the emphasis on maintaining fidelity to the existing programme.

Conclusions

First, this trial has provided strong and robust evidence that ROE did have a positive impact on children's behaviours in the directions expected immediately post test. More specifically, there is evidence that the programme enhanced children's prosocial behaviour and some evidence that it reduced difficult behaviour, above and beyond the typical effects associated with attending school.

Second, the trial has also provided clear evidence that, although ROE was originally developed in Canada, it is possible to deliver it extremely effectively and with fidelity in a different country and a different cultural context, in this case Northern Ireland.

Third, the trial found no evidence to support the hypothesised theory of change. It is not possible to conclude with certainty how ROE achieved positive behavioural effects without associated increases in social and emotional outcomes. However, it is clear from the qualitative process evaluation that the

children enjoyed the ROE lessons and that the lessons did, progressively, help to encourage the development of a collective sense of concern and caring for the baby, which may have resulted in a positive shift in the group norms (i.e. class norms) of prosocial behaviour. Peer groups play an important and influential role in the development of children's behaviour and attitudes, and they are an important social context within which individual development takes place. Further research is required to explore this possible explanation for behavioural change.

Fourth, the current ROE programme provides only limited opportunities to engage with parents. However, and as found through the process evaluation, there is significant interest among teachers and some parents for a greater degree of parental involvement in the programme. It is, therefore, recommended that consideration be given to incorporating greater parental involvement in the future.

Finally, the findings are not so positive in relation to the sustainability of initial gains in prosocial behaviour. In this respect, further work would be beneficial in terms of developing a more holistic and progressive curriculum that seeks to use evidence-based programmes such as ROE but in a way that is able to sustain and build on the short-term gains found in a developmentally appropriate way.

Trial registration

This trial is registered as ISRCTN07540423.

Funding

Funding for this study was provided by the Public Health Research programme of the National Institute for Health Research.

Chapter 1 Introduction

This report presents the findings from a cluster randomised controlled trial evaluation of the Roots of Empathy (ROE) programme. This chapter provides the background for the study and a description of the programme. The methodology for the evaluation is outlined in *Chapter 2*. The quantitative findings from the trial regarding the impact of the programme on pupil outcomes and the cost-effective analysis are reported in *Chapter 3* and the findings from the accompanying qualitative process study are set out in *Chapter 4*. Key issues emerging from the findings are set out in *Chapter 5*.

Rationale for current study

There is a growing consensus in academic and policy circles regarding the importance of attending to young children's social and emotional well-being. There is substantial evidence that links early social and emotional development to later academic performance¹ and a number of key health outcomes, such as stress and mental health.² Deficits in basic skills, such as the ability to identify emotions, tend to have wide-ranging implications, including being rejected by others and excluded from peer activities and being victimised.³ Such deficits are also related to lower peer-rated popularity and teacher-rated social competence.^{4,5} Chronic physical aggression during primary school also increases the risk of violence and delinquency throughout adolescence in boys.^{6,7} In turn, this can lead to destructive forms of emotion management, such as alcohol abuse.

In recognition of this, a comprehensive set of public health guidelines was published by the National Institute for Health and Care Excellence (NICE) in 2008, the aim of which was to encourage the promotion of social and emotional well-being in primary school children.⁸ According to the guidelines, child well-being not only is important in its own right but can also be a determinant of success in school and physical health. The guidelines recommend that schools create an ethos that supports positive behaviours for learning and successful relationships, and also provide an emotionally secure and safe environment that protects against bullying and violence and offers teachers and practitioners the support they need in developing children's social and emotional well-being.

However, perhaps the most significant recent development has been the publication of the Marmot Review in England.⁹ At the heart of the review's key recommendations is the policy objective of giving every child the best start in life. Of the six policy objectives identified by the review, this one was held up as its 'highest policy recommendation' and reflected the review's life course perspective. Alongside a call to increase the proportion of overall expenditure allocated to the early years, the review also placed an emphasis on reducing inequalities in the early development of physical and emotional health and cognitive, linguistic and social skills, and thus building resilience and well-being among young children. This should be done, according to the Marmot Review, through investment in 'high quality maternity services, parenting programmes, childcare and early years education to meet need across the social gradient' (p. 16).⁹

A second, linked, policy objective identified by the review is to enable all children, young people and adults to maximise their capabilities and have control over their lives. This, in turn, should be achieved by ensuring that schools, families and communities work in partnership to improve health, well-being and resilience. Among some of the key recommendations made in this regard is the need to prioritise developing the capacity of schools to address and improve children's 'social and emotional development, physical and mental health and well-being' (p. 18).⁹

Most recently, a report commissioned by the Early Intervention Foundation in the UK (March 2015) explored the relationship between social and emotional skills in childhood and long-term effects into adulthood.¹⁰ The authors found that self-control and self-regulation are the most important childhood social and emotional skills in relation to positive adult outcomes. Similarly, they found that self-perception,

self-awareness and social skills were important influences on many adult outcomes. There was no clear evidence linking motivation or resilience to adult outcomes; however, emotional well-being in childhood was found to be important for mental well-being as an adult. An equally important finding from this report was that social and emotional development is just as important as cognitive development, if not more so, in some respects, for future life.

Scientific background

A substantial body of evidence now exists to suggest that well-designed school-based prevention programmes can be effective in improving a variety of social, health and academic outcomes for children and young people.^{11,12} Several reviews have been conducted in the area of social and emotional learning (SEL) programmes and, although the types of intervention, participants and outcomes have varied between reviews, the consensus is that well-designed universal school-based programmes have a positive impact on child outcomes.^{13–16}

The most relevant and recent of these reviews is Durlak *et al.*'s¹⁷ meta-analysis, which focused exclusively on school-based universal SEL programmes and their impact on a number of pupil outcomes, including SEL skills, attitudes, positive social behaviour, conduct problems, emotional distress and academic performance. The analysis comprised 213 programmes and 270,034 pupils. The mean effect sizes for each outcome ranged from –0.22 (conduct problems) to 0.57 (SEL skills), which, the authors note, is consistent with effect sizes reported by other studies and reviews of similar programmes and outcomes. The most effective SEL programmes in this review (defined as those that had a significant and positive impact on all six outcomes) were those that did not experience implementation problems and, consistent with Payton *et al.*'s¹⁴ conclusions,¹⁴ also incorporated the following four recommended practices commonly referred to as 'SAFE':

- Sequenced – applying a planned set of activities to develop skills in a step-by-step fashion
- Active – using active forms of learning (i.e. role plays, behavioural rehearsal with feedback)
- Focused – devoting sufficient time to developing social and emotional skills
- Explicit – targeting specific social and emotional skills.

Durlak *et al.*¹⁷ concluded that SEL programmes tended to have a significant and positive impact on students' social and emotional competence, increase prosocial behaviour, reduce conduct and internalising problems, and improve academic performance. They also reported that in those studies that followed up participants, these effects remained statistically significant for at least 6 months post intervention.

However, only a small number of studies in this review (15%) reported follow-up data that met the inclusion criteria, and so little is known about the long-term effects of SEL programmes. Adi *et al.*,¹⁸ whose review informed the NICE guidelines, reinforced this view and observed that, although programmes teaching social skills and emotional literacy show promise, there remains a need for good-quality trials to assess these programmes' long-term effectiveness.

More recently, the Early Intervention Foundation's second of three reports on social and emotional skills in childhood focused on what works in the UK to promote such skills in childhood and adolescence.¹⁹ The authors found that school-based and targeted programmes were most effective, along with those interventions that adopted a 'whole school' approach, involving staff, parents and the wider community. The evidence of the effectiveness of UK out-of-school programmes was less clear-cut.

Existing impact evaluations of the Roots of Empathy programme

The ROE website (www.rootsofempathy.org) reports a number of evaluations that have been conducted to date. To provide further context for this present study, an attempt has been made to identify these existing

studies and to synthesise the data. Full details of the studies identified and of the methods used for the meta-analysis are provided in *Appendix 1*.

In total, 10 studies^{20–29} were identified that had reports that provided sufficient information to assess their eligibility for the meta-analysis. A further six studies were referenced but it was not possible to locate the full report or the data for these. Although the authors of these were contacted directly, the research team had not received a response at the time of writing. Of the 10 reports found, three were excluded because they did not meet the eligibility criteria (i.e. employed an experimental or quasi-experimental design, quantitatively measured at least teacher-rated prosocial and aggressive/difficult behaviour, and collected outcome data at both pre and post test).

Of the seven eligible studies,^{20–26} only one²⁰ was a (cluster) randomised controlled trial that also tracked children for 3 years following the end of the programme. The remaining six studies employed quasi-experimental designs with pre- and immediately post-test data only. Four of the studies,^{20–23} including the cluster randomised controlled trial, were conducted in Canada, two were conducted in Scotland^{24,25} and one was conducted in Australia.²⁶

A total of 4140 primary school aged children from 145 schools took part in the seven eligible studies.^{20–26} Sample sizes ranged between 132 and 785 children, with an average sample size of 591. All of the evaluations measured teacher-rated prosocial and aggressive behaviour using valid and reliable instruments. A range of other teacher and child rated outcomes were also measured, but this meta-analysis focuses only on synthesising the effects for the most commonly measured outcomes:

- teacher-rated *prosocial behaviour* immediately post test (all seven previous studies)
- teacher-rated *aggressive behaviour* immediately post test (all seven previous studies)
- child-reported *empathy* immediately post test (five studies)
- child-reported *emotional regulation* immediately post test (two studies).

Full details of the studies included and excluded, and also of the methods used for the meta-analysis, are provided in *Appendix 1*. The findings are summarised in *Table 1*. As can be seen, when the available data from the seven studies are pooled there is evidence that ROE is effective in leading to small improvements in prosocial behaviour [standardised mean difference (SMD) 0.13] and reductions in aggressive behaviour (SMD –0.18). However, and interestingly, there is no evidence to suggest that it is effective in improving other SEL outcomes among children, in this case empathy and emotional regulation.

As noted, only one evaluation²⁰ studied the longer-term impact of the programme. This is the only pre-existing randomised controlled trial for which there are data, and it suggests that after 3 years the intervention group had *poorer* prosocial behaviour than the control group [SMD –0.12, 95% confidence interval (CI) –0.17 to –0.07]. With respect to aggressive behaviour 3 years post intervention, the intervention group were displaying only slightly less aggressive behaviour than the control group (SMD –0.06, 95% –0.09 to –0.03) and, although statistically significant, this effect was much reduced compared with that observed immediately post test (SMD –0.25).

TABLE 1 Summary of meta-analyses of previous evaluations ($n = 7$) of the ROE programme

Outcome	Pooled sample	Pooled SMD (95% CI)
Teacher-rated prosocial behaviour	1895 ROE, 1617 control (seven studies)	0.13 (0.06 to 0.19)
Teacher-rated aggressive behaviour	1897 ROE, 1626 control (seven studies)	–0.18 (–0.33 to –0.03)
Child-reported empathy	1186 ROE, 861 control (five studies)	0.10 (–0.05 to 0.25)
Child-reported emotional regulation	699 ROE, 655 control (two studies)	0.03 (–0.08 to 0.14)

Overall, therefore, although the findings from existing evaluations of ROE are promising, they raise interesting questions regarding the apparently mixed effects of the programme immediately post test and also about whether or not such effects are sustained in the longer term. Moreover, the current evidence base is limited to only one randomised trial that is also the only study to date that has considered the longer-term effects of the programme. In addition, no study to date has included a cost-effectiveness analysis. This, then, provides the rationale for the present evaluation.

Objectives

The aims of the current evaluation are to:

- evaluate the immediate and longer-term impact of the ROE programme on social and emotional well-being outcomes among pupils aged 8–9 years
- evaluate the cost-effectiveness of the programme.

The purpose of the research is to answer the following research questions.

1. What is the impact of the programme post test and up to 3 years following the end of the programme on a number of specific social and emotional well-being outcomes for participating children?
2. Does the programme have a differential impact on children depending on their gender, the number of siblings they have and their socioeconomic status and/or the socioeconomic profile of the school?
3. Does the impact of the programme differ significantly according to variations in implementation fidelity found?
4. What is the cost-effectiveness of the programme in reducing cases of aggressive behaviour and increasing prosocial behaviour among school-aged children?

The full protocol for this trial, published in August 2011 before ethics approval was sought for the study and, thus, before the recruitment of schools and pre-testing, can be found at the National Institute for Health Research Evaluation, Trials and Studies website.³⁰

Chapter 2 Methodology

Introduction

This chapter begins by setting out the methodology for the trial in relation to sampling, outcomes and measures, data collection and analysis plan. It then describes the methodological approach adopted for the qualitative process evaluation and the approach being taken for the cost-effectiveness analysis. It concludes by identifying a small number of minor changes to the evaluation from the original published protocol.

Trial design

This study is based on a cluster randomised controlled trial and qualitative process evaluation undertaken in four of the five health and social care trust areas in Northern Ireland. The random allocation of schools to either the intervention or the control condition was carried out on a 1 : 1 basis. Ethics approval was granted by the School of Education, Queen's University Belfast, Research Ethics Committee on 2 September 2011.

Deviations of the evaluation from the original protocol

The protocol for this evaluation was published in August 2011, before ethics approval was secured and, thus, before the recruitment of schools and pre-testing. It has not been amended since and is available online.³⁰ The trial was also registered with an International Standard Randomised Controlled Trial Number (ISRCTN) in December 2011 (ISRCTN07540423). The original aims and objectives of the evaluation have not been altered and the overall approach to the research design in relation to the cluster randomised controlled trial, the process evaluation and the cost-effectiveness evaluation have also remained unchanged. The specified primary and secondary outcomes, with their accompanying measures, have also remained the same, as has the proposed analysis plan.

Within this, a small number of minor deviations to the original protocol, published in 2011, have been made, and these are detailed below.

Missing secondary outcome measures

It was not possible to collect data on two of the secondary outcomes specified in the original protocol: educational attainment and class detentions. With regard to education attainment, and at the start of the trial in 2011, the use of standardised InCAS literacy and numeracy tests became compulsory for children in Years 4–7 in Northern Ireland (www.cem.org/incas). Before this, there was no statutory testing in Northern Ireland primary schools. The evaluation planned to take advantage of this new statutory testing as a convenient means of collecting (directly from schools) literacy and numeracy attainment data for the sample immediately post test (June 2012) and for follow-up data sweeps in June 2013 and 2014 while the children were still in primary school. However, schools raised serious concerns about the reliability of the tests (and not related to this present trial), and these were abandoned the following year, 2013. Moreover, this had an impact on data collection in 2012. Unfortunately, therefore, only partial data were available from schools in 2012, with InCAS data available for only approximately 300 pupils in our sample collected in 2012.

Principals advised the research team to, instead, collect the results from the Progress in English and Progress in Maths tests that some schools use with their Year 6 children (our cohort in 2013 and at the 12-month follow-up). Unfortunately, not all schools used this test, and so the team have these data for only around 850 pupils in the sample immediately post test. There was no resource within the study budget for the purchase and administration of independent attainment tests and, for these reasons, it has not been possible to include a reliable measure of educational attainment.

In relation to the other outcome, the team originally intended to collect a class-level measure of behaviour via class detention rates. However, this proved not to be possible, as primary schools did not use detention as a means of punishment for poor behaviour, this being more common in post-primary schools. Primary school pupils may be suspended or expelled, but the instances of these events are extremely rare. For this reason, it was not possible to collect valid (or any) data on this outcome.

Missing and additional covariates for the main analysis

The original proposal stated that, for the main analysis, a series of covariates would be added to each statistical model representing the baseline (pre-test) scores for each of the outcome measures used, as well as measures representing the child's core characteristics. As detailed by the models in *Appendices 2 and 3*, this approach to the analysis was followed. However, there were a small number of covariates not included in these main models associated with data that were collected on the children's core characteristics. More specifically, these were:

- the parents' highest qualifications received
- the parents' occupations
- the number of siblings in the family home.

Data for these three variables were collected from the parents directly, but, given the lower response rates from parents, it was decided not to include these as covariates in the final model, as the number of missing data would have reduced the effective sample size by half. However, a separate sensitivity analysis in relation to the primary outcomes was undertaken to compare the findings of the main analysis with those of an alternative analysis that included these three additional covariates. This sensitivity analysis is provided in *Appendix 3*. As can be seen, the sensitivity analysis suggests that the exclusion of these three covariates did not have a notable impact on the findings presented in this report.

In addition, and following advice from the Trial Steering Committee, the decision was taken to add an additional covariate to all of the statistical models consisting of three dummy variables representing the location of each school in relation to the four health and social care trusts participating in the trial. Because the randomisation of schools occurred within each of the four trusts, it was felt appropriate to control for any possible design effects resulting from this by the inclusion of these dummy variables.

Missing assessment of external validity using propensity scores

Finally, the original protocol included a proposal to assess the external validity of the findings arising from the trial using propensity scores to compare the characteristics of trial participants with those of the population as a whole in Northern Ireland. Such an analysis would have required individual-level data not only for children participating in the trial but also for children from the wider population in the region. Unfortunately, the data required to undertake such an analysis were, subsequently, found not to be readily available and there was no resource in the project budget to cover the costs required to collect these. Therefore, the research team decided to assess the external validity of the trial through aggregate comparisons of the characteristics of the sample of children participating in the trial with those of the population as a whole.

Participants

Seventy-four primary schools (clusters), from four of the five trust areas in Northern Ireland, were originally recruited to and enrolled into the trial [Belfast Health and Social Care Trust (HSCT), South Eastern HSCT, Southern HSCT and Western HSCT] by trust personnel between March and June 2011. All primary schools were eligible to take part and all Year 5 children within each school were also eligible to participate. Before randomisation, school-level consent was sought from the principal of each participating school. Parental consent was sought post randomisation for each child participating in the trial. Teacher- and child-rated data were collected in the school setting and parent-rated data were collected via a postal questionnaire (see *Report Supplementary Material 1* for copies of the research instruments used).

Intervention

Roots of Empathy is among a small number of named universal school-based SEL programmes that has an existing evidence base regarding its effectiveness. It is a universal programme that has been developed and implemented in Canada, and it has only recently been introduced into the UK. It is delivered on a whole-class basis for one academic year (October to June). It consists of 27 lessons, which are based around a monthly classroom visit from an infant and parent, usually recruited from the local community, whom the class 'adopts' at the start of the school year. Children learn about the baby's growth and development through interactions with and observations of the baby during these monthly visits.

Each month, a trained ROE instructor (who is not the class teacher) visits the classroom three times for a pre-family visit, the visit of the parent and infant, and a post-family visit. Instructors undergo a total of 4 days' intensive training that is delivered directly by a specialist ROE trainer from Canada. The specialist trainer also provides ongoing mentoring support to all instructors via regular telephone calls. In addition, ongoing support is available to each instructor through the health and social care trust's lead ROE co-ordinator. Each ROE lesson takes place in the classroom, and the class teacher is present but not actively involved in delivery. The programme provides opportunities to discuss and learn about the different dimensions of empathy, namely emotion identification and explanation, perspective-taking and emotional sensitivity. The parent-and-infant visit serves as a springboard for discussions about understanding feelings, infant development and effective parenting practices. The intervention is highly manualised, and any adaptation or tailoring of either the content or the method of delivery is discouraged by the ROE organisation.

Roots of Empathy seeks to develop children's social and emotional understanding, promote prosocial behaviours, decrease aggressive behaviours, and increase children's knowledge about infant development and effective parenting practices. At the heart of the programme is the development of empathy among young children. Empathy is the capacity to recognise and, to some extent, share the feelings experienced by others. Baron-Cohen³¹ describes empathy as spontaneously and naturally tuning into the other person's thoughts and feelings. It is believed that the existence of empathy lays the basis for helping other people and for other forms of prosocial behaviour because it underpins the motivation to respond to the feelings of others. Similarly, it is suggested that the absence of empathy leaves a person to consider their own needs without reference to the feelings of others, which results in asocial or antisocial behaviour, depending on the degree of impact on the other person.

Baron-Cohen³¹ suggests that there are two major elements to empathy: cognitive (perspective-taking) and affective (sharing the feeling of the other person). The cognitive element of empathy is less problematic in some respects because the capacity for perspective-taking occurs as part of a wider developmental pattern of growth (as described by Piaget^{32,33}). The feeling element, on the other hand, is considered to develop mainly in response to close personal relationships, the prototype for which is the attachment bond between mother and child. The centrality of the attachment relationship was first established by Bowlby³⁴ and developed later by Ainsworth *et al.*³⁵ to include patterns of attachment between caregiver and child. For Ainsworth and many subsequent researchers, secure attachment is regarded as the basis for sound psychological development.

The means through which attachment has beneficial effects on development is still not fully understood. Fonagy *et al.*³⁶ argue that securely attached individuals tend to have more robust capacities to represent the state of their own and other people's minds. This ability to perceive and interpret human behaviour in terms of intentional mental states (e.g. needs, desires, beliefs, goals, purposes and reasons) is known as mentalisation. The concept of mentalisation is receiving increasing empirical support as a core process in the attachment relationship. It appears, however, that mentalisation can be acquired outside infancy and, indeed, there is a form of mentalisation therapy used in adults for which an evidential basis has been developed.³⁷

A characteristic of ROE is that it is a mentalisation-based programme with the principal aim of developing empathy in children. The labelling of feelings and the exploration of the relationship between feelings and behaviour is achieved through the mother–infant interaction that is observed by the children in the classroom. Clearly, the baby cannot communicate in words and can only express his or her feelings through behaviour. For this reason, the baby provides an ideal opportunity for the children to learn mentalisation skills through interpreting and labelling the baby's emotions and, by this means, to learn the affective and cognitive components of empathy, which will enable them to empathise with others. If and when children learn empathy, they then have the basis for developing positive social partnerships with others, as depicted in the logic model, shown in *Figure 1*, that has been developed by the present authors to summarise the implicit theory of change underpinning the programme.

Outcomes

The outcomes measured in this trial are based on the logic model (see *Figure 1*). The primary child outcomes are increases in prosocial behaviour and decreases in difficult behaviour as measured by the teacher-rated version of the SDQ. Additional data from alternative sources (parent- and child-rated SDQ) and alternative measures [teacher-rated Child Behaviour Scale (CBS)] were collected to allow the triangulation of the data and to confirm the reliability of the primary outcome measures (i.e. the teacher-rated SDQ). This is discussed further in *Chapter 3*.

The secondary outcomes largely reflect the key precursors expected to lead to behavioural change. The exceptions to this are bullying and quality of life, for which it is hypothesised that improvements are likely to flow from improved behavioural change. A description of the measurement of each outcome is given in *Table 2*.

Alongside completing the SDQ for their child, parents were asked to provide additional contextual information: their home postcode, the number and age of siblings, their education qualifications and their occupation. A proxy measure of socioeconomic status was determined from the Northern Ireland Multiple Deprivation Measure (NIMDM) 2010,⁴⁷ derived from the child's home postcode. Although entitlement to free school meals is often used as a proxy indicator of deprivation, concerns have been raised about its robustness, as it reflects only low income.⁴⁸ The NIMDM 2010, on the other hand, provides a relative measure of deprivation by collating information across a spectrum of factors (e.g. income deprivation, health deprivation, employment deprivation and living environment). The geographic areas corresponding to each home postcode are ranked according to overall deprivation. Rankings can range from 1 (most deprived) to 890 (least deprived), and in the current sample ranks ranged from 1 to 889.

Data collection

The research instruments used for data collection are provided in *Report Supplementary Material 1*. Initial pre-test data [time (T) 0] from the children, parents and teachers were collected in October 2011 across all participating schools before the first sessions of ROE were delivered in the intervention schools. The first (immediately) post-test data (T1) were collected in June 2012 and again, annually, at 12 (T2), 24 (T3) and 36 months (T4). The final sweep of data collection took place in June 2015 when children were 11–12 years of age and at the end of their first year in secondary school.

Teachers were asked to complete questionnaires, which included the SDQ and the CBS, for each participating child in their class at each time point. When children were attending primary school, their class teacher completed the questionnaire. At the final sweep of data collection, when children were completing their first year of secondary school, their form teacher (or the teacher who knew the child best) was asked to complete the questionnaire. Thus, a different teacher completed the questionnaire at each sweep of data collection.

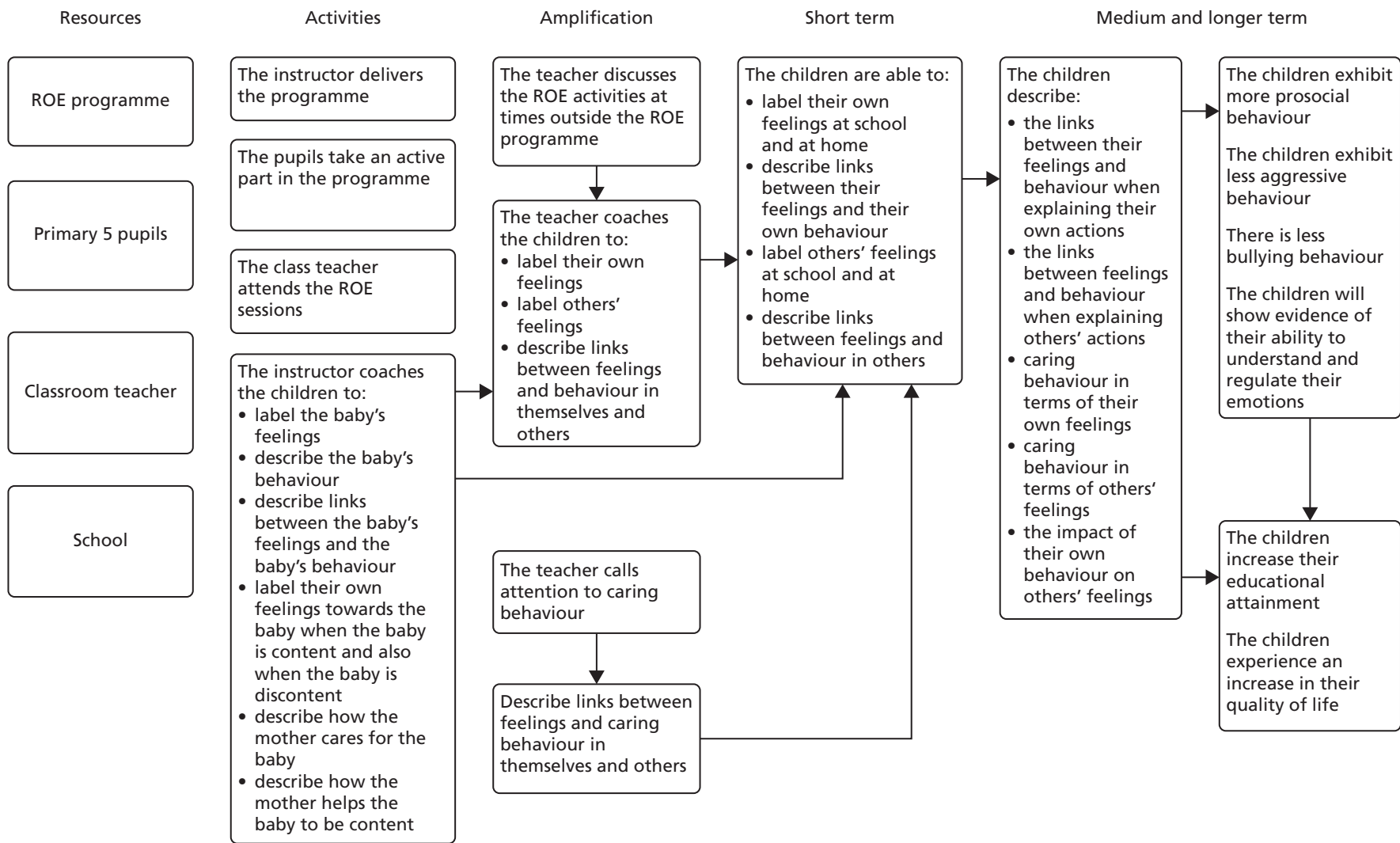


FIGURE 1 Logic model for the ROE programme.

TABLE 2 Description of the outcomes, measures and reliability pre and post test

Outcomes	Measures	Reliability (Cronbach's alpha)
Primary outcomes		
Prosocial behaviour	SDQ ³⁸	Teacher ratings
Difficult behaviour	<ul style="list-style-type: none"> The SDQ is a screening instrument used to detect mental health problems in children. It comprises 25 items, each rated on a 3-point scale (0 = not true, 1 = somewhat true, and 2 = certainly true). The items relate to five subscales covering distinct domains of psychological adjustment in children and adolescents; conduct problems, peer problems, emotional symptoms, hyperactivity and prosocial behaviours. The means of the subscales, with the exception of the prosocial scale, give a 'total difficulties' score. Mean composite scores on the total difficulties and prosocial subscales could range from 0 to 2. The SDQ was completed by both teachers and parents pre test, post test and at follow-up. It was also completed by pupils when they reached the appropriate age at T3 and T4 	<ul style="list-style-type: none"> SDQ prosocial subscale = 0.83–0.86 SDQ total difficulties = 0.88–0.89 CBS aggression subscale = 0.92–0.94 CBS prosocial subscale = 0.90–0.92
	CBS ³⁹	Parent ratings
	<ul style="list-style-type: none"> The CBS is a measure of children's behaviours with peers in the school context, and was completed by teachers. The 17 items that make up the 'aggression' (nine items) and 'prosocial behaviours' (eight items) subscales were used in the current evaluation, as these constitute the main outcomes. Responses to each item are on a 3-point scale (0 = not true, 1 = sometimes true and 2 = often true), and a mean composite score was computed for each subscale (range 0–2) 	<ul style="list-style-type: none"> SDQ prosocial subscale = 0.72–0.73 SDQ total difficulties = 0.86
Secondary outcomes		
Understanding of infant feelings	Infant Facial Expression of Emotions Scale ²¹	
Understanding of how to help a baby who is crying	<ul style="list-style-type: none"> This involved the children looking at a picture of an infant crying, and asking them to write down the possible reasons that the baby is crying, as well as ways to help a baby who is crying. Children were asked to give as many answers as they could think of, and responses were coded and summed for two scales: reasons a baby cries and ways to help a crying baby. Higher scores indicate a greater understanding of infant feelings 	
Recognition of emotions	ERQ ⁴⁰	Emotion recognition = 0.58–0.61
	<ul style="list-style-type: none"> This consists of 16 short vignettes, which were read aloud to the children, who were asked to correctly identify how a child would feel (from happy, sad, angry or scared). A score of 1 was given for each correctly identified feeling, and a mean composite score was computed across the 16 items to determine a total emotion recognition score, which could range from 0 to 1 (higher scores reflecting a greater recognition of emotions) 	

TABLE 2 Description of the outcomes, measures and reliability pre and post test (continued)

Outcomes	Measures	Reliability (Cronbach's alpha)
Empathy	IRI ⁴¹ <ul style="list-style-type: none"> The was adapted for use with children, first by Litvack-Miller <i>et al.</i>⁴² and again by Garton and Gringart.⁴³ Eighteen items reflect the two main components of empathy: affective (e.g. 'I want to help people who get treated badly') and cognitive (e.g. 'I sometimes try to understand my friends better by pretending I am them). Items are rated on a 5-point scale, ranging from 'not at all like me' (1) to 'very like me' (5). Mean responses to the 18 items were computed to give a total empathy score, which could range from 1 to 5 (higher scores indicating greater empathy) 	Total empathy = 0.85–0.86
Emotional regulation	CAMS ⁴⁴ <ul style="list-style-type: none"> This scale contains 11 items, covering three areas: inhibition of anger expression (four items, e.g. I hold my anger in); coping, or anger control (four items, e.g. I try to calmly deal with what is making me feel mad); and dysregulation of anger expression (three items, e.g. I say mean things to others when I'm mad). Responses to each item are on a 3-point scale (1 = not very often true, 2 = sometimes true and 3 = often true). All 11 items loaded onto 1 factor; a total mean score for 'anger management' was therefore calculated, with the dysregulation items reverse scored. Anger management scores ranged from 1 to 3, with higher scores reflecting greater anger management 	Anger management = 0.69–0.77
Bullying	Revised Olweus Bully/Victim Questionnaire ⁴⁵ <ul style="list-style-type: none"> Ten victimisation items from the junior version of the scale, suitable for use with primary school pupils, were used. Seven items pertain to different types of bullying behaviour (e.g. being made fun of, left out of activities, bullied physically or threatened), and children were asked to indicate how frequently, if at all, they had experienced each type of bullying in school over the past couple of months. Responses were given on a 5-point scale that ranged from 'it hasn't happened to me' to 'several times a week', and the mean of the seven items was computed to give a total bullying score. This score could range from 1 to 5, with higher scores reflecting experiencing more types of bullying in school more often. In addition to measuring the extent to which children were victims of bullying behaviour, the bully scale of this measure was also used in the final two data sweeps (T3 and T4) to determine the extent of bullying behaviour that children exhibited 	Bullying = 0.83–0.84
Quality of life	CHU9D ⁴⁶ <ul style="list-style-type: none"> This self-report instrument was developed for use with children aged 7–11 years, and measures nine dimensions of health-related quality of life (worry, sad, pain, tired, annoyed, school work, sleep, daily routine and ability to join in activities), each with five levels representing increasing levels of severity. The scores for each item were reverse coded, and a mean composite score was computed across the nine items. Scores could range from 1 to 5, and a higher score indicated better quality of life 	Quality of life = 0.71–0.74

CAMS, Child Anger Management Scale; CHU9D, Child Health Utility – 9D; ERQ, Emotion Recognition Questionnaire; IRI, Interpersonal Reactivity Index.

Consenting parents were contacted by post and asked to complete a questionnaire, which included the SDQ and background information on family composition, parental education and employment, and return it to the research team in a freepost envelope. Parents were given the option of completing the questionnaire via telephone interview, but none chose this.

Experienced field workers visited each school and administered questionnaires to the children on a whole-class basis. Fieldworkers were fully trained and co-ordinated by the research team. The children's questionnaires covered the secondary outcomes detailed in *Table 3* (emotional regulation, empathy, recognition of emotions, understanding of infant crying, bullying and quality of life). Children were asked not to confer, and this was ensured by the field worker and the class teacher. Each question was read aloud to the class and any difficult words/phrases were explained. Depending on the ability level of the group, testing took between 30 and 40 minutes. If a child was absent on the day of testing, efforts were made to return to the school at a later date and test these children separately. The procedure and materials were pilot tested by the research team during an earlier, feasibility study of ROE implementation. *Figure 2* presents a flow diagram of teacher, pupil and parent responses through the trial. As can be seen, barring the seven schools that withdrew before the start of the trial, retention rates were good overall, with 1278 pupils tested pre test (583 control and 695 intervention pupils) and 949 remaining in the study at the final 3-year follow-up (76.3%) and included in the analysis (424 control and 525 intervention pupils, i.e. 74.3%). However, parental engagement was less successful, with only 686 returning data pre test (53.7% of the sample of 1278 children tested), reducing to 506 at the end of the study that were included in the analysis (234 control and 272 intervention parents, i.e. 39.6%).

Sample size

The sample size calculation was based on the following assumptions.

- Previous evaluations of ROE, together with the wider meta-analysis of SEL programmes, suggested effects that would range in magnitude between $d = 0.22$ and $d = 0.57$.
- For the primary outcome measure (SDQ), typical intraclass correlation coefficients (ICCs) have been found to range between 0.05 and 0.15.
- With the inclusion of the relevant pre-test scores and other covariates, it is also reasonable to assume that the multilevel models used to estimate the effect sizes of the intervention will be able to account for approximately 20% of the variation in post-test outcome scores.

Thus, it was estimated that for the trial to be able to detect the lower bound anticipated effect size of $d = 0.22$ with between 85% power (for $ICC = 0.05$) and 60% power (for $ICC = 0.15$), a sample size of 70 schools with an average class size of 33 children [i.e. 630 children per arm (1260 in total)] would be required. For the highest estimate of $ICC = 0.15$, the trial would achieve sufficient power (80%) for effects of $d \geq 0.28$. These estimates were calculated using Optimal Design (version 2.0) (<https://sites.google.com/site/optimaldesignsoftware/>).

Randomisation

An independent statistician from the Northern Ireland Clinical Trials Unit undertook the (1 : 1) random allocation (stratified by health and social care trust area) of enrolled schools and assigned 37 schools to either the intervention or the control group (*Table 3*).

The schools that were randomly allocated to the intervention group received the ROE programme in their selected Year 5 class for one academic year (2011/12). When there were parallel classes in any specific school, one Year 5 class (8- to 9-year-olds) was randomly selected from these to take part in the trial.

The remaining schools in the control group did not receive the ROE programme but continued with the regular curriculum and usual classroom activity. Schools in the control group were placed on a waiting list

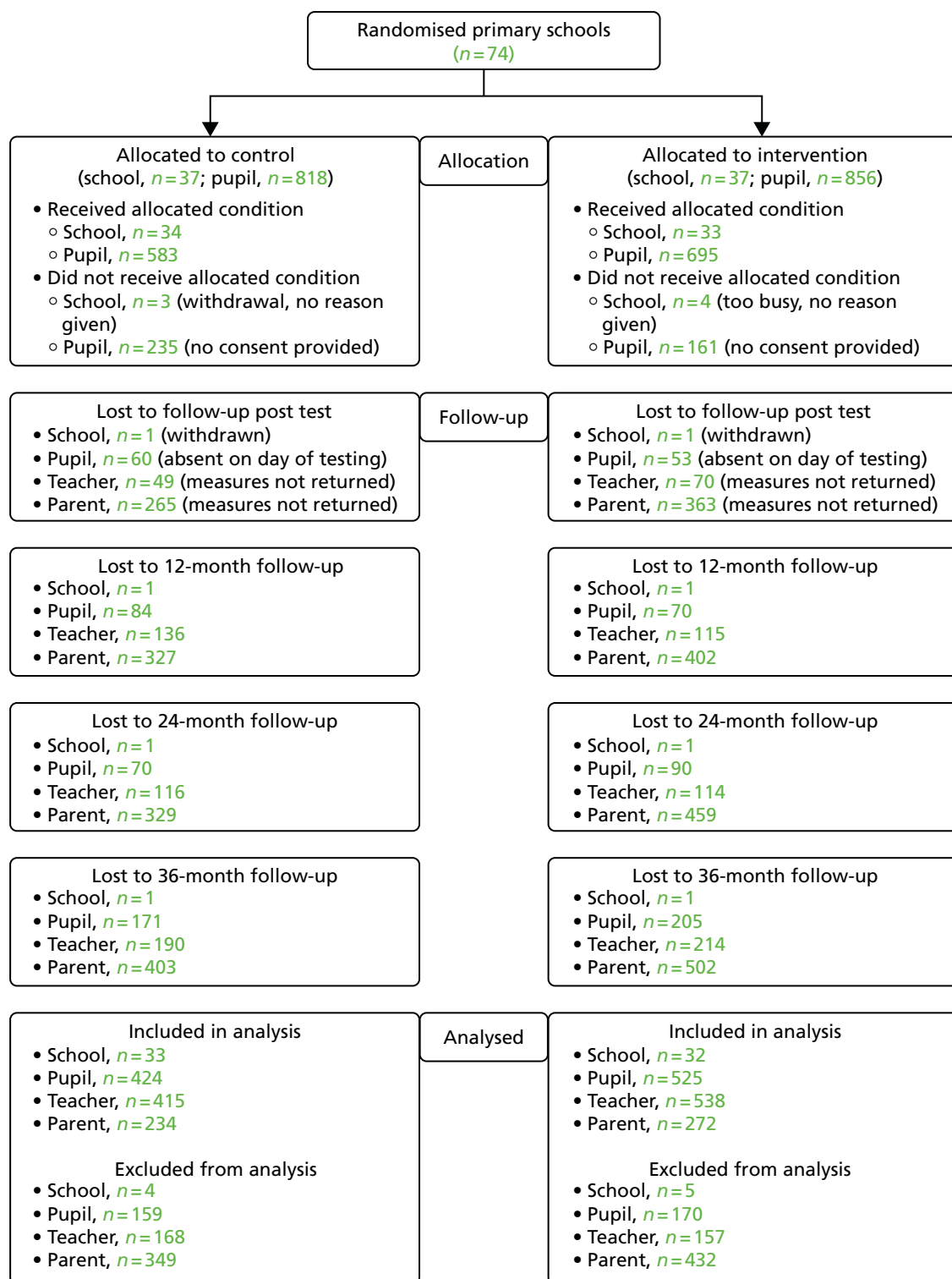


FIGURE 2 Flow diagram of recruitment and testing of children.

to receive the programme in 2012/13, but this was on the understanding that ROE would not be delivered to their current Year 5 cohort as they progressed through Years 6 and 7.

The Northern Ireland Clinical Trials Unit informed the research team of the allocation outcomes and the research team passed this information to the relevant HSCT personnel, who in turn informed the school.

TABLE 3 Number of intervention and control schools in each of the participating health and social care trusts at pre test and immediately post test

Region	Group, <i>n</i>		Withdrawn pre test, <i>n</i>		Withdrawn post test, <i>n</i>	
	Intervention	Control	Intervention	Control	Intervention	Control
Belfast HSCT	11	11	1	0	1	1
South Eastern HSCT	12	12	1	2	0	0
Southern HSCT	8	8	1	0	0	0
Western HSCT	6	6	1	1	0	0
Total	37	37	4	3	1	1

Statistical methods

Data were entered into IBM SPSS Statistics version 20.0 (IBM Corporation, Armonk, NY, USA) for preparation and preliminary exploration before being analysed using Stata[®] version 14.1 (StataCorp LP, College Station, TX, USA). Data preparation involved checking the proportion of missing data, and checking that minimum and maximum values were within the appropriate range. Descriptive statistics were generated for each variable, and the distribution was checked. The validity of measures [the SDQ,³⁸ the CBS,³⁹ the Interpersonal Reactivity Index,⁴¹ the Emotion Recognition Questionnaire⁴⁰ and the Child Anger Management Scale⁴⁴ (CAMS)] was assessed using factor analysis, and internal reliability was evaluated using Cronbach's alpha. The core characteristics (gender, parental education and familial deprivation) of the control and intervention groups were compared (controlling for clustering) using binary logistic multilevel modelling for categorical data and linear multilevel modelling for continuous data. Differences in mean scores for outcome variables between the control and intervention group were tested using multilevel models to control for effects of clustering.

Owing to the clustered nature of the data, the statistical analysis involved the use of multilevel models with children (level 1) clustered within schools (level 2). This was done separately for the immediately post-test data and then for each subsequent time point. For each of the outcome measures, a linear multilevel model was estimated, with the relevant post-test score being set as the dependent variable and its related pre-test score added as an independent variable together with a dummy variable for whether the child was a member of the control or the intervention group. A series of covariates, including child characteristics (gender and familial deprivation score) and mean pre-test scores for all other outcome measures, were also included in the models.

Before the analysis, all of the outcome variables (pre test and post test) were standardised, as were the covariates. The only exception to this was the dummy variable representing group membership, which remained the same (coded '0' for control group and '1' for intervention group). This meant that the constant in each of the main models represented the standardised group mean for the control group and the coefficient for this dummy variable represented the difference between that and the standardised group mean for the intervention group. As all of the variables had been standardised, this coefficient also represented the effect size associated with the programme for that particular outcome. The effect sizes reported in *Chapter 3* were calculated with the formula for Hedges' *g* and using the estimated post-test mean scores for the control and intervention groups from the statistical models, the standard deviations (SDs) for both groups pre test and their corresponding sample sizes. Not surprisingly, given the sample sizes involved, the effect sizes calculated were essentially the same as those estimated in the models (barring some minor differences due to rounding). Evidence of the effects of the programme is indicated by the statistical significance of the coefficient for the variable for intervention status.

The multilevel models were used to calculate the predicted mean post-test scores for the intervention and control group for the average child such that:

$$\begin{aligned} \text{Post-test score}_{ij} &\sim N(XB, \Omega) \\ \text{Post-test score}_{ij} &= \beta_0 + \beta_1(\text{group}_{ij}) + \beta_2(\text{pre-test score}_{ij}) + \beta_3(\text{gender}_{ij}) + \beta_4(\text{deprivation}_{ij}) \\ &\quad + \beta_{5-15}(\text{remaining 10 pre-test covariates}_{ij}) \\ &\quad + \beta_{16-18}(\text{dummy variables for Health \& Social Care Trust}_{0j}) + u_{0j} + e_{0ij} \\ [u_{0j}] &\sim N(0, \Omega_u); [e_{0ij}] \sim N(0, \Omega_e). \end{aligned} \quad (1)$$

Beyond these main models that were used to estimate the overall effects of the ROE programme, exploratory analyses were also undertaken to assess whether or not the programme was differentially effective for differing groups of pupils in terms of gender, familial deprivation and number of siblings. These analyses were undertaken by extending the main pre-test/immediately post-test models to include an interaction term between the contextual variable of interest and group membership. Evidence of differential effectiveness of the programme was indicated in relation to the statistical significance of this interaction term. Full details relating to all of the multilevel models estimated are provided in *Appendix 2*.

Sensitivity analyses

Two sensitivity analyses were undertaken to assess the reliability of the findings arising from the main analysis outlined above. The first involved comparing the effects of the programme that were estimated with regard to the two primary outcomes using the teacher-rated SDQ with the effects estimated using the alternative parent- and child-rated forms of the SDQ and also the CBS child-rated measures for prosocial behaviour and aggressive behaviour. This analysis assessed the robustness of the findings for the two primary outcomes and, also, with the parent- and child-rated measures, whether or not there was any evidence of bias introduced as a result of the use of teachers' ratings, given that the teachers were not blind to the children's participation in the evaluation.

The second sensitivity analysis explored how the findings may have been affected by attrition and, thus, by missing data. This was undertaken by rerunning the main analyses, conducted using only observed data, using multiple imputation.

Qualitative process evaluation

A qualitative process evaluation was conducted alongside the trial to provide in-depth qualitative data on both the implementation and the outcomes of the ROE programme. The delivery process of the programme was monitored and tracked across all schools and then a more detailed inquiry of underlying broad patterns outlined from across the schools was the focus of an in-depth case study approach conducted in six of the intervention schools.

Selection of the sample

A sample of six of the intervention primary schools were selected purposively from across four of the five health and social care trust areas in Northern Ireland. One of the selected schools declined to take part and another school was substituted, matched as closely as possible to the characteristics of the original school. The six schools selected were all intervention schools. The six case study schools were selected purposively to represent:

- the main types of schools (controlled, Catholic maintained and integrated)
- the trust areas (Southern, South Eastern, Belfast and Western)
- a mix of urban/rural schools
- size (large versus small)
- different catchment areas in terms of socioeconomic background.

School personnel

To explore the impact of the programme on the pupils and the school, interviews were carried out in all six schools with the ROE instructor, the class teacher and the principal (18 interviews). Interview schedules focused on the interviewees' perceptions of the following: the programme's perceived impact; the benefits of the programme; how the children responded to the programme; the value added to the class/school from the ROE programme; what worked well and what worked less well; the main challenges in implementing the programme; and the school's engagement strategy and experience of participation of parents with the programme. Interviews lasted no longer than 60 minutes.

Local programme co-ordinators

The four key point people from the health trusts were interviewed about their role in the implementation of the ROE programme. This included their experiences of schools in which programme implementation had been challenging or straightforward in terms of recruitment, curriculum, strengths and limitations of the programme, engagement of all participants and strategies used to ensure smooth roll-out of the programme in promoting the engagement of schools.

Volunteer mothers

The volunteer mothers from each of the six participating schools were interviewed about their experience of the family visits. The interview schedule focused on how they had heard about the ROE programme; what influenced their decision to participate; how comfortable they felt with the circumstances of the classroom visits; if they had experienced any challenges; their commitment to the programme; what they thought were the main benefits for the children, for them and for their baby; and how the children responded to their baby.

Children

The research team included staff who were very experienced in working with young children and it was proposed that children's views be sought about school and the ROE programme. This was undertaken on a group basis in each of the participating six schools. Six focus groups were conducted, with 6–10 pupils in each, and with specific children purposively selected to represent a range of observed responses to the programme (from resistance/non-response to active engagement). The group sessions were used for exploratory discussion of children's views of ROE, what they learned, what they liked and did not like about the programme and whether or not they talked to their families about it.

Parents

Parents in the intervention schools were invited to participate. They were asked about their views and experiences of the ROE programme and its impact on/contribution to their engagement with the child and the school; what benefits it brought and what the drawbacks were, if any; and whether or not they would recommend the programme to others.

Observational analysis

In total, there are 27 ROE lessons in the school year and there is a plan for every lesson. Nine themes are covered and each is addressed over three visits (pre-family visit, family visit and post-family visit). During the programme delivery (October 2011), classroom observations were conducted on three ROE lessons in each of the case study schools (18 observational visits). Three lessons were observed in each of the case study schools. Observational visits were selected to ensure that a variety of these lessons were observed in each of the case study schools. All of the lessons were tape-recorded and transcribed verbatim. During direct observation, field notes were taken to provide a descriptive summary of the lesson and record a time log of each lesson and written up in detail. The researcher interpreted the data immediately after collecting the field notes.

Ethics, consent and data analysis

Organisational consent was sought via a letter to the principal of each of the six case study schools, informing them of the purpose and nature of the research and outlining what taking part would mean for

them and their school. This letter was followed by a telephone call to each principal to address any concerns or queries and to determine principals' levels of interest in taking part. After principal consent had been obtained, parental and participant consent was sought.

To explain the study to the parents of the Year 5 children, the research team wrote a letter that was sent to parents via the school. The letter explained the study in non-technical terms and invited the parent and child to take part. It outlined the aims of the study, explained what was involved if the parent agreed that their child could take part and highlighted that fact that the participation of the child was voluntary.

Informed verbal consent was sought directly from the children. Children were provided with clear information on consent and confidentiality. Care was taken to make the information understandable, and the children had the opportunity to ask questions about the study. Following this, they were given the opportunity to consent to take part in or to withdraw from the study.

The data from the interviews were analysed using thematic analysis following the approach outlined by Braun and Clarke.⁴⁹ Thematic analysis is a flexible and descriptive method that allows the emergence of a narrative to formulate the important features relevant to the research questions. With prior consent, the qualitative interviews were tape-recorded and transcribed verbatim, and field notes were written up from the lesson observations. To perform the thematic analysis adequately, the coding programme MAXQDA 10 (VERBI GmbH, Berlin, Germany) was used. Six initial categories were created to reflect the research questions of this evaluation, namely:

1. benefits
2. programme content
3. mentoring and support
4. limitations
5. main challenges
6. parental involvement.

Themes that emerged were identified, and illustrative quotations have been used in the report. The observational data were analysed in a similar way to those from the interviews and focus groups, with a thorough process of reading, categorising, testing and refining that was repeated by the researcher until all emerging themes were compared against all of the observations.

Cost-effectiveness analysis

Economic evaluations of school-based population health interventions such as ROE are relatively uncommon, and yet there is growing consensus on the value of investing in children's health.^{50,51} By improving a child's overall health and well-being, that child may perform better in school, reduce their use of costly health-care services and, ultimately, be better prepared for and successful in adulthood in terms of labour and employment outcomes.^{50,51} Furthermore, having good social, emotional and psychological health can affect physical health and can also protect children against emotional and behavioural problems, violence, crime, teenage pregnancy and drug misuse.^{8,18} Beyond the health and social benefits to the individual, such outcomes have long-term economic impacts that need to be evidenced so that investment in such interventions can be justified.

The purpose of the economic evaluation was to answer the following research question:

- What is the cost-effectiveness of the programme in reducing aggressive behaviour and increasing prosocial behaviour among school-aged children?

The aim of this economic evaluation was to determine the cost-effectiveness of the ROE programme compared with usual classroom activity. Specifically, we aimed to conduct a:

- cost–utility analysis comparing costs and utilities of the two groups over a 3.75-year period
- cost-effectiveness analysis comparing costs and effects such as decreases in difficult behaviour and increases in prosocial behaviour as measured by the SDQ between groups.

Usual classroom activity was chosen as the most appropriate comparator, as this is what would normally occur in the absence of ROE.

Methods overview

The base-case analysis compared the ROE intervention group with the usual classroom activities control group in terms of (1) costs incurred over the 3.75-year period and (2) quality-adjusted life-years (QALYs) gained over the 3.75-year period. Data were collected at five time points:

1. pre test (baseline)
2. post test (after intervention completion)
3. 1-year follow-up post test
4. second-year follow-up post test
5. third-year follow-up post test.

The analysis had a time horizon of 3.75 years (45 months), which equates to 3 years’ follow-up after intervention completion, as described above. A cost–utility analysis was undertaken, with costs considered from a public sector perspective (2014 GBP) and health outcomes measured by QALYs. Health utilities were measured using the Child Health Utility – 9D (CHU9D),⁴⁶ a health-related quality-of-life measure designed specifically for children. All analyses were performed on individual patient-level data, taking clustering into account, and collected from the ROE trial. *Table 4* describes the data collected for the economic evaluation.

Costs and QALYs were discounted using NICE’s recommended public health economic evaluation discount rate of 1.5%.⁵² Missing data on costs and QALYs were handled by multiple imputation with chained equations.⁵³ Regression methods were used to obtain incremental cost and effect estimates. Multiple regression methods that ignore clustering (e.g. the within-school clusters as in this trial) can lead to biased coefficients and, especially, biased standard errors.⁵⁴ Multilevel models have been proposed as a method to address issues surrounding clustering in economic evaluation,⁹ and their use was explored. On recognition of the model being a poor fit for costs in this particular data set, regression with robust standard errors was conducted to adjust standard errors by indicating that observations within schools may be correlated but are independent between schools.

TABLE 4 Data from ROE trial collected for the economic evaluation

Data type	Description of data	Time points
Costs of intervention	Fees, training, personnel and materials to run ROE	During trial
NHS/PSS resource use	NHS/PSS service use and medications	FU2, FU3
Cost to society	Social worker, school nurse and police visits	FU2, FU3
Health-related quality of life	CHU9D	Pre test, post test, FU1, FU2, FU3
Demographics	Gender, school, NIMDM 2010, number of siblings	Pre test

FU, follow-up; PSS, Personal Social Services.

Incremental cost-effectiveness ratios (ICERs) were estimated by dividing the difference in mean costs between groups by the difference in mean effects between groups:

$$\text{ICER} = \frac{\Delta \text{Cost}}{\Delta \text{QALY}}. \quad (2)$$

The uncertainty surrounding the ICER was investigated by use of a non-parametric bootstrap of 1000 iterations. This uncertainty was then presented on the cost-effectiveness plane and summarised on the cost-effectiveness acceptability curve. ICER estimates were compared with the £20,000–30,000 per QALY threshold generally accepted by NICE.⁵⁵ To allow for uncertainty, a series of sensitivity analyses were performed. All analyses were conducted as intention-to-treat analyses and performed in Stata/SE version 14.1 (StataCorp, College Station, TX, USA).

Methods

Resource use

Resource use was identified through early discussions with the trial managers and their contacts with the school to identify likely resource use. Resource use was then measured over the duration of the trial and was made up of the following data collection: (1) resource use due to the delivery of the intervention including personnel, training, materials, fees and other costs; (2) NHS resource use, including service use and medications; and (3) societal costs such as social worker, school nurse and potential contacts with the police. These broad-ranging costs were considered from a public sector perspective, as per NICE public health guidance.⁵²

Cost of the intervention

Personnel costs (salary costs) were classified by NHS band and were taken from the 2011/12 Agenda for Change pay scale.⁵⁶ Personnel costs included key point people (band 7), who are health trust employees who co-ordinate ROE in each of the four participating trusts; administrative support (band 3); and ROE instructors (band 6). Salaries were based on mid-spine points for each respective band range, including 25% for on-costs, and adjusted for the time spent delivering ROE. An instructor fee was included, which was a one-time fee paid to each instructor.

The costs of time spent training instructors and of the training materials were also included. Other costs were made up of fees paid to the ROE programme in Canada for use of the programme in the UK. These included programme support costs, materials shipping, trainers and mentoring expenses. The programme fees were originally purchased in 2011 Canadian dollars and converted to the current price year using Purchasing Power Parities reported by the Organisation for Economic Cooperation and Development.⁵⁷ When required, costs were inflated to our base year of 2014 (GBP).

Annuitisation was carried out to spread the costs over the expected 5-year life span of the ROE intervention fixed costs. Annuitisation is typically performed for capital costs such as buildings and equipment; however, other costs such as training and materials may also be annuitised if they are incurred at the start of the programme and yet have a useful life longer than the initial period.⁵⁸ Training, materials and other programme costs were one-time costs and were annuitised over the expected life of the intervention.⁵⁸ The base-case assumption of the expected life of the intervention was assumed to be 5 years; therefore, costs were annuitised over 5 years at a discount rate of 1.5%. The equivalent annual cost was estimated using the annuitisation formula given below, where K = the initial outlay, E = the equivalent annual sum, n = the time period and r = the interest rate:

$$K = E \times \left[\frac{1 - (1 + r)^{-n}}{r} \right]. \quad (3)$$

Resource use

Resource use was collected at the second- and third-year follow-ups. To account for resource use over the entirety of the trial period, resource use questionnaires at the second-year follow-up asked parents to recall health and social care resource use from when their child started Primary 5, which relates to the beginning of the study. In the third-year follow-up, resource use questionnaires asked parents to recall their child's resource use from the previous 12 months. The resource use measured was deliberately broad-ranging and included various types of health and social service use, and potential contacts with the police. Resources were valued using UK national unit costs.⁵⁹

Quality-adjusted life-years

The QALY is a generic measure of health that combines life expectancy with health-related quality of life and is defined as a year lived in full health.⁶⁰ QALYs are calculated by weighting length of life by health-related quality of life. In this trial health-related quality of life was measured using the CHU9D, which is a generic preference-based measure specifically designed for use with children to estimate QALYs for economic evaluation of programmes/interventions for young people.⁶¹ It was originally developed for younger children (aged 7–11 years) and scored using weights based on UK adult general population values ($n = 300$).⁴⁶ Since then, an alternative value set has been developed for use with adolescents (aged 11–17) using weights based on Australian adolescent population values ($n = 590$).⁶² For our base-case analysis, the CHU9D was scored using the original UK value set. In this context, QALYs should be interpreted in the same way as the outcome of any population health intervention. ROE QALYs reflect the quality-of-life gains achieved from the intervention's aims to increase social and emotional understanding, empathy, promote prosocial behaviours and decrease aggressive behaviours.

Missing data

Health and resource use costs for children were measured using parental self-report. The health and resource use questionnaire (see *Appendix 4*) was posted home to parents, who were asked to fill it in and return it in a stamped, addressed envelope. Health and resource use data were available for the second- and third-year follow-ups. A descriptive analysis of missing data was first undertaken to identify an appropriate analysis method to deal with the missing data. The missing data analysis follows recommendations set out by Faria *et al.*⁶³ for handling missing data in a cost-effectiveness analysis.

Missing data mechanisms are often categorised using Rubin's framework for missing data.⁶⁴ 'Data missing completely at random' assumes that missing data do not depend on the observed and unobserved data values, and thus that the missing data vary independently of these. If data are missing completely at random, a complete-case analysis is valid. In complete-case analysis, only individuals with complete data at each follow-up point are included in the analysis. This is an inefficient use of the data because any individuals with missing follow-up data are dropped from the analysis.⁶³ Available case analysis makes more efficient use of the data by calculating costs and QALYs by treatment group at each follow-up point. They are then summed by treatment group over the whole time horizon of the study. A limitation is that different samples of costs and QALYs may be used, which can lead to non-comparability and affect the covariance structure.⁶⁵ 'Data missing at random' is a less restrictive assumption, as missing data depend only on the observed data and not the unobserved missing data. Multiple imputation is an appropriate analysis strategy for dealing with missing at random data. Data are not missing at random when the probability that data are missing depends on unobserved values.

A descriptive analysis of the proportion of missing values by group and in total was undertaken, along with the range, mean and SD of the observed data. Patterns of missing data were explored using the Stata 'misspattern' command. Logistic regression was performed to explore if baseline covariates were associated with the probability of data being missing. A dummy variable indicating missing data was created for overall costs and QALYs. Logistic regression was conducted with baseline covariates including gender, year group, multiple deprivation and number of siblings. A significant association between a baseline covariate and missing data indicates that data are not missing completely at random.⁶³ Dummy variables were created for costs and QALYs at each time point to explore the association between missing data and observed outcomes.

Each indicator variable was then regressed on all other costs and QALYs observed in each year (i.e. missing baseline QALYs were regressed on costs and QALYs in each subsequent follow-up). Data were then assumed to be missing at random, in which case multiple imputation is an appropriate method of analysis for missing at random data.

Multiple imputation

Multiple imputation was employed using chained equations to handle missing cost and QALY data. Costs were imputed at the total cost level and QALYs were imputed at the index score level for each time point. Owing to advances in computational feasibility, a rule of thumb has been proposed that 'the number of imputations should be similar to the percentage of cases that are incomplete'.⁵³ Missing data on resource use costs were particularly high, so 75 imputations ($m = 75$) were performed. Predictive mean matching was used for continuous, restricted range, and skewed cost and QALY variables. Predictive mean matching is useful as it avoids predictions that lie outside the bounds of each variable,⁵³ however, it can produce predictions that closely match observed values. The uncertainty in these values is incorporated into the mean costs and QALY estimates using Rubin's rules.

Multiple imputation was implemented separately by group allocation (ROE intervention and control), as is recommended to be good practice.⁶³ Covariates included in the imputation model were the same as those used during the estimation step and included gender, year in school, intervention allocation, number of siblings, school, trust and deprivation level. After imputation, three passive variables were created to allow total costs, total QALYs and QALY decrements to be classified as imputed variables to be analysed during the estimation stage. The total cost and QALYs variables generated were the sum of the imputed costs and QALYs at each time point. The QALY decrement was defined as the maximum QALYs that could possibly be accrued within the time frame minus the actual QALYs gained.

Analysis

Regression methods were used to estimate the incremental difference in cost and QALYs while simultaneously adjusting for baseline characteristics that were the same covariates used in the imputation model. Generalised linear models were selected because of their advantage over ordinal least squares and log models, in that they model both mean and variance functions on the original scale of cost.⁶⁶ They also take into account the typically skewed nature of cost and QALY data.⁶⁷ As cost data are typically right-skewed, a right-skewed gamma distribution is appropriate. As QALYs are typically left-skewed, the QALY decrement (described in *Quality-adjusted life-years*) was analysed with a gamma distribution. Thus, both costs and QALYs were analysed with a generalised linear model specifying a gamma family and identity link. Cost and QALY decrements were adjusted for the following covariates: gender, year in school, intervention allocation, number of siblings, school, trust and deprivation level. Baseline health-related quality of life was also included to adjust for any imbalance of health-related quality of life between groups.⁶⁸

The mean costs and QALYs for each group were presented using the method of recycled predictions.⁶⁶ Incremental costs and QALYs, along with their corresponding robust standard errors, were reported from results of the generalised linear model. The ICER was estimated, and uncertainty surrounding the estimate of incremental costs, QALYs and ICERs was investigated by use of a non-parametric bootstrap of the cost and effect pairs for 1000 iterations. This uncertainty was then presented on the cost-effectiveness plane and a 95% CI of the bootstrapped ICER was calculated. The results were summarised using a cost-effectiveness acceptability curve to reflect the probability of ROE being cost-effective at various willingness-to-pay thresholds. The thresholds varied from £0 to £50,000 per QALY, reflecting the range generally accepted to be considered cost-effective by NICE (£20,000–30,000 per QALY gained).

Clustering within economic evaluation

Roots of Empathy was a cluster randomised controlled trial, and so randomisation took place at the cluster (school) level rather than at the individual level. Therefore, it is important to take the effects of clustering into account in the economic analysis.⁵⁴ Cluster randomisation tends to reduce statistical power and precision⁶⁹ because, in the case of ROE, individual pupils from the same school will be more similar than

pupils from different schools. This non-independence is referred to as the intraclass correlation coefficient (ICC).⁷⁰ The ICC could be thought of as the proportion of variance due to between-cluster variation, or the correlation between members of the same cluster.⁵⁴ For sample size calculation, an ICC of 0.05 was assumed.⁷¹

Clustering was accounted for by use of a multilevel model⁷² and the true ICC was estimated. It was expected that a multilevel model may not actually be the best-fitting model for this analysis (owing to having collected cost at only two time points). With this in mind, the ICC was examined to determine if clustering had a design effect on the economic outcomes. If the ICC was < 0.01, then a more practical approach to reflect clustering would be employed by reporting robust standard errors⁷³ for the generalised linear model regressions.

Sensitivity analysis

Extensive sensitivity analyses were undertaken to allow for, explore and assess the uncertainty around the cost-effectiveness results in this economic evaluation. Thorough exploration through sensitivity analysis strengthens the external validity and generalisability of our results. All sensitivity analyses were derived from the base-case analysis described above and a description of each variation is provided in *Table 5*. To answer the main research question of the trial, outcomes were varied by conducting a cost-effectiveness analysis on the primary outcome, the SDQ. The SDQ has two components: the total difficulties score (which comprises difficult and aggressive behaviour) and the prosocial behaviour score. A cost-effectiveness analysis was conducted on both. For the cost-effectiveness analyses, differences in effect were measured as the difference in scores from year 3 to baseline by group (*Table 6*).

The cost of the intervention was a main cost driver, so annuitisation assumptions about the useful life of the intervention were varied to account for no annuitisation and annuitisation over a shorter useful life of 3 years versus 5 years. The discount rate was also varied to reflect a more traditional rate of 3.5% versus the 1.5% public health rate. The level of missing resource use and health-related quality-of-life data from the trial was particularly high, and so a sensitivity analysis was conducted to explore the uncertainty surrounding the missing at random assumption and use of multiple imputation. An available case analysis was conducted, assuming that data were missing completely at random, to assess the impact that multiple imputation had on the incremental costs and QALYs. The main analysis is referred to henceforth as the base-case analysis.

TABLE 5 List of sensitivity analyses

Sensitivity analysis	Element	Description of variation
0	Base case	Multivariate analysis of cost and QALY public sector perspective, 1.5% discount rate, child health utility, MAR assumption and multiple imputation
1	Outcomes	SDQ total difficulties (CEA)
2		SDQ prosocial behaviour (CEA)
3		CHU9D computed with alternative tariff
4	Costs	Training and material costs not annuitised
5		Training and material costs annuitised over 3 years
6	Discount rate	Use of more traditional 3.5% discount rate for costs and outcomes
7	Missing data	Available case analysis assuming MCAR

CEA, cost-effectiveness analysis; MAR, missing at random; MCAR, missing completely at random.

TABLE 6 The ICER for cost-effectiveness analyses on SDQ

Group	Total cost (mean)	Baseline score	Score at final follow-up (mean)	Difference in score	ICER
ROE	a	c	e	(e – c)	$(a - b) / [(e - c) - (f - d)]$
Control	b	d	f	(f – d)	
Difference	(a – b)				

Stakeholder engagement

Stakeholder engagement in this evaluation – particularly in relation to policy-makers, commissioners and programme providers, as well as teachers and pupils – has been a key element throughout this study. The purpose of such engagement has been to:

1. inform aspects of study design (data collection processes, procedures and dissemination)
2. raise stakeholders' awareness, support their involvement and develop their capacity to coproduce research into the practice of school
3. encourage active involvement in the interpretation of the trial findings and process evaluation
4. help identify the practical significance of the findings from the trial and implications for the further delivery of the ROE programme.
5. help plan a dissemination strategy, including a national dissemination seminar in Belfast.

This engagement has taken five forms, as described in the following sections.

Partnership meetings

From the outset of the trial, the research team has attended and fully engaged with the partnership meetings with staff from the health trusts, staff from education library boards and principals from schools participating in the trial. Each meeting has been attended by the lead health programme co-ordinator from the health trust, the pupil personal development officer for the education and library board and usually 6–8 principals from intervention schools. This forum has maintained a schedule of twice-yearly meetings throughout the trial (5 years). The research staff established links at the start and then continued to have consultations with the forum attendees throughout the research period. The aim has been to help influence the research at an early stage of development, to raise awareness of the research and to support the schools' involvement in the process.

Stakeholder members of the Trial Steering Committee

Alongside the above partnership meetings, key stakeholders, comprising staff from the health trusts and staff from education and library boards, as well as the Public Health Agency, have also contributed directly to the evaluation as members of the Trial Steering Committee. Critically, this has included contributing to the emerging interpretation of the findings and the development of the dissemination strategy.

Process evaluation

The research team has given particular prominence to the process evaluation element of the study. As described above, this has involved in-depth engagement with all of the key stakeholders to ascertain their experiences and perspectives of the programme. The rich qualitative insights gained from the teachers, ROE instructors, parents and pupils are demonstrated clearly in *Chapter 4* of this report.

End-of-project consultation meetings

Towards the end of the trial (May–June 2016), discussion groups were conducted in three of the intervention schools to provide preliminary feedback on the findings of the study in order to ensure stakeholder engagement in the interpretation and dissemination of the core findings. The consultations

involved nine interviews, and, in addition, a focus group was held with pupils. Overall, as part of this phase, the research team talked with three principals, two teachers, two ROE instructors, two parents and a group of eight pupils. A purposive quota sample of three schools was recruited, which, within the available budget and time frame, represented the different subsectors of the diverse primary school population in Northern Ireland. The sample was chosen to include schools of different sizes, schools from rural as well as urban locations and schools from three of the education and library boards. These key stakeholders continue to be consulted regarding the development of a regional dissemination strategy, including a national dissemination seminar in Belfast.

Dissemination events

A regional launch of the findings of this evaluation took place in September 2016. This attracted over 100 attendees representing participating schools and a range of voluntary and statutory organisations in the region. This event was planned with the Public Health Agency and in ongoing consultation with members from the Partnership Meetings and those recently engaged through the end of project consultation meetings.

A further regional event, aimed at schools, was held in November 2017, by the Centre for Evidence and Social Innovation at Queen's University Belfast. This event was also co-organised with the Public Health Agency and the Department of Education and included presentations on the findings of this evaluation and another evaluation of an SEL programme undertaken by the Centre for Evidence and Social Innovation. Further regional events are planned for 2018.

In addition, the PHA has funded members of the present research team to undertake a broader systematic review of school-based universal SEL programmes for children aged 3–11 years who have been registered with the Campbell Collaboration.⁷⁴ The findings of this present evaluation, together with the forthcoming findings of the systematic review, will be used to inform future policy and practice with regard to the promotion of children's social and emotional development in Northern Ireland.

Chapter 3 Results from the trial and cost-effectiveness analysis

Introduction

This chapter presents the findings from the trial element of the study and its associated cost-effectiveness analysis. It begins by describing the sample and comparing the differences between the intervention and control groups pre test. It then sets out the findings in relation to the impact of the programme immediately post test and then at each of the follow-up time points. It concludes with an outline of the findings of the cost-effectiveness analysis.

Participant flow

Before pre testing, seven schools withdrew from the trial: four from the intervention group and three from the control group. Sixty-seven schools participated pre test (October 2011). Two further schools withdrew post test, one each from the intervention and control groups, and, thus, 65 schools participated in the post-test data collection (June 2012). At the 12-month follow-up (June 2013), one further school (from the control group) did not take part in data collection; however, this school did not permanently withdraw from the study and will take part in the subsequent data sweeps. No other schools withdrew from the study. The flow of individual children and parents through the study is described in detail in *Figure 2*.

Recruitment

Seventy-four primary schools (clusters), from four of the five trust areas in Northern Ireland, were originally recruited to and enrolled in the trial (Belfast HSCT, South Eastern HSCT, Southern HSCT and Western HSCT) by health and social care trust personnel between March and June 2011.

In total, 1278 pupils aged between 8 and 9 years were recruited to the study: 695 in the intervention group and 583 in the control group. It can be seen from *Table 7* that the proportions of controlled, Catholic-maintained and integrated primary schools recruited to the sample were broadly representative of the population of Northern Ireland primary schools as a whole.

Table 8 shows the breakdown of the sample by gender, health and social care trust, geographic area (urban or rural) and primary school type (controlled, Catholic maintained, integrated or other).

TABLE 7 Schools in the sample, by type, compared with the Northern Ireland population in 2011/12

School type	Population, ^a n (%)	Sample, n (%)
Controlled	378 (44)	27 (40)
Catholic maintained	392 (46)	31 (46)
Integrated	42 (5)	6 (9)
Other	42 (5)	3 (5)
Total	854 (100)	67 (100)

^a Source: www.deni.gov.uk.

TABLE 8 Sample characteristics

Characteristic	Group, n (%)		Total, n (%)
	Control	Intervention	
Gender			
Male	310 (53.2)	347 (49.9)	657 (51.4)
Female	273 (46.8)	348 (50.1)	621 (48.6)
Class			
P4	43 (7.4)	38 (5.5)	81 (6.3)
P5	528 (90.6)	611 (87.9)	1139 (89.1)
P6	12 (2.1)	4 (6.6)	58 (4.5)
Trust			
Belfast	145 (24.9)	201 (28.9)	346 (27.1)
South Eastern	150 (25.7)	222 (31.9)	372 (29.1)
Southern	181 (31.0)	171 (24.6)	352 (27.5)
Western	107 (18.4)	101 (14.5)	208 (16.3)
Area			
Urban	330 (56.6)	363 (52.2)	693 (54.2)
Rural	253 (43.4)	332 (47.8)	585 (45.8)
School type			
Controlled	189 (32.4)	242 (34.8)	431 (33.7)
Catholic maintained	286 (49.1)	360 (51.8)	646 (50.6)
Integrated	85 (14.6)	77 (11.1)	162 (12.7)
Other	23 (3.9)	16 (2.3)	39 (3.1)
Total	583 (100.0)	695 (100.0)	1278 (100)

P, primary.

Baseline data

Table 9 compares the intervention and control groups in terms of the demographic characteristics of child gender, parental qualification and familial socioeconomic status (NIMDM 2010 ranking) pre test (T0). As can be seen, there were no notable differences between the groups in terms of these core characteristics.

Table 10 compares the intervention and control groups in terms of their scores pre test on the primary and secondary outcome measures. As can be seen, there were no notable differences between the groups for either the primary or the secondary outcome measures, apart from teacher-rated total difficulties, which were rated as higher in the intervention group. This, in turn, confirms that the randomisation process appears to have worked in producing two balanced groups. Any pre-test differences described above were taken into account and statistically controlled for during the analysis.

Finally, it emerged that there were pre-test differences between those parents who did and those who did not complete a questionnaire for their child immediately post test (T1). More specifically, parents who completed a pre-test questionnaire but did not return a post-test questionnaire had lower levels of educational attainment and were from areas of higher deprivation than those parents who returned a post-test questionnaire (Table 11 provides details). Furthermore, as detailed in Table 12, compared with

TABLE 9 Comparison between control and intervention groups on demographic characteristics

Characteristic	Group, n (%)	
	Control	Intervention
Gender		
Male	310 (53.2)	347 (49.9)
Female	273 (46.8)	348 (50.1)
Highest maternal qualification		
Below third level	202 (67.1)	246 (67.8)
Third level	99 (32.9)	117 (32.2)
Highest paternal qualification		
Below third level	179 (74.3)	227 (73.7)
Third level	62 (25.7)	81 (26.3)
Deprivation rank, ^a mean (SD)	380 (235)	444 (252)

^a Lower rank indicates a higher level of deprivation.

TABLE 10 Comparison between control and intervention groups on primary and secondary outcome measures pre test

Outcome	Group, mean (SD)	
	Control	Intervention
Primary		
Teacher-rated SDQ subscale		
Prosocial behaviour	1.59 (0.45)	1.58 (0.46)
Total difficulties	0.29 (0.29)	0.36 (0.34)
Measures associated with primary outcomes (for purposes of triangulation)		
Teacher-rated CBS		
Aggressive subscale	0.24 (0.43)	0.25 (0.45)
Prosocial subscale	1.59 (0.42)	1.59 (0.44)
Parent-rated SDQ subscale		
Prosocial behaviour	1.74 (0.34)	1.72 (0.31)
Total difficulties	0.36 (0.31)	0.40 (0.30)
Secondary		
Child understanding of infant crying		
Number of reasons an infant cries	3.35 (1.77)	3.52 (2.05)
Ways to help a baby who is crying	3.06 (1.65)	3.28 (1.76)
Recognition of emotions	0.83 (0.13)	0.82 (0.14)
Empathy	3.34 (0.78)	3.32 (0.75)
Child emotional regulation	0.97 (0.41)	0.99 (0.40)
Experience of being bullied at school	1.60 (0.76)	1.68 (0.74)
Quality of life (CHU9D)	0.84 (0.11)	0.84 (0.12)

TABLE 11 Comparison of demographic characteristics between those who completed pre-test questionnaire only and those who completed pre- and post-test questionnaires

Characteristic	Pre-test data only	Pre- and post-test data	Significance ^a
Gender, <i>n</i> (%)			
Male	60 (58.8)	597 (50.8)	$p = 0.08$
Female	42 (41.2)	579 (49.2)	
Highest maternal qualification, <i>n</i> (%)			
Below third level	130 (77.8)	318 (63.9)	$p < 0.01$
Third level	37 (22.2)	179 (36.1)	
Highest paternal qualification, <i>n</i> (%)			
Below third level	96 (83.5)	309 (71.3)	$p = 0.01$
Third level	19 (16.5)	124 (28.7)	
Deprivation rank, ^b mean (SD); <i>n</i>	349 (251); 605	475 (225); 649	$p < 0.01$

a Tested using linear multilevel models for continuous variables and binary logistic multilevel models for categorical variables, taking account of clustering at the school level.

b Lower rank indicates a higher level of deprivation.

TABLE 12 Comparison of outcome measures pre test between those who completed pre-test questionnaire only and those who completed pre- and post-test questionnaires

Outcome	Pre-test data only, mean (SD); <i>n</i>	Pre- and post-test data, mean (SD); <i>n</i>	Significance ^a
Primary			
Teacher-rated SDQ subscale			
Prosocial behaviour	1.39 (0.47); 75	1.60 (0.45); 1041	$p < 0.01$
Total difficulties	0.45 (0.37); 75	0.32 (0.32); 1041	$p < 0.01$
Teacher-rated CBS			
Aggressive subscale	0.40 (0.55); 77	0.23 (0.43); 1036	$p = 0.09$
Prosocial subscale	1.41 (0.44); 77	1.60 (0.42); 1036	$p = 0.01$
Parent-rated SDQ subscale			
Prosocial behaviour	1.69 (0.33); 176	1.74 (0.32); 510	$p = 0.09$
Total difficulties	0.47 (0.34); 176	0.35 (0.28); 510	$p < 0.01$
Secondary			
Child understanding of infant crying			
Number of reasons an infant cries	3.30 (1.87); 88	3.46 (1.94); 1081	$p = 0.01$
Ways to help a baby who is crying	3.35 (2.09); 88	3.17 (1.68); 1084	$p = 0.10$
Recognition of emotions	0.80 (0.16); 88	0.83 (0.13); 1092	$p = 0.05$
Empathy	3.46 (0.88); 89	3.32 (0.75); 1093	$p = 0.67$
Child emotional regulation	1.01 (0.40); 89	0.98 (0.41); 1091	$p = 0.93$
Experience of being bullied at school	1.64 (0.80); 90	1.65 (0.74); 1090	$p = 0.85$
Quality of life (CHU9D)	0.84 (0.13); 89	0.84 (0.11); 1091	$p = 0.92$

a Tested using multilevel models to account for clustering.

the children for whom there were both pre- and post-test data, those children for whom no post-test data were returned had:

- poorer prosocial behaviour, as reported by both teachers and parents (SDQ and CBS)
- greater difficulties, as reported by both teachers and parents (SDQ)
- higher levels of aggression, as reported by teachers only (CBS).

Outcomes and estimation

The two primary outcomes for this study were measured using the teacher-rated SDQ (prosocial behaviour and total difficulties) immediately post test. Data from the parent-rated SDQ and the teacher-rated CBS were used to provide triangulation and support for the reliability and validity of the teacher-rated SDQ. As can be seen from *Table 13*, after controlling for pre-test scores, children who participated in the ROE programme were rated by their teachers as *more* prosocial (effect size, $g = +0.20$; $p = 0.05$) and as exhibiting *less* difficult behaviour ($g = -0.16$; $p = 0.06$) than those in the control group. As the effect sizes indicate, the degree of the differences between the groups for both outcomes can be considered small, with only one approaching statistical significance.

TABLE 13 Summary of main effects immediately post test (T1)

Outcome	Group		Effect size (95% CI)	Significance ^a	ICC		
	Control	Intervention					
	Adjusted post-test mean (SD)	<i>n</i>	Adjusted post-test mean (SD)	<i>n</i>			
Primary							
Teacher							
Prosocial behaviour (SDQ)	-0.121 (0.979)	415	0.047 (1.018)	538	0.199 (0.005 to 0.394)	0.045	0.217
Difficult behaviour (SDQ)	0.098 (0.896)	415	-0.063 (1.068)	538	-0.162 (-0.330 to 0.007)	0.060	0.168
Secondary							
Child							
Reasons why a baby cries	-0.130 (0.917)	424	0.105 (1.060)	537	0.235 (0.048 to 0.421)	0.014	0.153
Ways to help a crying baby	-0.096 (0.963)	424	0.075 (1.026)	537	0.171 (-0.012 to 0.354)	0.066	0.158
Emotional recognition	-0.006 (0.979)	424	0.087 (1.017)	537	0.094 (-0.053 to 0.241)	0.211	0.061
Empathy	0.012 (1.018)	424	-0.052 (0.986)	537	-0.064 (-0.216 to 0.088)	0.410	0.061
Emotional regulation	-0.049 (1.015)	424	0.034 (0.988)	537	0.084 (-0.053 to 0.220)	0.229	0.070
Bullied (victim)	-0.083 (1.015)	423	0.039 (0.986)	537	0.122 (-0.036 to 0.280)	0.129	0.077
Quality of life	-0.005 (0.983)	411	0.012 (1.013)	525	0.017 (-0.118 to 0.152)	0.806	0.029

^a Significance of differences in mean scores calculated using multilevel models to take into account the clustered nature of the data.

Evidence of the validity of the teacher-rated SDQ as an outcome measure was provided in two main ways. First, the teacher-rated SDQ was significantly correlated with the scores obtained from other sources (parents) as well as from other measures (the CBS). Correlations between the three prosocial subscales and correlations between the three aggressive/difficulties subscales of the SDQ (parent and teacher rated) and the CBS were all statistically significant ($p < 0.001$) and ranged between 0.17 and 0.83.

Second, and as part of the sensitivity analysis referred to in *Chapter 2*, the findings from this main analysis using the teacher-rated SDQ for both primary outcomes were compared with the findings obtained by repeating the analysis using the alternative parent- and child-rated SDQ measures and also the measures of prosocial and aggressive behaviour gained from the CBS. These comparisons are summarised in *Table 14*. As can be seen, there is broad agreement between the findings across the different measures in relation to prosocial behaviour. In contrast, the level of effect found in the main analysis for the reduction in total difficulties does not appear to have been replicated using the parent-rated measure. It is, however, difficult to draw too many conclusions from this given the lower response rates of the parents and the known biases that this introduced to the parent sample, as described above. Moreover, the teacher-rated measure of aggression using the CBS is not wholly comparable with the total difficulties score of the SDQ and, thus, direct comparisons between the two need to be made with caution.

After immediately post test, and as can also be seen, a fairly consistent pattern was found when comparing the main effects across the other three follow-up time points. Overall, therefore, these comparisons do not raise any notable concerns regarding the reliability of the main measures for the two primary outcomes.

TABLE 14 Comparison of effect sizes for the primary outcomes used in the main analysis (teacher-rated SDQ) with those measured by parent- and child-rated SDQ and teacher-rated CBS

Outcome	Measure, effect size (Hedges' <i>g</i>) (significance level)			
	Teacher-rated SDQ	Parent-rated SDQ	Child-rated SDQ	Teacher-rated CBS
Immediately post test (T1)				
Prosocial	0.199 ($p = 0.045$)	0.185 ($p = 0.004$)	–	0.210 ($p = 0.040$)
Total difficulties (aggressive behaviour: CBS)	–0.162 ($p = 0.060$)	–0.058 ($p = 0.265$)	–	0.006 ($p = 0.942$)
Immediately post test (T2)				
Prosocial	–0.002 ($p = 0.988$)	0.097 ($p = 0.140$)	–	–0.019 ($p = 0.867$)
Total difficulties (aggressive behaviour: CBS)	–0.144 ($p = 0.223$)	–0.012 ($p = 0.867$)	–	–0.065 ($p = 0.522$)
Immediately post test (T3)				
Prosocial	0.048 ($p = 0.726$)	0.159 ($p = 0.044$)	0.034 ($p = 0.662$)	0.062 ($p = 0.665$)
Total difficulties (aggressive behaviour: CBS)	–0.133 ($p = 0.254$)	–0.078 ($p = 0.317$)	–0.070 ($p = 0.357$)	–0.044 ($p = 0.678$)
Immediately post test (T4)				
Prosocial	0.122 ($p = 0.200$)	0.005 ($p = 0.620$)	–0.042 ($p = 0.560$)	0.280 ($p = 0.006$)
Total difficulties (aggressive behaviour: CBS)	–0.142 ($p = 0.142$)	0.063 ($p = 0.438$)	–0.095 ($p = 0.364$)	–0.104 ($p = 0.243$)

Ancillary analyses

Secondary outcomes

With regard to the secondary outcomes, after controlling for pre-test scores, it can be seen from *Table 13* that children who participated in the ROE programme were able to report a greater number of reasons for why babies cry (effect size = +0.24; $p = 0.01$). It is important to note, however, that part of the intervention involves explicitly teaching children about how babies communicate and why they cry (it is one of the nine themes involving three lessons). It is conceivable, therefore, that this measure is treatment-inherent and biased in favour of the intervention group. Furthermore, the effect is small, given that the findings suggest that the children in the intervention group are recalling only less than one reason more than the children in the control group for why babies cry.

No evidence of any differences between the groups was found in relation to the other secondary outcomes. Full details of these models can be found in *Appendix 2*.

Exploratory subgroup analyses

Prespecified subgroup analyses were undertaken to explore whether or not the programme worked better according to:

- the child's gender
- the socioeconomic background of the child's family (measured using NIMDM ranking for the child's home address)
- the number of siblings.

Given the large number of significance tests undertaken in relation to exploring interaction effects for each of the above in relation to all of the primary and secondary outcomes listed, these analyses need to be treated with caution and considered as essentially exploratory in helping to identify potential patterns that may be useful to consider in future research. To provide an overview of the findings of these additional analyses, *Table 15* sets out the statistical significance for each of the interaction terms added to the

TABLE 15 Statistical significance of coefficients for interaction effects added to statistical models for child gender, family socioeconomic status and number of siblings, respectively, immediately post test (T1)^a

Dependent variable	Significance (p -value)		
	Gender	Socioeconomic status	Siblings
Primary outcomes			
Prosocial behaviour (SDQ)	0.255	0.088	0.106
Difficult behaviour (SDQ)	0.407	0.732	0.396
Secondary outcomes			
Reasons why a baby cries	0.254	0.959	0.507
Ways to help a crying baby	0.918	0.822	0.477
Emotional recognition	0.043	0.473	0.786
Empathy	0.935	0.229	0.002
Emotional regulation	0.264	0.078	0.539
Bullying (victim)	0.399	0.810	0.943
Quality of life	0.086	0.934	0.384

^a These models are as specified in *Appendix 2*, with two changes: the z-score for the dummy variable for gender replaced with the original dummy variable; and an interaction effect added for 'group × gender'. The significance levels above represent those of the interaction effects and provide a simple visual summary of whether or not there is evidence of subgroup differences in relation to the child's gender, family socioeconomic status and their number of siblings, and the pattern of such across the primary and secondary outcomes.

original models (as described in *Appendix 2*) for each of the three subgroup variables. This provides an indication of whether or not there is any potential evidence of subgroup differences in the effects of the programme for each of the outcome variables.

As can be seen, only two interaction terms were found to fall below the threshold of $p = 0.05$. Given that there were 27 tests in total, these findings could have occurred by chance. Moreover, and as can also be seen, there is no clear or consistent pattern to these findings to provide at least some suggestion that these may represent underlying differential effects. Therefore, it was decided not to explore the nature of these isolated subgroup differences any further.

In addition, it was found that ROE was delivered with high fidelity in all intervention schools (see *Chapter 4*). As a result, there was insufficient variation in fidelity to assess whether or not fidelity was associated with differences in outcomes achieved.

Further exploratory analysis

In addition to the above prespecified exploratory analysis, the following further analysis was undertaken in response to a number of queries raised by key stakeholders. This further analysis was not prespecified in the original analysis plan and should, therefore, be regarded as simply providing contextual data that may help in the interpretation of the findings from the main analysis. The analysis focused on three queries raised:

1. Is the baseline social and emotional functioning of children in Northern Ireland comparable to that of children in the UK, in Ireland and internationally?
2. Was there a difference in the duration of Personal Development and Mutual Understanding (PDMU) lessons delivered between control and intervention classrooms?
3. Does the programme work better for children who have poor prosocial behaviour to start with?

Is the baseline social and emotional functioning of children in Northern Ireland comparable with that of children in the UK, in Ireland and internationally?

This first query was raised in relation to the possibility that the PDMU element of the Revised Curriculum (which focuses on social and emotional development) might mean that baseline social and emotional functioning (i.e. SDQ prosocial behaviour and total difficulties scores) might be higher for children in Northern Ireland than for children in other countries. The implication of this is that the impact of ROE might not be as great here in Northern Ireland as it has been suggested to be in other countries. *Table 16* shows that the mean prosocial behaviour and total difficulties scores for children in the sample are, in fact, commensurate with those of children of a similar age across the UK, Ireland and the USA.

Was there a difference in the duration of Personal Development and Mutual Understanding being delivered between control and intervention classrooms?

Personal Development and Mutual Understanding is a statutory requirement of the revised curriculum. It focuses on encouraging each child to become personally, emotionally and socially effective, to lead healthy, safe and fulfilled lives and to become confident, independent and responsible citizens, making informed and responsible choices. The possibility was raised that teachers whose classes were receiving the programme might view it as contributing to PDMU and thus reduce the amount of time they spent on PDMU in class. If this is the case, then it is conceivable that ROE might be replacing some of the PDMU time in class rather than being delivered over and above PDMU.

Teachers were asked to estimate how long they spent delivering PDMU to their Primary 5 class every week (excluding ROE lessons for the teachers in the intervention schools). On average, and as summarised in *Table 17*, the intervention teachers reported delivering approximately 25 minutes *less* of PDMU per week than the control teachers ($p = 0.03$, tested using linear multilevel modelling to take account of clustering at the school level).

TABLE 16 Comparison of ROE sample at baseline and post-test SDQ mean scores with UK, Ireland and US national averages

SDQ subscale	ROE sample, mean (SD)					UK ^a	USA ^b	GUI ^c	ALSPAC ^d
	Baseline	Post test	FU1	FU2	FU3				
Teacher rated									
Prosocial	7.9 (2.3)	8.3 (2.1)	8.6 (2.0)	8.4 (2.2)	7.9 (2.3)	7.1 (2.4)	Not available	8.3 (2.1)	Not available
Total difficulties	6.6 (6.5)	6.1 (6.2)	6.0 (6.0)	5.5 (5.9)	5.8 (6.1)	6.3 (6.1)		5.9 (5.9)	
Parent rated									
Prosocial	8.6 (1.6)	8.8 (1.5)	8.8 (1.6)	8.8 (1.6)	8.7 (1.7)	8.6 (1.6)	8.8 (1.7)	8.9 (1.5)	8.2 (1.7)
Total difficulties	7.6 (6.0)	7.2 (5.6)	6.8 (5.8)	6.3 (5.5)	6.4 (6.1)	8.6 (5.7)	7.2 (5.8)	8.0 (5.3)	7.4 (4.8)

ALSPAC, Avon Longitudinal Study of Parents and Children; FU, follow-up; GUI, Growing Up in Ireland.

a UK norms are drawn from a nationally representative survey of child and adolescent mental health carried out by National Statistics and funded by the Department of Health.⁷⁵ The figures represent data for children aged 5–10 years ($n = 5855$).

b US normative data are available only for parent-report, and are drawn from the National Health Interview Survey conducted in 2001.⁷⁶ The figures represent data for children aged 8–10 years ($n = 2064$).

c Data are based on a representative sample of approximately 8500 9-year-old children across Ireland between 2007 and 2008 (the Growing Up in Ireland study).⁷⁷

d SDQ data were available for 7725 children at 7 years of age (parental report only) between 1998 and 1999 in the ALSPAC.⁷⁸

TABLE 17 Duration (minutes) of PDMU delivered in class per week by intervention and control teachers

Group	Minutes		
	Mean (SD)	Minimum	Maximum
Intervention	62.74 (35.44)	0	120
Control	87.42 (40.41)	45	300

Does the programme work better for children who have poor prosocial behaviour to start with?

Finally, the main analysis demonstrated that ROE was effective in improving prosocial behaviour. However, the question was raised of whether or not the programme works better for children with low levels of prosocial behaviour than for those who have high levels of prosocial behaviour. Prosocial behaviour was measured using three ratings, teacher SDQ, parent SDQ and teacher CBS, and there was no evidence that the programme worked any better for children with low prosocial SDQ scores at baseline (teacher or parent rated). There appeared to be a significant interaction between group allocation and prosocial behaviour as rated by the teacher using the CBS. However, given the fact that this was an isolated finding and not also the case for the SDQ scales, the exploratory nature of these analyses and also the increased likelihood of a type I error (false positive) due to multiple testing, this is likely to be a spurious and unreliable finding.

Primary outcomes at 12-, 24- and 36-month follow-up

Tables 18–20 summarise the findings of the analysis of the effects of the ROE programme at 12, 24 and 36 months following the end of the intervention, respectively. As before, full details of each of the models fitted are provided in *Appendix 2*. As can be seen, after controlling for pre-test scores, the potentially positive effect found immediately post test for prosocial behaviour appears to have disappeared after 12 months, with no notable differences between the scores of those in the intervention and control groups at any of the subsequent follow-up time points (at 12, 24 or 36 months post intervention).

TABLE 18 Summary of main effects at 12-month follow up (T2)

Outcome	Group				Effect size (95% CI)	Significance ^a	ICC
	Control		Intervention				
	Adjusted post-test mean (SD)	<i>n</i>	Adjusted post-test mean (SD)	<i>n</i>			
Primary							
Teacher							
Prosocial behaviour (SDQ)	-0.025 (0.979)	355	-0.027 (1.018)	481	-0.002 (-0.214 to 0.210)	0.988	0.134
Difficult behaviour (SDQ)	0.096 (0.896)	355	-0.048 (1.068)	481	-0.144 (-0.377 to 0.088)	0.223	0.160
Secondary							
Child							
Reasons why a baby cries	-0.023 (0.917)	387	0.007 (1.060)	513	0.030 (-0.205 to 0.265)	0.804	0.174
Ways to help a crying baby	-0.008 (0.963)	388	-0.017 (1.026)	508	-0.025 (-0.272 to 0.221)	0.842	0.173
Emotional recognition	0.063 (0.979)	388	0.060 (1.017)	515	-0.002 (-0.137 to 0.133)	0.976	0.014
Empathy	0.015 (1.018)	389	-0.030 (0.986)	515	-0.045 (-0.239 to 0.149)	0.650	0.074
Emotional regulation	-0.054 (1.015)	389	-0.036 (0.988)	515	0.018 (-0.158 to 0.195)	0.841	0.060
Bullying (victim)	0.003 (1.015)	388	0.007 (0.986)	515	0.004 (-0.146 to 0.155)	0.963	0.046
Quality of life	-0.098 (0.983)	385	0.088 (1.013)	507	0.185 (0.049 to 0.322)	0.008	0.032
a Significance of differences in mean scores calculated using multilevel models to take into account the clustered nature of the data.							

However, and in relation to total difficulties (as measured by the teacher-rated SDQ), the effect size immediately post test appears to have been maintained at the 12-month ($g = -0.14$), 24-month ($g = -0.13$) and 36-month ($g = -0.14$) follow-up time points. However, and because of the reduction in sample size due to attrition, this effect is not statistically significant and so needs to be treated with a degree of caution.

Secondary outcomes and exploratory analyses

As is also evident from these tables, no evidence of any differences between the intervention and control groups was found in relation to the other secondary outcomes at any of the follow-up time points (with the exception of quality of life at 12-month follow-up).

TABLE 19 Summary of main effects at 24-month follow-up (T3)

Outcome	Group				Effect size (95% CI)	Significance ^a	ICC
	Control		Intervention				
	Adjusted post-test mean (SD)	n	Adjusted post-test mean (SD)	n			
Primary							
Teacher							
Prosocial behaviour (SDQ)	-0.101 (0.979)	360	-0.052 (1.018)	488	0.048 (-0.222 to 0.318)	0.726	0.228
Difficult behaviour (SDQ)	0.119 (0.896)	360	-0.013 (1.068)	488	-0.133 (-0.360 to 0.095)	0.254	0.155
Secondary							
Child							
Reasons why a baby cries	-0.100 (0.917)	402	0.056 (1.060)	505	0.156 (-0.045 to 0.356)	0.128	0.130
Ways to help a crying baby	-0.050 (0.963)	402	0.054 (1.026)	504	0.104 (-0.100 to 0.307)	0.319	0.132
Emotional recognition	0.043 (0.979)	402	0.006 (1.017)	505	-0.036 (-0.177 to 0.105)	0.614	0.034
Empathy	0.056 (1.018)	401	0.026 (0.986)	505	-0.030 (-0.190 to 0.130)	0.713	0.085
Emotional regulation	-0.065 (1.015)	402	0.018 (0.988)	505	0.084 (-0.068 to 0.235)	0.280	0.074
Bullying (victim)	0.063 (1.015)	401	0.015 (0.986)	505	-0.047 (-0.182 to 0.088)	0.494	0.031
Bullying (bully)	0.038 (1.009)	397	0.033 (0.993)	492	-0.005 (-0.143 to 0.132)	0.941	0.029
Quality of life	-0.083 (0.983)	397	0.017 (1.013)	492	0.099 (-0.032 to 0.231)	0.139	0.017

^a Significance of differences in mean scores calculated using multilevel models to take into account the clustered nature of the data.

In addition, *Tables 21–23* summarise the evidence of any subgroup differences in relation to the effects of ROE with regard to the three prespecified comparisons for child gender, family socioeconomic status and number of siblings. As can be seen, the picture is essentially similar to that immediately post test, with very few interaction terms proving to be statistically significant. Moreover, of the terms that are significant, there is no clear or convincing pattern. Therefore, it can be concluded that there remains insufficient evidence of any subgroup differences in the effectiveness of the ROE programme at each stage of follow-up.

TABLE 20 Summary of main effects at 36-month follow-up (T4)

Outcome	Group		Adjusted post-test mean (SD)	n	Adjusted post-test mean (SD)	n	Effect size (95% CI)	Significance ^a	ICC
	Control	Intervention							
Primary									
Teacher									
Prosocial behaviour (SDQ)	-0.101 (0.979)	318	0.021 (1.018)	405	0.122 (-0.064 to 0.308)	0.200	0.116		
Difficult behaviour (SDQ)	0.123 (0.896)	318	-0.019 (1.068)	407	-0.142 (-0.332 to 0.048)	0.142	0.098		
Secondary									
Child									
Reasons why a baby cries	-0.059 (0.917)	327	-0.026 (1.060)	409	0.033 (-0.115 to 0.181)	0.661	0.031		
Ways to help a crying baby	0.000 (0.963)	327	-0.053 (1.026)	409	-0.053 (-0.211 to 0.104)	0.506	0.041		
Emotional recognition	0.014 (0.979)	326	-0.065 (1.017)	409	-0.078 (-0.246 to 0.089)	0.360	0.033		
Empathy	0.048 (1.018)	327	-0.012 (0.986)	411	-0.059 (-0.200 to 0.082)	0.410	0.000		
Emotional regulation	-0.030 (1.015)	326	-0.007 (0.988)	412	0.023 (-0.120 to 0.166)	0.755	0.056		
Bullying (victim)	0.052 (1.015)	326	0.024 (0.986)	406	-0.028 (-0.178 to 0.122)	0.713	0.004		
Bullying (bully)	0.037 (0.983)	322	0.032 (1.113)	406	-0.004 (-0.158 to 0.149)	0.956	0.000		
Quality of life	-0.109 (0.983)	322	0.002 (1.013)	406	0.111 (-0.052 to 0.274)	0.182	0.043		

^a Significance of differences in mean scores calculated using multilevel models to take into account the clustered nature of the data.

Sensitivity analysis for missing data

As noted earlier and summarised in the flow diagram for the study (see *Figure 2*), the trial has experienced some attrition of pupils and schools since baseline pre-testing. To test whether or not this attrition has introduced any bias to the findings, the main statistical models used to estimate the effects of the programme for the primary and secondary outcomes, across all time points (immediately post test to 36-month follow-up), were rerun with imputed data.

Multiple imputation was employed using chained equations, using Stata version 14.1. Imputed data sets were created using all of the outcome variables at all five time points (pre test through to 36-month follow-up), multiple deprivation scores and the fully observed variables for gender and trust location.

TABLE 21 Statistical significance of coefficients for interaction effects added to statistical models for child gender, family socioeconomic status and number of siblings, respectively, at 12-month follow-up (T2)^a

Dependent variable	Significance (<i>p</i> -value)		
	Gender	Socioeconomic status	Siblings
Primary outcomes			
Prosocial behaviour (SDQ)	0.672	0.514	0.003
Difficult behaviour (SDQ)	0.976	0.841	0.287
Secondary outcomes			
Reasons why a baby cries	0.862	0.502	0.164
Ways to help a crying baby	0.985	0.102	0.086
Emotional recognition	0.239	0.421	0.639
Empathy	0.342	0.730	0.178
Emotional regulation	0.735	0.346	0.719
Bullying (victim)	0.505	0.410	0.453
Quality of life	0.264	0.631	0.741

a These models are as specified in *Appendix 2*, with two changes: the z-score for the dummy variable for gender replaced with the original dummy variable; and an interaction effect added for 'group*gender'. The significance levels above represent those of the interaction effects and provide a simple visual summary of whether or not there is evidence of subgroup differences in relation to the child's gender, family socioeconomic status and number of siblings, and the pattern of such across the primary and secondary outcomes.

TABLE 22 Statistical significance of coefficients for interaction effects added to statistical models for child gender, family socioeconomic status and number of siblings, respectively, at 24-month follow-up (T3)^a

Dependent variable	Significance (<i>p</i> -value)		
	Gender	Socioeconomic status	Siblings
Primary outcomes			
Prosocial behaviour (SDQ)	0.384	0.429	0.807
Difficult behaviour (SDQ)	0.810	0.070	0.654
Secondary outcomes			
Reasons why a baby cries	0.247	0.221	0.853
Ways to help a crying baby	0.124	0.239	0.692
Emotional recognition	0.527	0.891	0.287
Empathy	0.802	0.760	0.083
Emotional regulation	0.413	0.403	0.694
Bullying (victim)	0.975	0.003	0.103
Bullying (bully)	0.724	0.006	0.553
Quality of life	0.775	0.045	0.722

a These models are as specified in *Appendix 2*, with two changes: the z-score for the dummy variable for gender replaced with the original dummy variable; and an interaction effect added for 'group*gender'. The significance levels above represent those of the interaction effects and provide a simple visual summary of whether or not there is evidence of subgroup differences in relation to the child's gender, family socioeconomic status and number of siblings, and the pattern of such across the primary and secondary outcomes.

TABLE 23 Statistical significance of coefficients for interaction effects added to statistical models for gender, socioeconomic status and number of siblings, respectively, at 36-month follow-up (T4)^a

Dependent variable	Significance (<i>p</i> -value)		
	Gender	Socioeconomic status	Siblings
Primary outcomes			
Prosocial behaviour (SDQ)	0.739	0.231	0.035
Difficult behaviour (SDQ)	0.677	0.078	0.925
Secondary outcomes			
Reasons why a baby cries	0.854	0.952	0.689
Ways to help a crying baby	0.166	0.547	0.901
Emotional recognition	0.902	0.721	0.398
Empathy	0.879	0.457	0.045
Emotional regulation	0.900	0.458	0.401
Bullying (victim)	0.235	0.746	0.975
Bullying (bully)	0.547	0.082	0.356
Quality of life	0.272	0.316	0.492

a These models are as specified in *Appendix 2*, with two changes: the zscore for the dummy variable for gender replaced with the original dummy variable; and an interaction effect added for 'Group*Gender'. The significance levels above represent those of the interaction effects and provide a simple visual summary of whether or not there is evidence of subgroup differences in relation to the child's gender, family socioeconomic status and their number of siblings, and the pattern of such across the primary and secondary outcomes.

Twenty imputations ($m = 20$) were performed for the purposes of the analysis. Multiple imputation was also performed separately by allocation (intervention and control groups).

Comparisons of the effects reported above from the main analysis at each time point with those estimated using multiply imputed data are provided in *Table 24*. As can be seen, the findings using the imputed data sets are broadly similar to those using just the observed data.

Cost-effectiveness analysis

Missing data

In total, 38% of resource use questionnaires were returned for the second-year follow-up and 29% were returned at the final third-year follow-up (*Tables 25 and 26*). Some variables did not provide any data as nearly 100% of the data were missing. Such variables were subsequently dropped from the analysis; these included days off work because a child was home from school, other resource use and medications.

Missing data followed a non-monotonic pattern (*Figure 3*) because data may be missing for an individual in one follow-up but then return in subsequent follow-ups. Here the missing completely at random assumption would be inefficient because data from subsequent follow-ups would not be utilised and all non-complete cases would be dropped.

Logistic regression

Deprivation level and number of siblings at baseline were found to be significant predictors of missing cost. Gender, deprivation level and number of siblings were all significant predictors of missing QALYs that

TABLE 24 Comparison of main effects estimated with the observed data only with the effects estimated using data sets with multiple imputation

Outcomes	Time point, SMD (statistical significance)							
	Immediately post test (T1)		12-month follow-up (T2)		24-month follow-up (T3)		36-month follow-up (T4)	
	Observed data set	Imputed data set	Observed data set	Imputed data set	Observed data set	Imputed data set	Observed data set	Imputed data set
Primary outcomes								
Prosocial behaviour (SDQ)	0.199 ($p = 0.045$)	0.203 ($p = 0.023$)	-0.002 ($p = 0.988$)	-0.019 ($p = 0.831$)	0.048 ($p = 0.726$)	0.006 ($p = 0.957$)	0.122 ($p = 0.200$)	0.142 ($p = 0.158$)
Difficult behaviour (S4DQ)	-0.162 ($p = 0.060$)	-0.151 ($p = 0.040$)	-0.144 ($p = 0.223$)	-0.107 ($p = 0.301$)	-0.133 ($p = 0.254$)	-0.142 ($p = 0.171$)	-0.142 ($p = 0.142$)	-0.189 ($p = 0.046$)
Secondary outcomes								
Reasons why a baby cries	0.235 ($p = 0.014$)	0.219 ($p = 0.011$)	0.030 ($p = 0.804$)	0.116 ($p = 0.246$)	0.156 ($p = 0.128$)	0.157 ($p = 0.083$)	0.033 ($p = 0.661$)	-0.010 ($p = 0.894$)
Ways to help a crying baby	0.171 ($p = 0.066$)	0.150 ($p = 0.094$)	-0.025 ($p = 0.842$)	-0.066 ($p = 0.510$)	0.104 ($p = 0.319$)	0.134 ($p = 0.136$)	-0.053 ($p = 0.506$)	-0.050 ($p = 0.490$)
Emotional recognition	0.094 ($p = 0.211$)	0.127 ($p = 0.080$)	-0.002 ($p = 0.976$)	0.025 ($p = 0.705$)	-0.036 ($p = 0.614$)	0.003 ($p = 0.962$)	-0.078 ($p = 0.360$)	-0.072 ($p = 0.417$)
Empathy	-0.064 ($p = 0.410$)	-0.022 ($p = 0.758$)	-0.045 ($p = 0.650$)	-0.052 ($p = 0.570$)	-0.030 ($p = 0.713$)	0.011 ($p = 0.898$)	-0.059 ($p = 0.410$)	0.014 ($p = 0.844$)
Emotional regulation	0.084 ($p = 0.229$)	0.113 ($p = 0.102$)	0.018 ($p = 0.841$)	0.055 ($p = 0.498$)	0.084 ($p = 0.280$)	0.136 ($p = 0.077$)	0.023 ($p = 0.755$)	-0.015 ($p = 0.835$)
Bullying (victim)	0.122 ($p = 0.129$)	0.086 ($p = 0.254$)	0.004 ($p = 0.963$)	0.032 ($p = 0.642$)	-0.047 ($p = 0.494$)	-0.037 ($p = 0.566$)	-0.028 ($p = 0.713$)	0.101 ($p = 0.186$)
Quality of life	-0.017 ($p = 0.806$)	-0.001 ($p = 0.989$)	0.185 ($p = 0.108$)	0.184 ($p = 0.010$)	0.099 ($p = 0.139$)	0.086 ($p = 0.186$)	0.111 ($p = 0.182$)	0.047 ($p = 0.556$)

TABLE 25 Variable descriptions and missing data percentages

Variable	Description (total, n = 1254; ROE, n = 672; control, n = 582)	Missing values, %			Range	Mean	SD
		Total	ROE	Control			
Baseline variables							
Gender	Male or female	0	0	0	0, 1	51.45% male	
Year Group	Year in school at trial entry	0	0	0	4, 5, 6	89% P5	
MD-rank	NIMDM	2	3	0	1–889	414.13	245.9
Siblings_PT0	Number of siblings at baseline	1	1	0	0–7	1.01	1.26
Outcome variables for health-related quality of life							
utility0	CHU9D pre test	13	10	16	0.3261–1	0.84	0.12
utility1	CHU9D post test	12	11	13	0.3261–1	0.85	0.11
utility2	CHU9D at 1-year follow-up	14	12	16	0.4582–1	0.84	0.1
utility3	CHU9D at 2-year follow-up	14	15	13	0.3261–1	0.85	0.1
utility4	CHU9D at 3-year follow-up	31	31	31	0.3929–1	0.87	0.1
Outcomes for cost-effectiveness							
total_QALYs	Total QALYs over 3.75 years ^a	45	43	48	1.70–3.61	3.09	0.26
total_costs	Total costs over 3.75 years ^a	76	78	75	£77–10,580	£899.04	£841.93

MD, multiple deprivation; P, primary.

^a Total QALY and costs refers to the sum of QALYs and costs over the 3.75-year trial period discounted at a 1.5% annual rate.

TABLE 26 Outcome variables for cost

Variable	Missing values, %			Range	Mean cost (£) (SD)
	Total	ROE	Control		
Intervention cost	0	0	0	NA	175.22 (NA)
GP_3	62	66	57	0–706	96.07 (102.56)
School Nurse_3	62	66	57	0–1209	9.65 (64.75)
A&E_3	62	66	57	0–345	29.26 (53.22)
Social Worker_3	62	66	57	0–1416	4.43 (65.43)
Speech therapist_3	62	66	57	0–1025	4.89 (52.51)
Occupational Therapist_3	62	66	57	0–261	2.44 (18.86)
Physiotherapist_3	62	66	57	0–1555	12.89 (107.42)
Educational psychologist_3	62	66	57	0–393	6.52 (33.31)
Psychiatrist_3	62	66	57	0–5252	10.74 (237.53)
Counselling/therapy_3	62	66	57	0–2332	20.35 (137.51)
Dentist_3	62	66	57	0–1247	253.53 (138.02)
Optician_3	62	66	57	0–202	27.00 (32.96)
Police_3	62	66	57	0–623	4.47 (46.62)
Hospital Stay_3	62	66	57	0–1564	21.12 (127.31)
Hospital Outpatient visit_3	62	66	57	0–2902	88.30 (277.06)
GP_4	71	72	70	0–652	44.09 (74.42)
School Nurse_4	71	72	70	0–595	39.07 (67.80)

TABLE 26 Outcome variables for cost (continued)

Variable	Missing values, %			Range	Mean cost (£) (SD)
	Total	ROE	Control		
Education Welfare Officer_4	71	72	70	0–102	0.48 (5.78)
A&E_4	71	72	70	0–204	15.83 (35.80)
Social Worker_4	71	72	70	0–465	2.10 (25.48)
Speech therapist_4	71	72	70	0–84	0.46 (6.19)
Physiotherapist_4	71	72	70	0–3064	16.77 (175.15)
Educational psychologist_4	71	72	70	0–155	1.48 (10.61)
Psychiatrist_4	71	72	70	0–1293	7.58 (84.40)
Counselling/therapy_4	71	72	70	0–919	15.88 (88.97)
Dentist_4	71	72	70	0–614	110.88 (75.35)
Optician_4	71	72	70	0–79	13.46 (14.06)
Police_4	71	72	70	0–307	4.98 (38.87)
Hospital Stay_4	71	72	70	0–5241	30.00 (289.11)
Hospital Outpatient visit_4	71	72	70	0–1787	51.83 (191.40)

A&E, accident and emergency; GP, general practitioner; NA, not applicable.

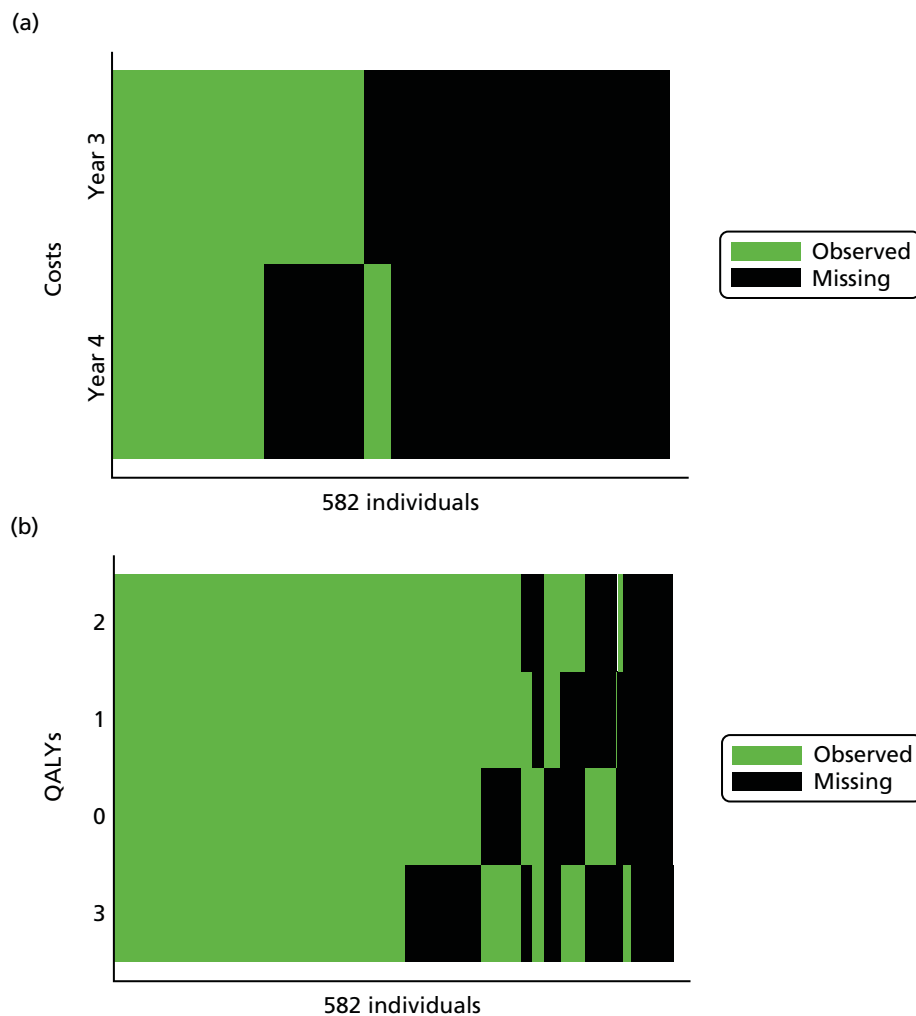


FIGURE 3 Pattern of missing data in (a) costs and (b) QALYs.

can rule out the missing completely at random assumption. For regressions that explored the association between missing data and observed outcomes, at least one covariate produced statistically significant results indicating that the data are unlikely to be missing completely at random and are, thus, assumed to be missing at random.

Clustering in economic evaluation

A simple multilevel model of cost was fit, but owing to issues with the design of the trial (i.e. resource use was only collected at second- and third-year follow-up), the data did not fit this type of model as there were only two time points for cost. The ICC was estimated for cost, and it was low, at 0.0055. The low ICC was deemed to have only a very small design effect for this outcome, so robust standard errors were reported within the generalised linear model regressions to account for clustering in the uncertainty estimates.

Costs and quality-adjusted life-years

The total cost of providing ROE over one academic year to 33 schools in four out of the five health and social care trusts throughout Northern Ireland was £133,866. The average cost was £4057 per school and £175 per pupil. *Table 27* summarises the intervention costs.

In addition to information on the costs of the intervention, information on public sector service use was collected. The resource use unit costs are reported in *Table 28*. Most resource use costs were valued using unit costs from the Personal Social Services Research Unit's *Unit Costs of Health and Social Care 2014*.⁵⁹ The mean cost (including intervention and resource use costs) for the ROE group was £1181 and the mean cost for the control group was £1028. The incremental cost was £153 (95% CI £14 to £292), significantly higher for ROE ($p = 0.032$). The additional cost of the intervention is the main cost driver in this incremental cost.

The mean QALY gain in the ROE group was 3.0908 versus 3.0748 for the control. The incremental QALY gain of 0.0160 (95% CI -0.0143 to 0.0462) was not statistically significant ($p = 0.300$). The results of the base-case cost-utility analysis, as well as those of the sensitivity analyses, are reported in *Table 29*.

TABLE 27 Cost of the ROE intervention

Item	Cost (£)
Key point people	51,419.28
Administrative support	25,793.46
Instructor time	37,231.17
Instructor training materials ^a	7092.94
Instructor materials ^a	3152.42
Instructor fee	5653.83
Other costs ^a	3522.59
Total cost	133,865.69
Cost per school	4056.54
Cost per pupil	175.22

^a Indicates costs that were annuitised over 5 years with a discount rate of 1.5%. These costs were also converted from CAD (2011) and inflated to year 2014 using Purchasing Power Parities. All other costs were accrued in GBP and inflated to 2014 GBP using UK Purchasing Power Parities.

TABLE 28 Unit costs of public sector service use

Variable	Unit cost (£)	Source
GP	46.00	PSSRU 2014 ⁵⁹
School nurse ^a	63.00	PSSRU 2014 ⁵⁹
Education welfare officer ^a	27.00	PSSRU 2014 ⁵⁹
A&E	72.00	<i>NHS Reference Costs 2013/14</i> ⁷⁹
Social worker ^a	41.00	PSSRU 2014 ⁵⁹
Speech therapist	89.00	PSSRU 2014 ⁵⁹
Occupational therapist	113.00	PSSRU 2014 ⁵⁹
Physiotherapist	81.00	PSSRU 2014 ⁵⁹
Educational psychologist ^a	41.00	PSSRU 2014 ⁵⁹
Psychiatrist	228.00	<i>NHS Reference Costs 2013/14</i> ⁷⁹
Counselling/therapy	81.00	PSSRU 2014 ⁵⁹
Dentist	65.00	PSSRU 2014 ⁵⁹
Optician	21.10	Northern Ireland sight test fee MOS/294 ⁸⁰
Police ^a	325.00	PSSRU 2014 ⁵⁹
Hospital stay (number of nights)	326.00	PSSRU 2014 ⁵⁹
Hospital outpatient visit	189.00	PSSRU 2014 ⁵⁹

A&E, accident and emergency; GP, general practitioner; PSSRU, Personal Social Services Research Unit.
 a Indicates societal cost.

Cost-effectiveness

The ICER was £9571 per QALY gained (95% CI –£87,776 to £106,676) (see *Table 29*). Uncertainty around this estimate was explored through bootstrapping. The cost-effectiveness plane is presented in *Figure 4*. At a cost-effectiveness threshold of £20,000, ROE had an 83.1% probability of being cost-effective. This probability rises to 90.1% at a threshold of £30,000. The cost-effectiveness acceptability curve is presented in *Figure 5*.

Sensitivity analysis

The cost-effectiveness analysis of the SDQ total difficulties score resulted in an ICER of £197 per one-unit decrease in the total difficulties score. In this sensitivity analysis, the costs were not significantly different ($p = 0.149$) but the effects were ($p = 0.024$). The available case analysis explored the uncertainty around multiple imputation and the missing at random assumption to assess the impact that multiple imputation had on incremental cost and QALY estimates. The available case analysis mean costs were £1132 for ROE and £894 for the control. The incremental cost of £238 was statistically significantly higher for ROE ($p = 0.01$). The mean QALY gain was 3.0932 for ROE and 3.0811 for the control. The incremental QALY gain was not statistically significantly higher, at 0.012137 ($p = 0.544$). The ICER was £19,626 per QALY (95% CI bootstrap –£149,124 to £144,577). The probability that ROE is cost-effective is 77.3% at a £20,000 threshold and 86.7% at a £30,000 threshold. The rest of the analyses did not have statistically significant effects at the final follow-up but the programme still had a high probability of being cost-effective (see *Table 29*).

TABLE 29 Cost-effectiveness results

Analysis	Mean costs (£)				Mean effects				ICER (£ per QALY)	95% CI of bootstrapped ICER (£)	Probability that ROE is cost-effective, ^b %
	ROE	Control	Incremental costs (95% CI)	Robust standard error ^a	ROE	Control	Incremental effects (95% CI)	Robust standard error ^a			
SA0	1181	1028	153 (14 to 292)	70.9487	3.0908	3.0748	0.0160 (-0.0143 to 0.0462)	0.0154	9571	-87,776 to 106,676	83.1 (90.1)
SA1	1170	1063	107 (-38 to 252)	73.7271	1.1686	0.6272	0.5414 (0.0718 to 1.011)	0.2394	197 ^c	77 to 471	100 (100) ^d
SA2	1192	1038	154 (12 to 297)	72.3742	-0.5469	-0.5743	0.0274 (-0.3487 to 0.4034)	0.1917	5630 ^e	-23,402 to 29,140	96.7 (97.5) ^d
SA3	1187	1026	161 (14 to 307)	74.5814	2.9693	2.9546	0.0147 (-0.0228 to 0.0522)	0.0191	8398	-95,861 to 142,246	84.4 (90)
SA4	1251	1028	222 (82 to 362)	70.9255	3.0908	3.0748	0.0160 (-0.0143 to 0.04623)	0.0154	13,909	-125,331 to 150,800	75.2 (84.2)
SA5	1193	1028	165 (25 to 304)	70.9398	3.0908	3.0748	0.0160 (-0.01436 to 0.0462)	0.0154	10,309	-93,718 to 114,187	82.1 (88.6)
SA6	1119	965	154 (17 to 290)	69.455	2.9637	2.9477	0.0159 (-0.0128 to 0.0446)	0.0146	9660	-94,523 to 112,977	82.5 (89.8)
SA7	1132	894	238 (58 to 419)	92.1694	3.0932	3.0811	0.0121 (-0.0271 to 0.0514)	0.02	19,626	-149,124 to 144,577	77.3 (86.3)

SA, sensitivity analysis.

a Adjusted for 66 clusters in school.

b At £20,000 per QALY (£30,000 per QALY).

c ICER per unit decrease in SDQ total difficulties score.

d Illustrative only: a hypothetical £20,000 (£30,000) threshold per unit increase/decrease in SDQ scores.

e ICER per unit increase in SDQ prosocial behaviour score.

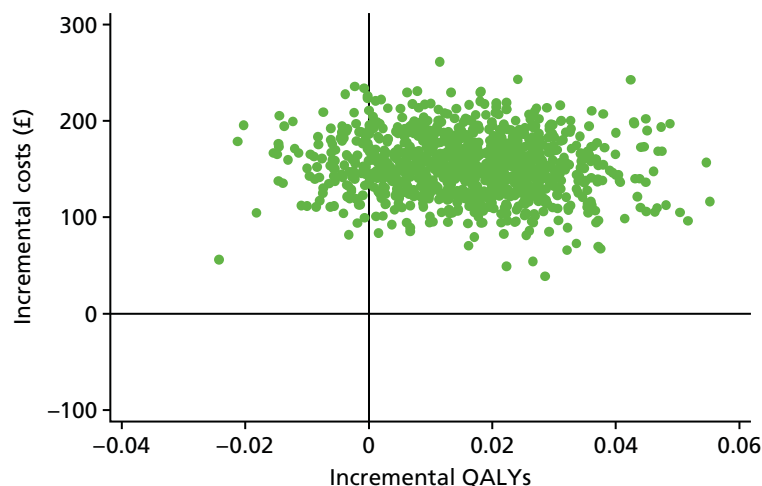


FIGURE 4 Cost-effectiveness plane representing 1000 bootstrapped cost and QALY pairs.

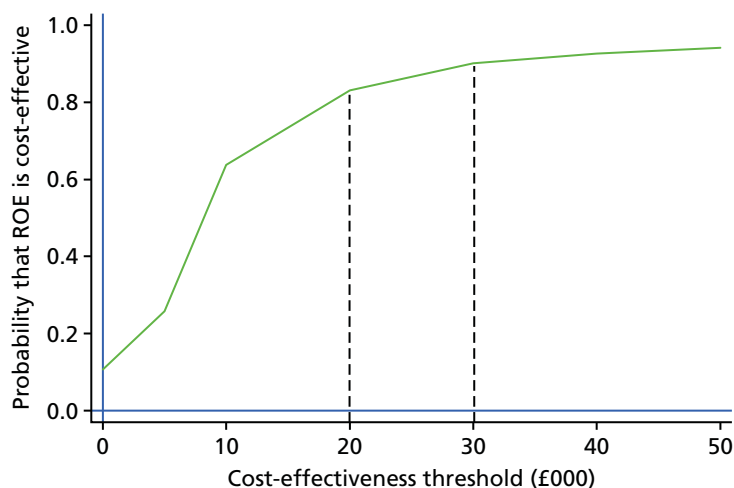


FIGURE 5 Cost-effectiveness acceptability curve showing probability of ROE being cost-effectiveness compared with the control.

Discussion

During the trial period, the base-case analysis indicated that the ROE intervention incurred a mean additional cost of £153 (95% CI £14 to £292) per pupil. Utility, as measured by the CHU9D instrument and combined with duration to calculate QALYs, showed no significant QALY difference between the groups at an incremental QALY gain of 0.0160 (95% CI -0.0143 to 0.0462). However, these analyses do not capture any spillover effects, such as QALY impacts on parents and siblings, and other children coming into contact with participants. Given the direction of QALY gain, these impacts are likely to be positive, if anything, and thus spillover effects would most likely strengthen the cost-effectiveness result. QALY gains in other areas of child health-related utility research are often small and insignificant; however, economic evaluation methods still use such estimates to explore the probability of cost-effectiveness when combined with the cost of achieving these gains. When applied across a population, even small QALY gains can be highly cost-effective. A recent study looking at a family-based childhood obesity treatment used the EuroQol-5 Dimensions youth version (EQ-5D-Y) to measure QALYs.⁸¹ The authors reported a non-significant QALY gain of 0.03 (95% CI -0.04 to 0.10). Another recent study for an asthma intervention in children used adult EuroQol-5 Dimensions QALY estimates.⁸² They found a difference in mean QALYs of -0.00017 (95% CI -0.00051 to 0.00018).

This research adds to the evidence from other studies, which have used other outcome measures including mental health, empathy, perspective taking and SDQ, showing that ROE is effective immediately post intervention.^{20–26} However, most studies had no follow-up post test and the only published study that did follow up pupils (3 years post test) similarly found no significant differences in effect after 3 years of follow-up.²⁰ Although the QALY differences between the arms of this randomised controlled trial were not statistically significantly different, the majority of the incremental points lie in the north-east quadrant (see *Figure 4*), indicating a more costly yet more effective intervention. This leads to a high probability of ROE being cost-effective within the £20,000–30,000 per QALY threshold.

The cost-effectiveness analysis of the SDQ total difficulties score (sensitivity analysis 1) was the only effect that was statistically significantly different at the final follow-up between groups. This perhaps reflects the fact that the SDQ is the most sensitive for detecting changes in social-emotional well-being, the main outcome ROE intends to improve.

The CHU9D is appropriate for a QALY framework, but many of its dimensions would not have been affected by ROE (e.g. pain and daily routine). Therefore, its appropriateness for detecting change in social-emotional well-being is questioned. It does, however, capture a generic health improvement. Its nine dimensions – worried, sad, pain, tired, annoyed, school work/homework, sleep, daily routine and ability to join in activities – capture an overall improvement in functioning. One of the hypothesised health outcomes of ROE is that it decreases aggressive and bullying behaviour, so if fewer children are being bullied that may be evidenced in the worried, sad, pain, annoyed, sleep and ability to join in activities dimensions of the CHU9D. The CHU9D is the only health-related quality-of-life instrument designed for children and valued by adolescents, and, thus, it was the best choice for measuring QALYs in children. Other health-related quality-of-life measures for children exist; however, they are usually adapted from an existing adult measure (16D),⁸³ are valued using adult values (the EQ-5D-Y⁸⁴ and the Health Utilities Index Mark 2⁸⁵) or have not been valued at all but are mapped to an adult measure [Pediatric Quality of Life Inventory⁸⁶ (PedsQL)]. This is because it has typically been very difficult to elicit children's health preferences because of ethical and cognitive difficulties. Time trade-off would involve asking children about death, and the ethics of such an activity is questioned. It is also a cognitively challenging task that may not be appropriate for children. It should be noted that the use of different tariffs to score the CHU9D does result in different cost-effectiveness results, with the use of the alternative tariff⁶² resulting in an ICER of nearly £1200 lower than the original tariff.⁴⁶ The use of annuitisation and the assumptions around the useful life of the intervention do have an impact on the cost-effectiveness results. Sensitivity analysis 4, when there was no annuitisation, resulted in an ICER of £13,909 versus £10,309 and £9571 when costs are annuitised over 3 and 5 years, respectively. In this study, the choice between a 1.5% and 3.5% discount rate minimally affects the cost-effectiveness results.

The available case analysis demonstrated the most conservative estimate, with greater incremental costs and lower incremental QALYs resulting in the highest ICER estimate and lowest probability of cost-effectiveness for ROE. However, the ICER of £19,626 for the available case analysis is still well within the limits that NICE typically considers cost-effective. In fact, all reported results of the sensitivity analyses would typically be considered cost-effective by NICE; however, it is important to note that the threshold of £20,000–30,000 per QALY gained is from a NHS and Personal Social Services perspective. If ROE were to be rolled out to schools across Northern Ireland, it is likely that the cost of providing the programme would largely fall on schools or local education authorities, and their willingness to pay for the programme may be very different from that expressed by the current NICE-supported threshold.

The health and medical fields have long used cost–utility analyses to aid policy decision-making. Without such analyses, there is a risk that decisions will be made based on emotional appeal, absolute intervention cost and political pressure.⁸⁷ This cost–utility analysis (base-case analysis or sensitivity analysis 0) provides initial evidence that school-based population health interventions are feasible, are likely to be cost-effective according to current thresholds and can be employed to aid decision-making.

Limitations

Ideally, data on resource use would have been collected at each data collection time point. It was recognised that recall bias was likely as a result of the long recall periods for estimating resource use expenditure; however, the alternative was to completely forego collecting any resource use information for the trial. The lack of consistently collected resource use data was the main limitation of this cost–utility analysis, which also had a limiting effect on the choice of analytical methods employed.

The available resource use data were also limited by large numbers of missing data. Variables with the largest numbers of missing data may have been affected by a survey design effect as they were all questions that were self-reported using free-form text. Therefore, a detailed descriptive analysis was employed to determine the appropriate assumptions around the missing data, and missing data were subsequently handled using multiple imputation. Future evaluation work of school-based population health interventions should be mindful of potentially large numbers of missing data, particularly those that are collected from parents by post.

It would have been useful to explore the longer-term impacts of ROE by modelling potential impacts over the child's lifetime. However, there is a paucity of longer-term evidence using the main outcomes of our analysis, the SDQ and CHU9D, especially the latter, which is a relatively new generic health-related quality-of-life measure. Additionally, the lack of statistically significant difference in effects at the third-year follow-up meant that any potential longer-term benefits would have significant assumptions and uncertainty attached.

Conclusions

Overall, and in relation to the two primary outcomes, there is evidence that the ROE programme has achieved small positive effects in relation to increasing prosocial behaviour ($g = +0.20$) and reducing difficult behaviour ($g = -0.16$) immediately post test. Moreover, these effects are consistent with those found from the meta-analysis of other existing evaluations of ROE ($+0.13$ for prosocial behaviour and -0.18 for aggressive behaviour). Interestingly, the lack of evidence of effects in relation to secondary outcomes also appears consistent with previous studies.

In relation to the longer-term impact of the programme, although the positive effect for prosocial behaviour has disappeared 1 year on from the completion of the programme, the size of the effect in relation to the reduction in difficult behaviour has been sustained across the 3-year follow-up. However, and because of a reduced sample size, this effect is no longer statistically significant and thus cannot be cited as evidence, in itself, of the programme being effective in reducing difficult behaviour 3 years on from its delivery.

With regard to the secondary outcomes, it is notable that there is no evidence that the ROE programme has had any effects in improving outcomes above and beyond those achieved by schools continuing as normal. One issue that emerged is the fact that the ROE programme appears to have replaced the traditional activities that schools used to deliver the PDMU element of the Northern Ireland curriculum, rather than being delivered *in addition* to these.

Finally, and with regard to the economic evaluation, this study shows that, within the current thresholds for the value of a QALY, ROE is likely to be a cost-effective school-based population health intervention. However, important additional sensitivity analyses relating to the total budgetary impact of rolling out this intervention, assumptions about ROE intervention lifespan and longer-term quality-of-life benefits are required to draw definitive conclusions about longer-term cost-effectiveness. In addition, future studies are needed to compare ROE interventions with alternative interventions aiming to achieve the same social and emotional well-being gains.

Chapter 4 Process evaluation

Introduction

This chapter presents the findings from the qualitative process evaluation that was undertaken to examine the implementation and fidelity of the ROE programme. More specifically, it seeks to:

- address how the programme was delivered across different sites, identifying any variations in implementation and any other relevant factors for which differences may be evident (e.g. whether or not all lessons were covered, commitment of volunteer mothers, timetable and resources)
- provide insights into elements of the programme that tended to work or not, and the reasons why
- document the experiences and perspectives of key stakeholders on the programme, the extent to which the programme was delivered as planned and the reasons for the findings to subsequently emerge from the main trial.

The chapter provides evidence from the in-depth observational data and an overview of all of the key stakeholders' views about and perceptions of the implementation and fidelity of the ROE programme relating to site observations, programme content, mentoring and support, programme limitations and challenges. This is followed by a description of the stakeholders' perceptions of the benefits of the ROE programme and suggestions for how the programme may be further improved, after which there is a final section on parental engagement with the programme.

Implementation and fidelity

Overall, the programme was found to have been delivered with high fidelity. Of the 33 schools delivering the intervention, all successfully completed the nine themes and 27 lessons, and no instructors or volunteer mothers left during the implementation phase. The programme co-ordinators also reported that all participating schools completed the programme, covering all of the lessons with a trained ROE instructor. This tallied with findings from the analysis of information recorded by instructors on lessons covered, length of time and whether or not the teacher was present. Detailed site observations were conducted on three full sessions at all six case study schools (18 observational visits) during the programme delivery. The observation schedule involved observing a variety of lessons and collecting information related to adherence to the intervention, dosage or exposure received by participants, quality of delivery of the lesson, participant responsiveness and programme differentiation.⁴⁵ The schedule was designed in conjunction with the content of the manual. In terms of fidelity, it was observed that instructors were dedicated and worked hard to maintain fidelity, that fidelity to the manual was very high and that there did not seem to be many occurrences of informal adaptation.

The passionate commitment of staff involved with implementing the ROE programme emerged from the interviews with the key stakeholders. In all of the case study schools, although the ROE instructor was the person delivering the lessons, they reported receiving support from the class teacher and school principal. This is evident in the following quotations:

Our principal is so committed to it that it wasn't an issue. I have them an hour every Wednesday morning. In that hour I can fit in my preparation and photocopying and delivering it.

Instructor

The teacher and principal in my school are amazingly supportive of the programme.

Instructor

The process evaluation raised a few issues relating to the ROE instructor's role. In several cases the ROE instructor was also a teacher in the school, and most stakeholders, for a variety of reasons, considered this an advantage. First, it was considered that if the instructor was familiar with the school timetable and pressures, then this would facilitate a more seamless delivery of the programme:

Our instructor teaches part-time in the school. She knows my class well and is familiar with the ones that are troublesome. That really helps because she takes no nonsense from them.

Teacher

We meet up most days and have a chat about how things are going. It works well because sometimes we want to discuss a particular child or issue.

Teacher

Second, effective communication between instructor and teacher was highlighted as important to maximise the benefits of the programme for the children. It was suggested that, where the instructor was also a teacher at the school, there would be more opportunities to meet informally and talk about the programme. This was highlighted as a major advantage because of the constraints of the school timetable, and would, in turn, lead to better outcomes for the children. As one teacher stated:

It's great because the instructor always lets me know in advance what the next lesson is. We also get the opportunity to chat after the lessons about anything that arises with the children. This can only be good for improving the benefits of the programme for the children.

Teacher

Another instructor echoed this sentiment:

We can have a chat in the staff room over coffee about any issues that arise in the ROE lessons and this is great because we are always under pressure for time.

Instructor

On a related note, most of the programme co-ordinators raised concerns that some schools did not put teachers forward for the role of ROE instructor because of limited time and financial resources. In general, when the instructor was not from the same school, the programme still ran smoothly. However, because ROE was in its first year of implementation, there were concerns about the teacher not knowing in advance what themes were being covered in lessons as this reduced their ability to tie the themes in with other lessons in the class. Again, it was considered an advantage if the instructor was a teacher in the school, as this allowed for more opportunities to talk about and reflect on the ROE lessons. These concerns are reflected in the following comments from a programme co-ordinator:

Some schools are very reluctant to put a teacher forward to be an instructor due to financial pressures and time constraints. I suppose they have to release the teacher for 4 days training in the school year and then they have to free them up for possibly 2 hours per week to prepare and facilitate the lessons. I try to explain to them that after the first year then they have their own ROE trained instructor within the school who does not require any further training just mentoring.

Programme co-ordinator

Although all of the instructors reported being on target to deliver the programme, some raised concerns that, because this was the first year of the ROE's implementation, there were a lot of lessons to learn and a lot of ground to cover. Both instructors and teachers suggested that in future years this might not be an issue, as the teachers in the school would become familiar with the lesson themes. One teacher summarised this well, saying:

I think when I become more familiar with the lessons, I can tie them in with some of my lessons, for example artwork or science. Because this is the first year in the school for the ROE, I am learning myself.

Teacher

An instructor echoed this sentiment:

This year there is a lot for all of us to learn, as we are all new to it. The manual is very easy to follow and does give a lot of guidance and advice. Next year hopefully I won't have to refer to the manual as much.

Instructor

Programme content

Most of the references interviewees made to the ROE curriculum and teaching materials were very positive. Key stakeholders had many suggestions about what helped to make the ROE programme work well. The nine lesson themes were considered very strong, and most of the teachers, principals and instructors suggested that these crossed paths with all of the subjects on the revised Northern Ireland curriculum, with the exception of physical education. Teachers mentioned that the following areas of the curriculum overlapped: language and literacy, mathematics and numeracy, PDMU and the arts. This is reflected in the following comments:

It is cross-curricular; all that stuff about smoking and drinking, healthy eating will be covered again at some stage.

Teacher

If I was going to deliver it again, and I hope I do, then I know as a teacher I can make it have significantly enhanced impact. For example, at various stages in the nine cycles of the Roots of Empathy programme, it touches everything that we do in Year 5.

Instructor

So I try to run on a sort of twin-track concurrent basis in maths and literacy and ICT and link to everything with the ROE lessons but that's not easy as this is the first year. But next year if we did that it would be consistent and [there would be] constant reinforcement of the messages of the ROE lessons. Next year, I now know from my own notes how it progresses, so I can – right, I won't do weighing in kilograms in the second half of the Christmas term, I'll do it in the first half because it links with this is when the baby is born and now the baby's 3 months or maybe 2 months. And I think that would be very useful, and I know that's shared with my colleagues.

Teacher

Most respondents suggested that the ROE curriculum complemented the PDMU learning and was considered a key building block in enriching children's pastoral experience in the school. The importance of this is reflected in the following comments:

Well, it mostly ties in with our PDMU personal development and mutual understanding programme, it covers that so well.

Instructor

We need scores higher, higher, higher, data, data, so unfortunately, schools are now focused on data. But having said that, children will not succeed in Key Stage 1 and Key Stage 2, unless they are emotionally stable, and if their homes are good, and there's understanding, there's empathy there, and the children are pastorally looked after. That goes without saying it. The key building block of a good school is the pastoral experience. And thankfully in our school, ours was described as outstanding, so it's something we're very proud of and the ROE is helping to build on that.

Principal

The instructors and teachers were very positive about the materials used to support the ROE programme, with particular reference made to the fact that the books were very diverse, multicultural and inclusive. Instructors made comments such as the following:

There's such a diverse range in the books between cultures and whether it's, you know, male, female, all that sort of thing. You know, everything there brings in every aspect of inclusion in my opinion, and I think that the books are so good that way that, you know, there's some things with children with glasses, with children that are black, with children that are Chinese.

Instructor

The stories are brilliant. A wee bit American, but the kids don't mind. They know about recess and all that, so they're all happy with that.

Instructor

I thought the books were wonderful and the children got the chance to see how other cultures lived.

Teacher

A few instructors suggested that some of the books were aimed at a lower age group but that each lesson theme had a wide selection of books to choose from and this usually helped to overcome the problem:

There are such a variety of books to choose from. Even if one seems not age appropriate for the classes there are plenty more to choose from. The books are excellent.

Instructor

Sometimes I think this story is going to be too childish for the class, but then I go ahead and give it a go and they love it. The stories are all so new to them.

Instructor

Children, particularly some of the boys, concurred with the above comments; they suggested that the stories were more suited to younger children and referred to some of the stories as not age appropriate. This is reflected in the following comments:

They're good enough but like after a while you could think that they're quite babyish or something.

Child

Some of the books are really for much younger children because like, you know, all the stuff in them and it's just really not right . . . like you enjoy it.

Child

When asked about the stories told in the ROE lessons, the children were very enthusiastic, and most of them could recall the name of their favourite story and recount some of the details:

I like it because it's sort of like if you closed your eyes you'd be able to . . . it's like the book is actually talking to you and it's like it's just happening in your head. Instead of you having to read it.

Child

Yeah, Daniel's Cape and it was about a little boy who lost his cape and he didn't feel brave without it but when he found it and he had a lot of fun.

Child

The story, I think it was called Jamaica and the Boots.

Child

Teachers reported that the children greatly enjoyed having stories read to them and that this was something that seldom happened at the Primary 5 stage. One teacher explained this well:

Well P5s [Primary 5s], when you read the story to them they listened and you showed them the pictures which in a lot of ways is not age appropriate when you think about it but the kids loved this. I think is because there so much emphasis on our kids now to be reading and to be doing, doing, doing.

Teacher

A few teachers described parts of the programme content as not very realistic, but most instructors and teachers interviewed were very positive about the content and, in particular, about how it conveyed very simple, uncomplicated messages to the children and how the repetition of activities reinforced the learning outcomes. A few instructors and teachers highlighted that there was a lot of repetition with the artwork tasks and that sometimes the boys, in particular, would become bored with this. This is reflected in the following comments:

I just thought that the pictures and the stories and the underlying meaning were very simple. And yes, OK, maybe it wasn't very advanced but it was meant to be simple.

Instructor

And the messages were simple, and it's very repetitive, but that's how we learn.

Teacher

And going round the artwork, and again there was a lot of artwork and sometimes the boys don't enjoy this as much as the girls.

Instructor

Most interviewees, including the children, responded well to singing the greeting and goodbye songs to the baby at the start and at the end of every lesson, respectively. This was considered a very special time in the lesson for all of the children and they would look forward to seeing how the baby would react to their welcome. Some teachers commented that this was a very happy occasion and that very few of the children did not enjoy this. In some of the classes it was reported that some boys felt awkward about singing the song to the baby at the beginning of the year, but this really changed as the lessons progressed and they built up a relationship with the baby:

All the boys and girls in my class just love it when the baby arrives and they sing the song.

Teacher

We are always in great form when the baby and his mum arrive. We love singing the song.

Instructor

At the beginning, some of the boys in class just didn't sing the song, they just stood there, sort of shy or awkward. But that has all disappeared. They sort of forget themselves now and just automatically join in and don't look around to see who is looking at them.

Instructor

Mentoring and support

Fidelity to ROE is supported by a mentoring and monitoring system for instructors. Instructors undergo a total of 4 days' intensive training, delivered directly by a specialist ROE trainer from Canada, before they commence the programme and midway through the academic year. The specialist trainer also provides ongoing mentoring support via regular telephone calls to all instructors during the school year that the programme is running. In addition, ongoing support is available to each instructor from each health and social care trust's lead ROE co-ordinator.

Programme co-ordinators and ROE instructors were asked how they found the training and support that they received from ROE Canada. All who were interviewed were positive about this, particularly about the ongoing support from and communication with their mentors. For example, they made reference to the commitment demonstrated by their mentors, how quickly they received responses to queries and problems, and how well informed they felt. The programme co-ordinator was felt to be a great source of support for instructors when implementing the programme. The overwhelming majority of instructors highlighted this as a main influence on the high fidelity of the programme, and because all programme co-ordinators were trained ROE instructors, the instructors felt that they had the necessary experience and first-hand knowledge of the programme:

Yeah, the support from Canada was very, very good. I mean she's been very thorough and she had trouble getting me so many times but she never gives up. She's very, very good, and you really do have the sense that if you asked her anything she would get straight back to you. Even though she's away in Canada, she's just the end of the e-mail and the end of the phone, and when she came and she did the observation with us she got back to me then and I'd say it was a month later just to say, have you been able to do this and remember these were the things that you were going to work on? You know, like nice friendly, gentle reminders. And the training, the mid-year training was very good.

Instructor

The training was very good, very intensive. And you got all these files, you're thinking – but it was very good, I think, the way it was delivered. When we were up doing things the first or second day, we were up teaching lessons.

Instructor

Limitations of the programme

Teachers' perspectives of the limitations of Roots of Empathy

Although teachers were, on the whole, very positive about the ROE programme, they did report a number of perceived limitations, with three key issues reported in particular. First, some mentioned the fact that the programme would not continue after the current year:

It's just that it's only 1 year. I think that's the limitation. I think that it needs to be picked up on again. I think actually the other staff here aren't doing it to . . . like, the P6 teachers should have more of an awareness of what's going on to be able to carry that on. Because I would worry that it would just be something that happened in P5.

Teacher

Second, teachers referred to the programme as being quite rigid and unimaginative at times, with too much repetition of certain activities, such as artwork:

It can be very prescribed and set in stone.

Teacher

I feel that maybe the baby end of things is very good and a lot of the lessons are very good, but some of the activities are very repetitive. They're drawing a picture a lot and discussing their feelings and that.

Teacher

There's one child in the class who actually said – he probably is on the autism spectrum and he said – 'But I just did that'. You know, because he just says it as it is, 'But I did that before'. And I could see where he was coming from, because a lot of the drawings or whatever they have to do, they could be repetitive.

Teacher

I know the person giving them would expect that maybe they would draw a different picture or discuss about it a different scenario but very often some of the children would use the same scenario so therefore they're drawing the same thing. So I think it falls down a wee bit there.

Teacher

I don't know what child care courses are like, right, but I do feel that a lot of it is sort of child caring, minding the baby and watching the baby, and I just feel that our curriculum addresses a lot of these areas. It sort of in a way is a wee bit repetitive and with a little bit of imagination it could be extended in a worthwhile way.

Teacher

Finally, the consensus from all of the teachers was that parents did not know much about the programme and yet were crucial to its success. This is reflected in the following comment:

It would be lovely for parents to actually see more or hear more; get more feedback, so they can see what their actual children are learning. Because, really, they've got their questionnaire, which gives them their information, and that's really all they know about it.

Teacher

Another teacher commented that:

Some parents ask me about the baby because their child goes home and talks about it. There is no real direction on how we should involve parents, so that's a bit ad hoc.

Teacher

School principals' perspectives of the limitations of Roots of Empathy

When asked about the limitations of ROE, many principals had very strong concerns about the lack of resources in schools and insufficient future funding for sustaining the programme. Other principals supported this, as they felt under pressure to release teachers from their usual duties so that they could take on the instructor role, and also because they had to find funding for cover for those teachers' classes. Principals also considered time constraints a drawback of the programme; for example, they noted the need for preparation time so that the lessons could be built into the timetable, as well as an opportunity for the instructor and teacher to discuss each lesson in advance and then reflect on it afterwards. Some of the principals felt that it might not always be possible to give adequate time to this, owing to the pressures of other workloads on the instructor and teacher:

I would say the limitations are that I can't afford to have the likes of [instructor] to every class in the school. If I had my way and I had the money, I would have her deliver it in every class. I might not be able to get enough babies but I could see benefits of this right throughout the school. It would be getting the money to do this and then sustain the programme.

Principal

Resources, and funding, and release of teachers, and substitute cover, and back to that whole issue of resources and funding to facilitate that, without appearing to be negative, which I don't want to be, because the programme is so positive.

Principal

Well the only negatives I would say here would be the funding, the possible funding. Well I'm thinking of the bigger picture, not our school, about overemphasis on data and league tables, etc., rather than this.

Principal

A limitation discussed by one principal was that, owing to the fidelity of the programme, the school felt restricted in being able to involve parents. Another principal stated:

Limitation as in it would be good to be able to communicate more with the parents and share activity but, at the minute, we can't because everything has to be vetted, you send. . . if you want to do anything in the local school newsletter, it has to be sent away and that really is not very practical so I think if it was really shared a lot more in the class newsletters going out, it would really strengthen the whole process but because it's just kept so insular, I would say that is a limitation of it.

Principal

Instructors' perspectives of the limitations of Roots of Empathy

Most of the instructors had views on the programme's limitations. Some of the issues discussed were time constraints, the prescriptive nature of the programme, school issues, age appropriateness of materials, and cultural adaptability of the programme.

That it's very rigid.

Instructor

You have to stick to it.

Instructor

One limitation is when they do all start to talk, where do you cut it off? When you've only got a certain timeslot in your day, and you don't like to say, 'OK'. And I had to say it a few times, 'OK, the last one, the last one.' That's the problem with it too, just time constraint.

Instructor

I think for the age group that I'm using, I think it's a bit too childish. There were things about losing teeth and although the children have been through it, I think there is maybe room where you could use some lessons from the older age group for the P5 age group.

Instructor

The books, well that book today, because it was American, very American like starting off, I don't what. And like when we were going through different things like diapers and pacifier, and there were words that they knew from TV. But I think if we were going to use it in Northern Ireland, we might need to put another slant on it.

Instructor

Yeah. The limitations are, I suppose, trying to make sure that it's not just the 1 hour, and that it's right through the rest of the week.

Instructor

Well I think one of the difficulties is the content, there's so much content, and the lessons aren't that long. I always feel I'm whizzing through things a lot. To me that's the main drawback really.

Instructor

One of the instructors stated that she did not see any limitations of the programme:

No, none whatsoever.

Instructor

Many of the instructors noted that most of the limitations were related to the programme being in its first year of implementation. They saw this first year as a 'learning curve' and felt that implementation in the second year would be easier in terms of time and familiarity with the materials:

I think after the first year I will be more familiar with the lesson and will actually have all my lesson plans available. I will not have to spend hours preparing and writing.

Instructor

Next year I think I will be able to focus more on the interactions with the baby and the children. This year I am very comfortable with the programme but sometimes you are checking your notes because you don't want to leave anything out.

Instructor

A fundamental issue raised by two of the instructors was the possibility that a child could make a disclosure and that very little guidance and instruction is given about this:

I know, and in some parts, when we're talking about all this nurture and things, some children have said, 'Oh, that wouldn't happen in my house', or, 'We just leave the baby'. And sometimes there's a thing with it, disclosures and things. If a child in the middle of that, I think there might need to be a bit more done. I know a teacher knows.

Instructor

And I suppose anybody who's in the social work background would know that. But maybe there just needs to be a wee bit more awareness that there could be a disclosure, and how that's dealt with by an outsider coming into it.

Instructor

Main challenges

There were common themes in most of the schools regarding the main challenges of implementing the programme, with only one school claiming that this did not bring any challenges and that all of the stakeholders involved were passionately committed to the programme.

Choosing a ROE family was done through the joint work of the principal and the key point person from the trust in each of the case study schools. All of the principals expressed the positive view that this was an excellent opportunity to involve a family from the local community in the programme. Guided by the programme co-ordinator from the trust, school personnel sought a local family who were willing to, and would be in a position to honour their commitment to, visit the classroom on the agreed dates and times for the duration of the programme. This consisted of nine visits in total, approximately one every 3 weeks, each of which lasted between 30 and 40 minutes. Only one of the case study schools identified this as the biggest challenge when implementing the programme:

The biggest challenge for the school was trying to track down a parent, with the child the right age who was prepared to do it.

Principal

Each intervention school was responsible for selecting a suitable school staff member to train to be a ROE instructor. In the event that the school had difficulty identifying a potential instructor, the programme co-ordinator from the relevant health and social care trust identified a suitable instructor from the community. Most of the schools were able to recruit a member of the staff, and only a few experienced

difficulties. In these latter schools, such recruitment was considered a major challenge. As one programme co-ordinator explained:

That was fine, that was fine. As I said, it did cause some confusion at the beginning when there was an instructor that wasn't known to the school and a researcher that wasn't known to the school and different bodies appearing at the door and different faces. And I was getting to know some of the schools as well. Now, in quite a few of the schools I had contacts so I knew, you know, but I have to say in some of the schools they were saying, 'Oh such and such was here', and I thought they were here to be instructor and they were here to do the research.

Programme co-ordinator

Benefits of the programme

Children's perspectives of the benefits of Roots of Empathy

When asked what they had learned from the programme, the overwhelming majority of the children from all of the groups enthusiastically referred to the baby. The children spontaneously linked learning outcomes from the programme with concepts relating to child development, milestones and caring for the baby. As illustrated by the following representative sample of comments taken from across the focus groups, the children were keen to convey their knowledge of the key principles of looking after the baby, such as safety, crying, communicating, and meeting and greeting:

It's about learning how a baby would grow up and seeing a baby grow up.

Child

Different changes from a baby and like its milestones and what it can achieve.

Child

It teaches you how to care for a baby and all that.

Child

How to make your home baby proof.

Child

That babies' teeth comes 6 or 5 weeks after they are born.

Child

Don't put a baby at the top of cot just if they have a blanket they could wriggle down and it could suffocate them.

Child

Never shake a baby because it can damage their brain and they can have problems like spinal injuries or something like that.

Child

Many children across the focus groups displayed understanding of a range of reasons why a baby might cry. In a few of the groups, children indicated that the programme had taught them about this and most children in all of the focus groups talked freely about why a baby might cry and gave solutions for helping the baby to stop:

And it tells you what the baby wants and what's happened to him when he cries.

Child

If a baby cries it means that she needs something or something is wrong.

Child

It's that just she's sore or it's like that she's hungry or she needs the toilet or something else.

Child

Without prompting, some children in only a few of the groups related the learning from the ROE lessons to helping them to understand the feelings of others. This then led some to reiterate some responses and triggered other children to talk freely about how they had learned about feelings and emotions:

It's like walking in someone else's shoes.

Child

It's trying to understand what other people feel like.

Child

I can understand my sister more when she has to do her homework and then I need her to help me I have to let her do what she needs to do first.

Child

Our different emotions and things like that.

Child

It's helped me to understand my baby brother and sister better.

Child

Teachers' perspectives of the benefits of Roots of Empathy

All of the teachers commented on their role in the ROE lessons as being unique, as they were observers rather than active participants. One teacher commented that she felt that she was in a privileged position as an observer in her classroom, as she was able to observe behaviours from the children that, as a class teacher, she would never have been able to because the school timetable was so busy. This teacher stated that:

I wasn't really sure what my role was, but now I can see that just by being able to sit and listen and watch the children in my class is something I rarely get to do! One child in my class was talking about the content of her dreams during the lesson and I know there are difficulties going on in her home life. That really signalled to me to maybe just watch out and maybe even follow that up.

Teacher

All but one of the class teachers reported experiencing positive benefits and outcomes in the classroom and that there had been a change in some children's behaviour as a result of the lessons. The following comments from teachers reflect this:

I think they show more empathy.

Teacher

I see a softer side come out in them.

Teacher

I can see progression with the children in my class.

Teacher

When the teachers were asked what they thought the main benefits of the programme were for the children, nearly all felt that the majority of children had benefited, with only one teacher stating that so far she could see few benefits for her class, although she was optimistic that these could become evident in the future. This important point is found in the following quotation:

They struggle to name their feelings and to generally express them and it comes out in other ways. I can see that they enjoy the ROE and having the baby in and we do talk about feelings more now. So there is little bit of improvement but it is still not great but I think there could be benefits further down the line.

Teacher

Some of the changes in behaviour suggested by the teachers were that children seemed to be more empathetic and were more able to identify and reflect on their own feelings and those of others, and that many of the children were showing a softer side to their personality. This is illustrated by the following comments:

The main thing would be that the children are more in touch with their feelings, and they are certainly more open to talk about their feelings than they were at the beginning of the programme and they do tend to see other people's point of view more easily now than they did. I think there are less rows in the playground.

Teacher

They are a bit more considerate now of each other and really don't want to end up in the principal's office after calling someone a name on the playground.

Teacher

Some of these teachers made a link between the 'experiential learning' aspect of the programme and the children being more open in their discussions about their feelings. All of the teachers mentioned that the baby as the 'teacher' was a lever for the children to learn through observation and interaction with the baby and its mother, as reflected in the following comments:

The baby is the teacher in the class and when it comes in and it shows you its emotions, like say you and it like laughs and moves, it's showing it's like happy and its emotions to the children.

Teacher

This is real life. Very powerful.

Teacher

The baby is a 3D figure, she is real and the children really get this as opposed to learning these lessons from a book.

Teacher

The children just love the baby.

Teacher

As demonstrated by the following quotations, teachers also mentioned that children were more able to express their feelings, label their feelings with appropriate language and show more consideration to others:

I do feel they are showing slightly more empathy with each other. So if there was a child on the fringes, they know the right answers. Whether they actually do the right thing, you know, they're still children, they still have to be prompted, so they know what they should be doing but I wouldn't necessarily say that they're automatically doing it any more than they would have done in previous years that didn't have this programme.

Teacher

They're definitely more open to their feelings to say, 'I feel sad', or, 'I'm happy', or why they're happy.
Teacher

They do explore their feelings and in that end, yeah, their language would have changed.
Teacher

A key observation expressed by many the teachers was with regard to the perceived change in boys. The teachers were surprised at how well the boys engaged with the programme, particularly in being more open when talking about their feelings. As one teacher cogently put it:

It's the boys that are showing their feelings a lot more. Girls will always have shown their feelings, and at that age group of girls, they tend to be very competitive and one of them wants to lead and be the leader and all that, so in many ways that hasn't changed, which is a bit of a disappointment. But the boys are opening up more, that's all the way I could put it.

Teacher

Similarly, other teachers commented on this important point:

Yes I think the boys have come round to the programme. I don't think they have always been engaged.

Teacher

Some of the wee boys will just come out with one word like 'sad'.

Teacher

At the beginning the boys wouldn't hold the doll but now they just nurse it so naturally in the absence of the baby.

Teacher

Some teachers felt that younger children would have benefited from receiving the programme at an earlier age as it may have helped those who are now struggling to talk about their feelings and this quite often manifests itself in behavioural issues. One teacher said that she noticed only small positive changes in the classroom and that most of these benefits were seen in the boys. She continued to report that many of the children in the class, particularly the boys, had major behavioural problems and found it very difficult to express themselves; during the period of the fieldwork, one boy had been at risk of being expelled from the school for engaging in challenging behaviours that required disciplinary action. Many children had poor language skills and their emotional literacy levels were considered low or non-existent. The Primary 5 teacher expressed her exasperation by referring to the class as 'a difficult class':

I think actually 'Seeds of Empathy' would be something that I would be very interested in looking at, to get in younger, because we're having kids in primary school who maybe don't have the experiences of playing with other children than they might have done in the past as they are on PlayStations [Sony, London, UK], etc. So I think there can be difficulties for kids in school when it comes to turn-taking and co-operative play. I think it would be good for them to have the experience of that.

Teacher

Another teacher, who also thought that younger children would benefit from the programme, said:

I think the younger children in the school would benefit from this. The earlier these programmes are given the better the outcomes for them.

Teacher

The teachers were asked if they had noticed any changes in the classroom environment in general as a result of the programme. There were mixed reactions to this question, with some of the teachers commenting that, in general, the children were a little more sensitive and a lot of trivial arguments that used to take place between the children had stopped. As one teacher explained:

I don't see as many wee silly arguments or they don't keep going on and on, oh for example, 'such and such a one fell out with me'. We used to try to work these problems through PDMU and we've related it now to ROE and they can identify with that.

Teacher

A few teachers who had said that some of the children in their class had behavioural problems commented that the programme had made a difference to these children. One teacher mentioned that there had been a serious issue with bullying in the class and commented on how the ROE lessons had helped the children to think more about their actions:

There were bullying issues going on, and I really feel that the programme highlighted those and it didn't make anybody stand out, but it gave them a chance to reflect on what that was. And it worked as well for the person that was maybe doing the bullying or the teasing as it worked for the weaker person which I think is great.

Teacher

One teacher commented that the programme was a great tool to refer to if children were misbehaving in class. A few of the teachers felt that it was too early to see changes but again they were hopeful that there would be some before the end of the school year. This is reflected in the comment below:

I think it's helped a bit but I think you'll see more of it near the end, like already I've seen from February.

Teacher

Children learning emotional language was considered a major strength of the programme. One teacher shared that most of the children in her class did not have the language to express their feelings and instead often expressed them using anger and hyperactivity:

A lot of the children in my class don't have the words to describe feelings even if they wanted to. They may know primary feelings like happy or sad, but as they often express sadness in anger, happiness in hyperactivity, and anger in temper, then translating their behaviour back to its possible underlying emotion can be helpful. It is difficult to communicate feelings with no basic emotional vocabulary.

Teacher

A few teachers commented on the fact that they did not really know what to look out for in terms of benefits to the children or how they should be building on the ROE lessons afterwards, and that this was definitely a weak aspect of the programme. Such findings raise key questions for this research study in terms of how teachers view their role. The following teachers' comments reflect this:

I think at the beginning of the programme the P5 teacher has nil visibility on what's coming, and from my understanding it's deliberately kept that way as the teacher is not really supposed to engage other than scribe and assist the instructor. I do think that's to the detriment of the programme, because if I had access to – and I understand that I can now but I couldn't before – if I could have access to the handbook and even know, right, next week or for the next 4 weeks we're going to be concentrating on keeping baby safe, I can link that into health and safety from a mathematical point of view, why does a measurement have to be exact if we're going to be measuring doses of medicine for example.

Teacher

This is all new to me so I am not sure what the lesson is going to cover in advance. I could probably tie it in with other parts of my lessons with the children if I did know.

Teacher

I think we should have access to the manual, even a photocopy. We could help prepare the children for the lesson and certainly reflect with them afterwards if we had more information about the lesson. They are very strict about that. It is really only the instructor that has access to the manual.

Teacher

School principals' perspectives of the benefits of Roots of Empathy

School principals were asked if they had observed any outcomes or benefits of the ROE programme in the children. All six principals from the case study schools said that they had observed positive benefits in terms of, for example, less aggressive behaviours, a more caring attitude, improved concentration and fewer bullying incidents. As one principal stated:

But I have definitely seen . . . because I would do dinner duty out there and, yeah, I've seen more of a caring . . . maybe a bit of caring coming. It can be a difficult for them, being able to express themselves.

Principal

Again, the gender difference was recognised by the principals, who made particular comment on improvement in behaviour in a few boys who were seen as 'troublesome'. Those changes included children having fewer disputes in the playground and therefore being sent to the principal's office less often. Some of the words principals used to describe the changes in the children's behaviour and actions since the programme were 'more caring', 'willingness' and 'softness'. This is reflected in the following comments:

There are one or two boys who'd be devilish and there's nothing wrong with that but they always get involved in disputes. But I've definitely seen that there's a willingness among them to be more considerate now. I must say on the playground now, I do the dinner duty every day so I'm out there and behaviour has improved.

Principal

We have found definitely that there's softness has come, and it probably was there but boys are now not afraid to show it, if you know what I mean. Boys are maybe articulating their views more you know, on a playground or in a situation where both feel that they can say their piece and then it's over and hopefully then the bullying will end or, you know, it'll certainly help to stamp out the bullying.

Principal

Three of the school principals discussed how they felt the programmes fitted very well into the school curriculum, in particular how it could build and strengthen the pastoral care aspect:

Where a lot of time, as much as we do circle time, and you were talking through different situations and so on, it's a much more sophisticated form of that.

Principal

The key building block of a good school is the pastoral experience. And thankfully in our school, ours was described as outstanding, so it's something we're very proud of. And we're always trying to enrich that way of working in schools with ROE, or the newest programme.

Principal

Parts of ROE feed into PDMU and improve our pastoral experience.

Principal

Instructors' perspectives of the benefits of Roots of Empathy

When asked about the benefits of the ROE programme for the children, all of the instructors gave very positive views. Their responses included comments on perceived improvements in a range of outcomes, including the children's ability to show empathy, emotional literacy, behaviour in class and the playground, knowledge on child development and parenting skills:

I think there's more tolerance in the original sense of the word tolerance not the condescending version of tolerance. I think they are a wee bit more tolerant.

Instructor

The playground, for example, the way they would normally behave. But then if I have to deal with an issue, they're able to very quickly tell me how the other person would be feeling, and they can empathise, which some of them wouldn't have done before. So it's certainly a way in to dealing with issues.

Instructor

I feel they're more receptive to talking about feelings too. They do seem to, through the course of the year, talk a lot more about feelings to me as an instructor, and I think the teachers would say that they see some of the children doing that. For some of those children they're learning to speak up a bit more in the classroom learning, which you mightn't see in other programs or other bits of the classroom delivery.

Instructor

The instructors said that it was difficult to identify improved language or behaviour at this stage:

But in terms of their development of empathy and their development of emotional literacy it's hard to know really at this stage.

Instructor

The instructors talked about the family visit being a very special and positive experience for the children. The experiential learning gained from the family visit was recognised as this also encouraged the children to talk more about feelings:

Certainly the baby visits are the highlight, and that is where you can get more out of the children, and more language.

Instructor

In my opinion in our school the children, they just drink it all up. You know, so they can't wait for it to come. And their retention of the information just amazes me, you know. Whether it's because it's something new or whether it's because, you know, it's parallel with the baby coming in – there's a real live baby that they can demonstrate what they've learned with or what they've learned before they can then demonstrate it with the baby. It's just that they think it's brilliant.

Instructor

I think they now project their thoughts and feelings onto the baby, it gives a really nice way of understanding when it's OK if the baby is feeling frightened because I sometimes feel frightened. And perhaps it's going along the lines of the bully or the unhelpful or the nasty kind of person, that's probably where it's more targeted trying to understand, because they're older and they see the baby younger and crying they might think, 'Oh that's . . . I don't want to be that type of person. I want to be a nice caring kind of person'.

Instructor

One instructor said that even though the children in her class were poorly behaved most of the time in their normal class as reported by their teacher, they behaved reasonably well during the ROE class and, in particular, during the family visit. She also felt that this particular class had great difficulty describing or talking about their feelings and one reason for this was that they did not have the language to describe how they felt. This instructor felt that these children had progressed to using one word to describe how they thought the baby was feeling because of the learning from the ROE programme, and this was a very positive outcome for the class:

The class are a difficult class. Their behaviour is a lot better with me in the ROE and I really don't have many negative situations. I don't know whether there has been an improvement yet in their overall behaviour but certainly some children has more empathy for others. A lot of our children only use or answer questions with one word. We find that happens right throughout the school, they just say one word and then very often when we hear just the one-word answers we accept it. We are now trying to encourage more, like sentences from the children but it is hard. Some of the wee boys will just come out with one word like sad.

Instructor

Interestingly, all of the instructors commented that there were more notable changes in the boys in terms of being more open when talking about their feelings, and in the relationship that they developed with the baby:

The boys do engage with the baby so well even though they are all so macho.

Instructor

It's the boys that – I just feel that the boys are showing their feelings a lot more. Girls will always have shown their feelings, and at that age group of girls, they tend to be very competitive and one of them wants to lead and be the leader and all that, so in many ways that hasn't changed which is a bit of a disappointment. But the boys are opening up more, that's all the way I could put it.

Instructor

Some of the bigger boys in P6 initially were saying, 'We're not singing', but now they do.

Instructor

The boys love it.

Instructor

Parents' perspective of the benefits of Roots of Empathy

There was a mixed response from parents when they were asked if they had noticed any changes in their child's behaviour because of ROE. Among the parents interviewed, opinions were evenly split among those who noticed a change in their child's behaviour, those who did not and those who were unsure. Parents who noticed positive benefits of the ROE lessons on their child's behaviour, with the exception of one, had sons. Several parents said that they had not noticed any difference in their child's behaviour because of the programme, and these tended to be mothers of daughters. They said they were not sure if this was because they did not really know much about the programme and therefore they were not really looking out for any changes in their child or if they had just not really thought about it. Parents had mixed views as below:

I think the main benefit of the programme, for my son, was that he has just become more thoughtful about the children around him. He would have been selfish at times but I noticed him starting to say sorry more to me for bad behaviour and becoming more affectionate.

Parent

Definitely it had a positive effect on him.

Mother referring to impact of the ROE programme on her son

My child is a very kind child and she would be very – I would say she would have quite a lot of empathy anyway. I'm sure it has benefited her to do it because I think the more things you do with them like this the better. I don't know that I've seen massive differences, she would always be a very kind child; if someone was hurt or whatever she would always have taken the child in and looked after them and stuff. So she was already like that. That would kind of be her personality.

Parent

She loves babies. Her cousins . . . I mean, she's wonderful with her younger cousins and, you know, all the uncles and aunts will always say that. In comparison to any other cousins her age she has always got an awful lot of time for them. Whether that's something that's been enhanced by the programme or not, I don't know, but she's definitely very good with younger children.

Parent

The benefits communicated by parents were related to increases in prosocial behaviours only. No parent suggested that his or her child was less aggressive because of their participation in the programme. Some of the prosocial behaviour changes included being more thoughtful, being more tolerant, showing increased affection and being apologetic:

Yes I think he is a bit more thoughtful to his siblings.

Parent

He is now wanting to nurse his baby cousin a lot more and is really quite affectionate. He is a very kind child anyway.

Parent

I think she maybe finds it a bit easier to say sorry now. Not sure if that is because she is getting more mature or whether its to do with what she is learning in school.

Parent

Well what I did notice from him was that when he was cheeky, he would come to me shortly after and apologise for what he had done and give me a hug, which he hadn't done before. And that he would come and say, 'I'm really sorry for shouting' or 'I'm sorry for doing that' and give me a hug.

Parent

As illustrated below, one mother felt that her son had benefited greatly from the interaction with the baby, as he was an only child with autism who had a very low tolerance of babies, particularly when they cried:

I think it has been very good for my child because he is an only child and doesn't get to see babies much. My son has autism and he doesn't really like babies especially when they cry so I think this programme has helped him to be more tolerant.

Parent

Among the parents interviewed there did not appear to be a high awareness of the content of the programme or of what the outcomes or benefits may be for their children. However, all parents were aware that there was a baby visiting the class, because their children had gone home and talked about it or they had read about it in the school bulletin or on the school noticeboard. When asked what they know about the ROE programme, parents responded:

Ah I think it is lovely having the baby in the class. I'm not sure what it's suppose to do.

Parent

Maybe indirectly. I'm not sure what I'd be looking for. I mean, she's a very empathetic child anyway, she's a very caring child anyway and I am not sure what the programme is suppose to be teaching them.

Parent

Is it about how to look after a baby? How to be a good parent?

Parent

Some of the parents reported that because their child had siblings they were used to being considerate towards other family members:

She's the eldest. OK. So she has younger siblings. OK. Yeah. So whenever my daughter was in P1 we had a baby in the house so you know, she knows and remembers all about bathing the baby, changing the baby's nappy, the baby's crying, needs to be fed, you know, comforting the baby, all that sort of stuff she's went through. So I wouldn't say that it has made an awful lot of difference with her but I would think that maybe with my youngest child who has never had a baby in the house it may make more of a difference.

Parent

I didn't really see any real changes with regards to my daughter. Usually she'll come home and she'll talk about the baby and, you know, what stage it's at or it's growing or it's got bigger or whatever. I wouldn't really say that it has made any difference in her manner or her view towards babies or anything like that, more so because of the fact that . . . Well, [she's] 9 now and she has a smaller brother and a sister.

Parent

In contrast, other parents suggested that the programme had helped their child to be more considerate to siblings:

He didn't. He wouldn't be much of a talker but he would always have been quite an empathetic child anyway, but I suppose yeah it did help. So I suppose towards his other siblings he would be more.

Parent

Well I think the main benefits of the program, for my son anyway, was really . . . obviously when the baby came in. He has brothers and a sister younger than him so he would be used to some exposure to children and young kids. He got really into it in terms that he would talk about it and he would tell me everything that happened with the baby, I almost felt as if I knew the baby myself. So I think he did get very into it and was very excited about the programme and was looking forward to baby coming in to the classroom. So I definitely think that it did give him a bit more empathy for the children.

Parent

One parent expressed concern that her child had been bullied during the year that the ROE programme was running in her class. She was exasperated by the fact that the programme did not appear to be making any difference to the bullies' behaviour or how the school was dealing with it:

I guess my issue with the ROE last year was that my child was still very badly bullied at the time by another child in the class, and we weren't listened to at all by the school. It wasn't a problem but there was a problem, a very big one, which they've now realised actually. They have now recognised it and they're acting on it now. So I kind of was very disappointed because in my head I thought, 'They've got this ROE thing. This is great; this is working with the children, it's teaching them that when you're nasty to someone else or whatever how will they feel? It's kind of building that in'. And yet I felt that nothing was happening for my child. The reality was Yeah. 'We've got this lovely programme, it's great. When the baby comes in and isn't that super?'. And it is, but why is nothing happening to stop this happening to my child.

Parent

In summary, the ROE programme was perceived by most key stakeholders in the case study schools to have positive benefits for the children; the children seemed to show more empathy, open up more about their feelings and, overall, have fewer disputes in the playground. Only one school felt that the children would have benefited more from the programme being implemented at an earlier age; this was because in that school, by the time the programme started, a lot of children had established very tough exteriors and did not have a wide emotional vocabulary. This meant that it was very difficult for these children, particularly the boys, to start talking about emotions and feelings. Teachers, principals, instructors and parents all commented that more benefits had been seen among the boys. The children mainly identified learning about babies and their development with the programme. Parents were generally unsure about the programme and the benefits it may have for their children.

Parental involvement in Roots of Empathy

School personnel's perspectives on parental involvement

There was a disparity between the case study schools in terms of how, or if, they involved parents with the programme. However, all schools stated that they would have liked to involve parents more. When asked how they involved parents in the programme, school personnel and programme co-ordinators referred to a kind of 'ad hoc' involvement, as there was no overall strategy in the manual for engaging parents:

I don't think we have done anything. There are posters around the school so the parents may see those but I don't think there has been anything else.

Teacher

It took longer really for parents to find out what the programme was about, but I did have a call from one mum saying about feelings and her little girl had been involved in a very traumatic situation before that and she didn't like this talk about feelings, she didn't want to talk about feelings, and the mummy was saying, 'What's all this talking about feelings, and she's not coping very well and it's bringing stuff back for her'. So I went to see the mum and spoke to her and said to her, 'OK, and this is hard but obviously she hasn't dealt with the stuff that's going on then', you know. So she was still getting counselling for this particular situation at this time, so I said, 'Look, can we stick through it and we'll see?'. So the outcome on that was OK and the child did stick through it but initially the mummy wanted to take her out.

Instructor

I know that the schools would have written to all parents to tell them that their children were participating in the programme not asking for permission, but I do know that when it came then to the research some of the parents wouldn't allow their children to participate in the research, am I right in saying that?

Programme co-ordinator

Well it's very limited now I think. Certainly in the schools that I work in halfway through I do a wee news sheet and just take a photograph of the baby and say, 'This is a letter to go home to your parents to say, "Look, this is our baby" '.

Instructor

The required strict fidelity to the programme and the media guidelines supplied by programme developers were considered 'off-putting' reasons for deciding against writing to parents or inviting them to take part in the programme during the year. Conversely, a barrier mentioned by school personnel was that parents might be too busy to be involved or that some could be struggling with literacy and so would have difficulty reading information sent home. Some schools identified opportunities during the year to inform parents of the programme's progress, and these schools tended to align this communication with their

own practice for engaging parents. For example, some of the case study schools held school assemblies and distributed school bulletins that parents had access to, and the ROE updates were included in these:

I remember in one of the newsletters, I think the principal put a little bit about the Roots of Empathy programme on that. The name of the baby and that he would be visiting the school, once a month and a little bit about the programme. This was at the start of the year and I don't think there has been anything lately. I know that some schools have talked about having a little assembly.

Instructor

We would have open assemblies, whereby our parents are invited in to see the children just into their classrooms. They just come in and sit in the chairs, and the children talk about what they've been doing, or emphasis, celebration of the feast, or whatever it is that they're doing and the teacher has included it in that. And the board up so it's very visible. So parents can see it when they're coming into the room. But I'm sure there's other ways. But even if we have a meeting, you might only get a few parents that would come.

Principal

I think it was something like, if you're going to send home something, it has to be approved by Canada, so that's time-consuming.

Teacher

I am sure it would be an advantage but you would have some parents who would be illiterate and they wouldn't really understand maybe what it's all about. Certainly it would be good if they could talk to them about it. I would say that some of the parents would actually have forgotten that it is still going on. Perhaps that something the school could look at to try and promote Roots of Empathy and maybe the principal could put monthly updates in the newsletter as to what the themes are and what size baby is, etc., to encourage maybe that home-school link.

Principal

I haven't had any feedback from any parents. But it'll be interesting to see what the parents' view is.

Instructor

A recurring theme on parental engagement was that parents would benefit from knowing more about the ROE programme. Key stakeholders considered it important for parents to know more about the programme because they felt that this would encourage parents to be more involved and improve outcomes for the children by reinforcing lessons learned from the programme at home:

Yeah, it would be nice to send home a newsletter to say look, this is what we've done, even just like a monthly one or like every 2 month one so at least parents then are a bit more involved, like if I'm sure if I did ask the parents they probably wouldn't ever call you, do you know what you mean, but then I suppose you can't really invite them in because it's hard to get them in during the day.

Teacher

I think it's always good if you're embedding something that's really forced home or that they can talk about it and I think that angle of it, at the minute, is somewhat lacking.

Instructor

I think it's something you need to do because what you're also saying to the parents is, this is a really important part of what we do, the whole pastoral social emotional dimension, because sometimes parents need to be educated to the fact that, I mean you're saying that education is not just about the three Rs, it's also about all this other stuff, and then some parent would reflect on challenges they're

thinking on as a parent, because actually, you teach your kids right and wrong, there's a way of dealing with things in terms of how they manage behaviour at home.

Principal

I know that there have been a few issues round parental participation and I know one of the things that was pointed out at one of the original instructors' meetings was that they felt that maybe going forward with the programme a good thing to do would be to bring the parents in advance of the programme and do some sort of a general presentation with them. Some of the schools did that, others of the schools didn't, they just wrote out.

Instructor

Parents' perspectives on parental involvement

There was a mixed response from parents when they were asked how much they knew about ROE. Many of those interviewed knew that their child was taking part in the programme but there did not appear to be a high level of awareness about the content of the programme or what the outcomes or benefits may be for their children. However, all parents were aware that there was a baby visiting the class, mostly as a result of their children going home and talking to them about it or from reading about it in the school bulletin or on the school notice board. Parents shared what they knew about the ROE programme:

The only reason I know this is taking place is because my child comes home and talks about the baby all the time.

Parent

I think the programme is about how to look after a baby, am I right?

Parent

Some of the parents mentioned that being part of the research programme meant that they had received written information about it. There was unanimous agreement from the parents, school personnel and programme co-ordinators that greater parental knowledge about the programme would perhaps increase children's chances of positive outcomes:

I think I got information from Queen's University about taking part in this research. The letter told us what Roots of Empathy was about.

Parent

I think the only time I got any written information about this was from Queen's University. I thought they were running this in school.

Parent

Children's perspectives on parental involvement

According to the children, their parents were really interested in the programme and the ROE baby was often discussed at home:

Well the first week of the ROE I took part I went home and told my dad. And my dad ever since has just known about and he's asked me every time I come home on Friday, how was the baby?

Child

I always tell my parents about how the baby is doing, when he got his first tooth and things. They loved to hear about it.

Child

Although this was less common, a few children said that their parents had no interest in hearing about the programme or the baby:

My mum and dad aren't very interested they don't say much.

Child

Some children said that their parents' knowledge about and interest in the programme were clear from the fact that they came home and talked about the baby or stemmed from the fact that they were familiar with the ROE family from their own community:

Yeah because my mum knows the baby from church.

Child

My mum knows the mum and baby because the mum taught my sister when her teacher was having a baby.

Child

Some parents approached school staff to talk about the ROE programme, which the staff commented was positive:

I had one parent who had been on a behavioural management programme with the board previously and she's personally came and said she has seen a change in her child but she wanted to talk and I, you know, there's only so much then you can share, read the notice board, talk about the noticeboard with your child, get him to talk about it so the few parents that I have had come to me about it have been positive but again, they would have liked a little bit more continual.

Principal

One teacher mentioned that many fathers had relayed that their sons had been talking a lot about the baby at home:

Yeah. I've had more feedback from dads about how their sons have responded, to me, because it's the brotherhood all together. A lot of the mums have said, 'Oh, my daughter's really enjoying it', but again, I think, and I may be stereotyping myself, but I think girls are already imbibed with that maternal thing.

Teacher

Conclusion

Most comments from all the key stakeholders were positive regarding the benefits of the programme, and focused on improvements in the children. The feedback from the ROE instructors (including the four trust programme co-ordinators who were trained ROE instructors) on the delivery of the ROE programme was also very positive overall. The programme co-ordinators and instructors all reported few difficulties in bringing schools, mothers and babies or instructors on board. Overall, the interviews with the co-ordinators revealed a strong sense of engagement and support for the project from schools, teachers and instructors, which was echoed, without exception, in the interviews with the instructors and teachers. Instructors had been helped and supported by the class teachers. However, one issue that all interviewees appeared to agree on was parents' varied level of interest in and awareness of the programme.

Chapter 5 Discussion and conclusions

Introduction

This final chapter draws out several key conclusions from the overall findings presented in the previous chapters and considers their implications for future research and practice.

Key findings

Effectiveness of Roots of Empathy

Immediately post test (T1), there is evidence that the programme is achieving a positive effect in relation to both of the primary outcomes. In particular, participation in the ROE programme is associated with an increase in prosocial behaviour ($g = +0.20$; $p = 0.045$) and a decrease in difficult behaviour ($g = -0.16$; $p = 0.060$); however, as this latter finding is only approaching statistical significance, it needs to be interpreted with a degree of caution. This being said, both findings are consistent with the small number of other evaluations undertaken of the ROE programme, for which the pooled effect sizes are 0.13 and -0.18 , respectively.

One year after the end of the programme (T2), the findings from this present study indicate that the effects for prosocial behaviour have disappeared ($g < 0.01$), and this continues to be the case at the subsequent follow-up points (at 24 and 36 months). Interestingly, the size of the effect in relation to the decrease in difficult behaviour remains fairly stable across the 3 further years but, with the reduction in sample sizes because of attrition, these are no longer statistically significant (T2.: $g = -0.14$, $p = 0.22$; T3, $g = -0.13$, $p = 0.25$; and T4, $g = -0.14$, $p = 0.20$). The consistency of this difference over time would suggest that the programme may be having a sustainable effect in terms of reducing difficult behaviour. However, because these effects are no longer statistically significant, this finding must be treated with caution and requires further verification.

With regard to secondary outcomes, other than the programme-specific outcome of the understanding of infant feelings, no differences were found immediately post test (T1) or at any of the follow-up time points between the children who participated in the ROE programme and those in the control group. This finding also seems to be consistent with other existing evaluations of ROE. It is important to stress that this does not mean that the programme was not effective at improving these secondary outcomes. Rather, it simply demonstrates that the programme was *no more effective than the existing curriculum*, and especially PDMU, in relation to these outcomes.

The additional exploratory subgroup analyses did not provide any convincing evidence that the programme tended to have differential effects with regard to gender (boys or girls), socioeconomic background or number of siblings. As the programme was delivered consistently with high fidelity across intervention schools, it was not possible to assess whether or not variations in fidelity were associated with variations in outcomes achieved.

Cost-effectiveness of Roots of Empathy

Overall, it is estimated that the average cost of delivering ROE is £4057 per school and £175 per pupil. Against generally accepted national guidelines, the findings of this present study suggest that ROE is a cost-effective intervention. In particular, NICE suggests that interventions costing the NHS $< £20,000$ per one-unit increase in QALYs are cost-effective. It also suggests that those costing between £20,000 and £30,000 may be cost-effective. For the present evaluation, it was found that ROE had an 83.1% chance of being cost-effective at the £20,000 per QALY threshold and a 90.1% chance at the higher threshold of £30,000 per QALY.

Programme delivery and stakeholder perspectives

In relation to programme delivery, it is notable that the ROE programme was delivered with high fidelity, with all lessons being delivered in all of the intervention schools. This was seen as the result of the clearly defined structure of the programme and the strong training and ongoing support provided to ROE instructors in schools. The programme was also very well received overall and it was felt to include good resources and be linked in closely with the Northern Ireland curriculum, particularly the element on PDMU. Beyond this, five key issues emerged from the qualitative process evaluation.

1. Some believed that it would be beneficial if another teacher in the same school was the ROE instructor, which could allow stronger communication and planning between the instructor and the class teacher.
2. There was a perception that the delivery of the programme in the first year may have been a little more challenging, especially in those schools whose ROE instructor was not a teacher within that school. Relatedly, there was a belief among some that there would be enhanced opportunities in future years once the instructor and teacher had greater experience of and knowledge about the programme.
3. There was concern regarding the resources required to deliver the programme, especially if the instructor were to be one of the teachers in the school, and whether or not this would be sustainable in the longer term.
4. There was concern that the programme lasts for only 1 year and is not followed up in subsequent years. In addition, and relatedly, some thought that it would be worthwhile to build the key knowledge and skills among children at an earlier age and before the programme took place, with some mentioning the 'Seeds of Empathy' programme.
5. The relative lack of involvement of or engagement with parents in the programme and how this may have been arisen partly because of the emphasis on maintaining fidelity to the existing programme.

Limitations

This study is the largest evaluation of ROE to date, and one of only two studies that has used a randomised controlled trial design and has measured the longer-term effects of the programme up to 3 years from its completion. In one respect, given the large-scale nature of this field trial, the overall levels of retention have been good. Barring the seven schools that withdrew before the start of the trial, 76.3% of the pupils who were pre-tested in 2011 remained in the study until the final follow-up data sweep in 2015. However, this level of attrition still represents a limitation of the present trial.

Moreover, the initial engagement and subsequent attrition levels of parents have been notably lower. Initially, only just over half (58.0%) of the parents of children who were tested pre test completed and returned questionnaires. This reduced to 31.6% at the end of the study in 2015. The impact of this in relation to potentially introducing bias to the study was assessed by comparing the findings from the main analysis, based on observed data only, with those based on data sets that dealt with missing data through multiple imputation. Overall, these sensitivity analyses suggest that the levels of attrition found for the study did not appear to introduce any notable biases. However, the levels of attrition have still had two key negative impacts on the study:

- A reduction of the statistical power of the trial to detect the smaller effects of the programme. This was particularly notable in relation to the primary outcome associated with the reduction in difficult behaviour, where a consistent effect was found over the 3 years following the end of the programme but where the findings were not statistically significant.
- The lack of data on participant health and social resource use that were collected during the early stages of the trial, and then the high proportion of missing data, largely as a result of the low retention rate of parents in the study. This, in turn, represents a significant limitation to the cost-effectiveness analysis.

Furthermore, another limitation of the present design that should be noted is the fact that the measures used in relation to the two primary outcomes were based on teacher ratings of the children's behaviour and hence were unblinded to condition. It is possible that this introduced some bias, especially immediately post test (T1). However, it is unlikely to have introduced notable bias for subsequent time points given that the children had transferred to new classes and were assessed by different teachers, at a considerable time (at least 12 months) after the end of the programme.

Generalisability

The external validity of the trial was good, given the size of the sample and the broadly representative nature of the schools that took part compared with the population of schools as a whole. As such, we can be reasonably confident in generalising these results to other children who might participate in the ROE programme in Northern Ireland. The study design – a cluster randomised trial – was robust in detecting an unbiased effect of the intervention, and the trial was sufficiently powered to detect effects in the region of $g = 0.22$. The trial was registered with the ISRCTN and a full protocol was published before the data were collected and analysed. As noted earlier, there have been no significant deviations from this protocol. The randomisation was conducted by an independent organisation and at baseline the control and intervention groups were balanced with respect to observable characteristics.

There was some attrition of schools at the start of the trial, but there was no evidence of differential attrition between the intervention and control groups. There was also attrition at the pupil level, which has the potential to create unbalanced groups and introduce some bias into the results. The intervention was well manualised and fidelity was high; however, there was some evidence that teachers in the intervention group used ROE as a replacement activity for the PDMU content of the curriculum rather than as an addition. This may have unintentionally resulted in less 'clear blue water' between the control and intervention treatments, which could in turn diminish the magnitude of any differences observed between the groups on the measured outcomes. The outcome measures were carefully chosen to ensure that they were valid, reliable and suitable for use with children of this age group (aged 8–9 years).

Interpretation

There are five key themes to draw out from the findings reported above. First, the trial has provided strong and robust evidence that ROE had a positive impact on children's behaviours in the directions expected. More specifically, there is evidence that the programme enhanced children's prosocial behaviour and some evidence that it reduced difficult behaviour, above and beyond the typical effects associated with attending school. In relation to this, and especially with regard to interpreting the size of the effects found, it should be noted that these were associated with the delivery of the programme in its first year. As indicated in some of the feedback from the teachers reported above, it is possible that ROE will achieve stronger effects in future years and with new cohorts of children as the schools and instructors gain more experience and a greater understanding of the programme, and thus potentially embed key learning further into other aspects of the school curriculum. Unfortunately, this remains a hypothesis at this stage and is not something that is possible to address through the design of the current trial.

Second, and alongside providing evidence of the effectiveness of programme, the trial has provided clear evidence that although ROE was originally developed in Canada, it is possible to deliver it extremely effectively and with fidelity in a different country and cultural context, in this case Northern Ireland. Although there were some concerns raised regarding the structured and prescribed nature of the programme and its perceived 'Americanised' influences, these were not found to present significant obstacles to its successful delivery. Indeed, and on the whole, the programme was found to be very well received by teachers, principals, parents and children.

Third, although ROE was found to be broadly effective in relation to its primary behavioural outcomes, the trial found no evidence to support the theory of change hypothesised to underpin this. In particular, although the children who participated in ROE were found to have an increased awareness of the reasons for a baby crying, the trial found no evidence that this translated into an increased ability to recognise emotions, increased empathy or increased emotional regulation compared with children in the control group. It should be noted that the measures of emotional recognition (Emotion Recognition Questionnaire) and emotional regulation (CAMS) did have low reliability (Cronbach's alpha of 0.58 and 0.69, respectively) and this may have resulted in a loss of sensitivity to detect change. Despite this, the evidence from this current trial suggests that the ROE programme appears to have had a positive effect on children's behaviour without having any measurable impact on these hypothesised precursors. Indeed, the raw mean scores for empathy (as measured through the Interpersonal Reactivity Index) were found to *reduce* from pre test to post test for both groups of children. The mean scores pre test were 3.34 (SD 0.78) for the control group and 3.32 (SD 0.75) for the intervention group. However, and post test, these scores had reduced to 3.15 (SD 0.76) for the control group and 3.13 (SD 0.69) for the intervention group (effect size changes of $g = -0.25$ and $g = -0.26$, respectively).

It is not possible to conclude with certainty how ROE has achieved positive behavioural effects without associated increases in social and emotional outcomes. However, it is clear from the qualitative process evaluation that the ROE lessons were enjoyed by the children and that they did, progressively, help to encourage the development of a collective sense of concern and care for the baby, which may have resulted in a positive shift in the group norms (i.e. class norms) of prosocial behaviour. Peer groups play an important influential role in the development of children's behaviour and attitudes and are an important social context in which individual development takes place.⁸⁸⁻⁹⁰ Chang's⁹¹ social context model suggests that group norms – and the extent of their influence on behaviour and its consequences – will differ between contexts. For this reason, it is important to acknowledge the school context in which this study is located, where compliance and co-operation are viewed as highly desirable and so the behaviour and attitudes of prosocial groups are likely to be valued by the teacher. Chung-Hall and Chen⁸⁸ found that for children aged between 9 and 11 years, the positive sequelae of belonging to a high prosocial group included being liked by classmates, performing well in school and more positive perceptions of their own social and behavioural competence. They concluded that the relationship between social emotional functioning and (aggressive or prosocial) behaviour may well be as a result of group (aggressive or prosocial) behaviour and that the mechanism through which the peer group influences individual behaviour and attitudes may well be through norm-based group processes (e.g. social learning, mutual regulation, within group assimilation or group reputational effects). Thus, and according to Chang,⁹¹ 'the social norm of a behavior facilitates peer acceptance of the behavior'.

Fourth, the current ROE programme provides only limited opportunities to engage with parents. However, and as found through the process evaluation, there is significant interest among teachers and also some parents in greater parental involvement in the programme. The enhanced engagement of parents would present a challenge in relation to maintaining the fidelity of the programme, which is currently very high. However, given the ecological nature of children's development, including their social and emotional learning, it would make sense to explore how parental engagement could be enhanced as an explicit element of the programme.

Finally, although there are some encouraging signs that ROE may achieve sustainable effects in reducing difficult behaviour, the findings of the current trial are not as positive in relation to the sustainability of initial gains in prosocial behaviour. With this in mind, and reflecting the views of teachers, there is a need to consider how this 1-year programme can become part of a longer-term curriculum and strategic approach in schools to the development of pupils' social and emotional learning. In this respect, further work would be beneficial in terms of developing a more holistic and progressive curriculum that seeks to use evidence-based programmes such as ROE, but in a way that is able to sustain and build on the short-term gains found in a developmentally appropriate way.

Acknowledgements

The research team are indebted to everyone who took part in this study, including all of the school principals, teachers, pupils, parents, ROE co-ordinators, ROE instructors and, of course, volunteer mothers and babies. Thank you.

We are very grateful to the members of our Trial Steering Committee for their insights, advice and guidance throughout the study. In particular, we would like to express our thanks to Mary Black, Professor Steve Higgins, Maurice Meehan and Dr Ben Styles.

We would like to thank Dr Harry Rafferty (School of Psychology, Queen's University Belfast) for his input during the initial stages of the study, particularly in relation to the design of the study and the development of the theory of change tested by the trial.

We would also like to express our gratitude to the Public Health Agency Research and Development Office for its continued support and advice throughout the study.

Contributions of authors

Professor Paul Connolly (Professor, Education) directed the study and had overall responsibility for the design, delivery, analysis, interpretation, reporting and management of all aspects of the trial, process evaluation and cost-effectiveness evaluation.

Dr Sarah Miller (Lecturer, Education) was the trial manager and co-ordinated the day-to-day running of the trial and process evaluation. She contributed to all aspects of the design, delivery, quantitative data preparation, data analysis, interpretation, reporting and management of the trial and process evaluation.

Professor Frank Kee (Professor, Public Health) contributed to the design of the trial and the interpretation of the results at all stages of the study.

Dr Seaneen Sloan (Research Fellow, Education) co-ordinated the quantitative (parent, teacher and child) data collection across all schools, co-ordinated data entry and contributed to the preparation of interim reports.

Ms Aideen Gildea (Research Assistant) conducted the process evaluation, which included data collection, data analysis, interpretation and write-up. She also contributed to the co-ordination of trial data collection and data entry.

Dr Emma McIntosh (Reader, Health Economics) oversaw and conducted the cost-effectiveness evaluation, including design, data analysis, interpretation and reporting.

Ms Nicole Boyer (Research Assistant, Health Economics) contributed to all aspects of the cost-effectiveness evaluation, including design, data analysis, interpretation and reporting.

Professor Martin Bland (Emeritus Professor) contributed to the design of the trial – in particular the statistical analysis section – and the interpretation of the results at all stages of the study.

Data sharing statement

All available data, once fully anonymised and prepared for sharing (anticipated March 2018), will be available from the corresponding author.

References

1. Petrides KV, Frederickson N, Furnham A. The role of trait emotional intelligence in academic performance and deviant behavior at school. *Pers Individ Dif* 2004;**36**:277–93. [https://doi.org/10.1016/S0191-8869\(03\)00084-9](https://doi.org/10.1016/S0191-8869(03)00084-9)
2. Ciarrochi J, Deane FP, Anderson S. Emotional intelligence moderates the relationship between stress and mental health. *Pers Individ Dif* 2002;**32**:197–209. [https://doi.org/10.1016/S0191-8869\(01\)00012-5](https://doi.org/10.1016/S0191-8869(01)00012-5)
3. Lemerise EA, Arsenio WF. An integrated model of emotion processes and cognition in social information processing. *Child Dev* 2000;**71**:107–18. <https://doi.org/10.1111/1467-8624.00124>
4. Leppänen JM, Hietanen JK. Emotion recognition and social adjustment in school-aged girls and boys. *Scand J Psychol* 2001;**42**:429–35. <https://doi.org/10.1111/1467-9450.00255>
5. Mostow AJ, Izard CE, Fine S, Trentacosta CJ. Modeling emotional, cognitive, and behavioral predictors of peer acceptance. *Child Dev* 2002;**73**:1775–87. <https://doi.org/10.1111/1467-8624.00505>
6. Nagin D, Tremblay RE. Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Dev* 1999;**70**:1181–96. <https://doi.org/10.1111/1467-8624.00086>
7. Broidy LM, Nagin DS, Tremblay RE, Bates JE, Brame B, Dodge KA, *et al.* Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: a six-site, cross-national study. *Dev Psychol* 2003;**39**:222–45. <https://doi.org/10.1037/0012-1649.39.2.222>
8. NICE. *Social and Emotional Wellbeing in Primary Education*. Public health guideline 12. London: NICE; 2008. URL: www.nice.org.uk/guidance/ph12 (accessed 23 March 2015).
9. Marmot M. *Fair Society, Healthy Lives: The Marmot Review. Strategic Review of Health Inequalities in England Post-2010*. Executive Summary. London: Department of Health; 2010.
10. Goodman A, Joshi H, Nasim B, Tyler C. *Social and Emotional Skills in Childhood and their Long-term Effects on Adult Life*. London: University College London; 2015.
11. Greenberg MT. School-based prevention: current status and future challenges. *Effective Education* 2010;**2**:27–52. <https://doi.org/10.1080/19415531003616862>
12. Greenberg MT, Weissberg RP, Utne O'Brien M, Zins JE, Fredericks L, Resnik H, Elias MJ. Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *Am Psychol* 2003;**58**:466–74. <https://doi.org/10.1037/0003-066X.58.6-7.466>
13. Browne G, Gafni A, Roberts J, Byrne C, Majumdar B. Effective/efficient mental health programs for school-age children: a synthesis of reviews. *Soc Sci Med* 2004;**58**:1367–84. [https://doi.org/10.1016/S0277-9536\(03\)00332-0](https://doi.org/10.1016/S0277-9536(03)00332-0)
14. Payton J, Weissberg RP, Durlak JA, Dymnicki AB, Taylor RD, Schellinger KB, *et al.* *The Positive Impact of Social and Emotional Learning for Kindergarten to Eighth-Grade Students: Findings from Three Scientific Reviews*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning; 2008.
15. Sutton PW, Love JG, Bell J, Christie E, Mayrhofer A, Millman Y, *et al.* *The Emotional Well-Being of Young People: A Review of the Literature*. Aberdeen: Robert Gordon University; 2005.

16. Wilson SJ, Lipsey MW. School-based interventions for aggressive and disruptive behavior: update of a meta-analysis. *Am J Prev Med* 2007;**33**(Suppl. 2):130–43. <https://doi.org/10.1016/j.amepre.2007.04.011>
17. Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, Schellinger KB. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev* 2011;**82**:405–32. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>
18. Adi Y, Killoran A, Janmohamed K, Stewart-Brown S. *Systematic Review of the Effectiveness of Interventions to Promote Mental Wellbeing in Children in Primary Education*. Report 1: Universal Approaches (Non-Violence Related Outcomes). London: NICE; 2007.
19. Clarke AM, Morreale S, Field CA, Hussein Y, Barry MM. *What Works in Enhancing Social and Emotional Skills Development During Childhood and Adolescence*. Galway: WHO Collaborating Centre for Health Promotion Research, National University of Ireland, Galway; 2015.
20. Santos RG, Chartier MJ, Whalen JC, Chateau D, Boyd L. Effectiveness of school-based violence prevention for children and youth: cluster randomized field trial of the Roots of Empathy program with replication and three-year follow-up. *Healthc Q* 2011;**14**:80–91. <https://doi.org/10.12927/hcq.2011.22367>
21. Schonert-Reichl K, Smith V, Zaidman-Zait A. *Effectiveness of the Roots of Empathy Program in Fostering the Social-Emotional Development of Primary Grade Children*. Vancouver, BC: University of British Columbia; 2002.
22. Smith V. *Roots of Empathy Whole Schools Evaluation Report: Examining Variability of Program Implementation of the Roots of Empathy*. Edmonton, AB: University of Alberta; 2008.
23. Schonert-Reichl KA, Smith V, Zaidman-Zait A, Hertzman C. Promoting children's prosocial behaviours in school: impact of the 'Roots of Empathy' program on the social and emotional competence of school-aged children. *School Mental Health* 2012;**4**:1–12. <https://doi.org/10.1007/s12310-011-9064-7>
24. MacDonald A, Bell P, McLafferty M, McCorkell L, Walker I, Smith V, et al. *Evaluation of the Roots of Empathy Programme by North Lanarkshire Psychological Service*. Airdrie: North Lanarkshire Psychological Service Research; 2013.
25. Wrigley J, Makara K, Elliot D. *Evaluation of Roots of Empathy in Scotland 2014-15: Final Report for Action for Children*. 2016. URL: www.actionforchildren.org.uk/media/6048/final_report_v12f.pdf (accessed 11 May 2016).
26. Kendall G, Schonert-Reichl K, Smith V, Jacoby P, Austin R, Stanley F, Hertzman C. *The Evaluation of Roots of Empathy in Western Australian Schools 2005*. Perth, WA: Telethon Institute for Child Health Research; 2006.
27. Rolheiser C, Wallace D. *The Roots of Empathy Program as a Strategy for Increasing Social and Emotional Learning. Program Evaluation Final Report*. Toronto, ON: ROE; 2005.
28. da Costa JL, Shultz L. *Reducing Bullying Behaviour: City Schools' Experiences Adapting Roots Of Empathy to their Contexts*. Edmonton, AB: University of Alberta; 2006.
29. Cain G, Carnellor Y. Roots of Empathy: a research study on its impact on teachers in Western Australia. *J Stud Well* 2008;**2**:52–73.
30. Connolly P, Rafferty H, Maguire C, Miller S, McIntosh E, Kee F, et al. *A Cluster Randomised Controlled Trial Evaluation and Cost-Effectiveness Analysis of the Roots of Empathy Schools-Based Programme for Improving Social and Emotional Wellbeing Outcomes Among 8–9 Year Olds in Northern Ireland*. Protocol. National Institute for Health Research Public Health Research Programme (PHR 10/3006/02); 2001. URL: www.nets.nihr.ac.uk/projects/phr/10300602 (accessed 7 January 2018).

31. Baron-Cohen S. *The Essential Difference*. London: Allen Lane; 2003.
32. Piaget J. *The Origins of Intelligence in Children*. New York, NY: International University Press; 1952.
33. Piaget J. *The Construction of Reality in the Child*. New York, NY: Basic Books; 1952.
34. Bowlby J. *Attachment*. 2nd edn. New York, NY: Basic Books; 1983.
35. Ainsworth M, Blehar M, Waters E, Wall S. *Patterns of Attachment*. Hillsdale, NJ: Erlbaum; 1978.
36. Fonagy P, Gergely G, Jurist E, Target M. *Affect Regulation, Entalization, and the Development of the Self*. London: Karnac Books; 2002.
37. Allen G, Fonagy P. *The Handbook of Mentalization-based Treatment*. Chichester: Wiley; 2006. <https://doi.org/10.1002/9780470712986>
38. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 1997;**38**:581–6. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
39. Ladd GW, Profilet SM. The Child Behaviour Scale: a teacher-report measure of young children's aggressive, withdrawn and prosocial behaviours. *Dev Psychol* 1996;**32**:1008–24. <https://doi.org/10.1037/0012-1649.32.6.1008>
40. Ribordy SC, Camras LA, Stefani R, Spaccarelli S. Vignettes for emotion recognition research and affective therapy with children. *Journal Clin Child Psychol* 1988;**17**:322–5. https://doi.org/10.1207/s15374424jccp1704_4
41. Davis MH. Measuring individuals differences in empathy: evidence for a multidimensional approach. *J Pers Soc Psychol* 1983;**44**:113–26. <https://doi.org/10.1037/0022-3514.44.1.113>
42. Litvack-Miller W, McDougall D, Romney DM. The structure of empathy during middle childhood and its relationship to prosocial behavior. *Genet Soc Gen Psychol Monogr* 1997;**123**:303–24.
43. Garton AF, Gringart E. The development of a scale to measure empathy in 8- and 9-year old children. *Aust J Educ Devl Psychol* 2005;**5**:17–25.
44. Zeman J, Shipman K, Penza-Clyve S. Development and initial validation of the Children's Sadness Management Scale. *J Nonverbal Behav* 2001;**25**:187–205. <https://doi.org/10.1023/A:1010623226626>
45. Olweus D. *The Revised Olweus BullyVictim Questionnaire*. Mimeo. Bergen: Research Center for Health Promotion, University of Bergen; 1996.
46. Stevens KJ. Valuation of the Child Health Utility 9D Index. *PharmacoEconomics* 2012;**30**:729–47. <https://doi.org/10.2165/11599120-000000000-00000>
47. Northern Ireland Statistics and Research Agency. *Northern Ireland Multiple Deprivation Measure 2010*. Belfast: Northern Ireland Statistics and Research Agency; 2010.
48. Northern Ireland Assembly. *Free School Meal Entitlement as a Measure of Deprivation*. Research and Library Service Briefing Paper. Belfast: Northern Ireland Assembly; 2010.
49. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;**3**:77–101. <https://doi.org/10.1191/1478088706qp063oa>
50. Belli PC, Bustreo F, Preker A. Investing in children's health: what are the economic benefits? *Bull World Health Organ* 2005;**83**:777–84.
51. Heckman JJ, Masterov D. *The Productivity Argument for Investing in Young Children*. Chicago, IL: University of Chicago; 2007. <https://doi.org/10.3386/w13016>
52. NICE. *Methods for the Development of NICE Public Health Guidance (Third Edition)*. London: NICE; 2014. URL: www.nice.org.uk/process/pmg4/chapter/introduction (accessed 7 January 2018).

53. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;**30**:377–99. <https://doi.org/10.1002/sim.4067>
54. Eldridge S, Kerry S. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. West Sussex: Jon Wiley & Sons, Ltd; 2012. <https://doi.org/10.1002/9781119966241>
55. NICE. *Guide to the Methods of Technology Appraisal 2013*. London: NICE; 2013. URL: www.nice.org.uk/process/pmg9/chapter/1-foreword (accessed 12 August 2016).
56. Royal College of Nursing. *NHS Agenda for Change Pay Scales – 2011/2012*. URL: www.rcn.org.uk/_data/assets/pdf_file/0005/372992/004106.pdf (accessed 13 May 2016).
57. Organisation for Economic Cooperation and Development. *Purchasing Power Parities (PPP) (Indicator)*. URL: <http://dx.doi.org/10.1787/1290ee5a-en> (accessed 2 September 2015).
58. Hale J, Cohen D, Ludbrook A, Phillips C, Duffy M, Parry-Langdon N. *Moving from Evaluation into Economic Evaluation: A Health Economics Manual for Programmes to Improve Health and Well-being*. URL: <http://orb.essex.ac.uk/hs/hs915/health%20economic%20evaluation%20manual.pdf> (accessed 21 April 2016).
59. Curtis L. *Unit Costs of Health and Social Care 2014*. Canterbury: Personal Social Services Research Unit, University of Kent; 2014.
60. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd edn. Oxford: Oxford University Press; 2005.
61. Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. *Qual Health Res* 2010;**20**:340–51. <https://doi.org/10.1177/1049732309358328>
62. Ratcliffe J, Flynn T, Terlich F, Stevens K, Brazier J, Sawyer M. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. *PharmacoEconomics* 2012;**30**:713–27. <http://dx.doi.org/10.2165/11597900-000000000-00000>
63. Faria R, Gomes M, Epstein D, White IR. A Guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *PharmacoEconomics* 2014;**32**:1157–70. <http://dx.doi.org/10.1007/s40273-014-0193-3>
64. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Hoboken, NJ: Wiley & Sons, Inc.; 2002. <https://doi.org/10.1002/9781119013563>
65. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing . . . presumed at random: cost-analysis of incomplete data. *Health Econ* 2003;**12**:377–92. <https://doi.org/10.1002/hec.766>
66. Glick H, Doshi J, Sonnad S, Polsky D. *Economic Evaluation in Clinical Trials*. Oxford: Oxford University Press; 2007.
67. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. *J Health Serv Res Policy* 2004;**9**:197–204. <https://doi.org/10.1258/1355819042250249>
68. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ* 2005;**14**:487–96. <https://doi.org/10.1002/hec.944>
69. Bachmann MO, Fairall L, Clark A, Mugford M. Methods for analyzing cost effectiveness data from cluster randomized trials. *Cost Eff Resour Alloc* 2007;**5**:12. <https://doi.org/10.1186/1478-7547-5-12>
70. Donner A, Bikett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol* 1981;**114**:906–14. <https://doi.org/10.1093/oxfordjournals.aje.a113261>

71. Henderson M, Jackson C, Bond L, Wilson P, Elliot L, Levin K. *Social and Emotional Education and Development (SEED): A Stratified, Cluster Randomised Trial of a Multi-component Primary School Intervention that follows the Pupils' Transition into Secondary School*. URL: www.nets.nihr.ac.uk/_data/assets/pdf_file/0020/81713/PRO-10-3006-13.pdf (accessed 2 July 2014).
72. Gomes M, Grieve R, Nixon R, Edmunds WJ. Statistical methods for cost-effectiveness analyses that use data from cluster randomized trials: a systematic review and checklist for critical appraisal. *Med Decis Making* 2012;**32**:209–20. <http://dx.doi.org/10.1177/0272989x11407341>
73. Gomes M, Ng ES, Grieve R, Nixon R, Carpenter J, Thompson SG. Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Med Decis Making* 2012;**32**:350–61. <http://dx.doi.org/10.1177/0272989x11418372>
74. Connolly P, Miller S, Mooney J, Sloan S, Hanratty J. *Universal School-Based Programmes for Improving Social and Emotional Outcomes in Children Aged 3–11: A Systematic Review and Meta-Analysis*. Title Registration. The Campbell Collaboration Library. URL: www.campbellcollaboration.org/library/universal-school-based-programmes-social-emotional-outcomes-ages-3-11.html (accessed 5 March 2018).
75. Meltzer H, Gatward R, Goodman R, Ford F. *Mental Health of Children and Adolescents in Great Britain*. London: The Stationery Office; 2000. <https://doi.org/10.1037/e622732007-001>
76. Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz DS. The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *J Am Acad Child Adolesc Psychiatry* 2005;**44**:557–64. <https://doi.org/10.1097/01.chi.0000159157.57075.c8>
77. Williams J, Greene S, Doyle E, Harris E, Layte R, McCoy S, et al. *Growing Up in Ireland: National Longitudinal Study of Children – The Lives of 9-Year-Olds. Report 1*. Dublin: The Stationary Office; 2009.
78. Wiles NJ, Northstone K, Emmett P, Lewis G. 'Junk food' diet and childhood behavioural problems: results from the ALSPAC cohort. *Eur J Clin Nutr* 2009;**63**:491–8. <https://doi.org/10.1038/sj.ejcn.1602967>
79. Department of Health. *NHS Reference Costs 2013–14*. 2014. URL: www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014 (accessed 7 January 2018).
80. Business Services Organisation. *MOS Library*. URL: www.hscbusiness.hscni.net/services/2377.htm (accessed 7 January 2018).
81. Robertson W, Fleming J, Kamal A, Hamborg T, Khan KA, Griffiths F, et al. Randomised controlled trial evaluating the effectiveness and cost-effectiveness of 'Families for Health', a family-based childhood obesity treatment intervention delivered in a community setting for ages 6 to 11 years. *Health Technol Assess* 2017;**21**(1). <http://dx.doi.org/10.3310/hta21010>
82. Julious SA, Horspool MJ, Davis S, Bradburn M, Norman P, Shephard N, et al. PLEASANT: Preventing and Lessening Exacerbations of Asthma in School-age children Associated with a New Term - a cluster randomised controlled trial and economic evaluation. *Health Technol Assess* 2016;**20**(93). <http://dx.doi.org/10.3310/hta20930>
83. Apajasalo M, Sintonen H, Holmberg C, Sinkkonen J, Aalberg V, Pihko H, et al. Quality of life in early adolescence: a sixteen-dimensional health-related measure (16D). *Qual Life Res* 1996;**5**:205–11. <https://doi.org/10.1007/BF00434742>
84. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res* 2010;**19**:875–86. <https://doi.org/10.1007/s11136-010-9648-y>

85. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care* 1996;**34**:702–22. <https://doi.org/10.1097/00005650-199607000-00004>
86. Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D utility scores from the PedsQL™ generic core scales. *Pharmacoeconomics* 2014;**32**:693–706. <https://doi.org/10.1007/s40273-014-0153-y>
87. Siegel JE. Cost-effectiveness analysis and nursing research – is there a fit? *Image J Nurs Sch* 1998;**30**:221–2. <https://doi.org/10.1111/j.1547-5069.1998.tb01295.x>
88. Chung-Hall J, Chen X. Aggressive and prosocial peer group functioning: effects on children's social, school and psychological adjustment. *Soc Dev* 2010;**19**:659–80. <https://doi.org/10.1111/j.1467-9507.2009.00556.x>
89. Cairns R, Cairns B. *Lifelines and Risks: Pathways of Youth in Our Time*. New York, NY: Cambridge University Press; 1994.
90. Rubin KH, Bukowski WM, Parker JG. Peer Interactions, Relationships and Groups. In Eisenberg N, William D, Lerner RN, editors. *Handbook of Child Psychology: Volume 3 – Social, Emotional and Personality Development*. 6th edn. Hoboken, NJ: John Wiley & Sons Inc.; 2006. pp. 571–645.
91. Chang L. The role of classroom norms in contextualizing the relations of children's social behaviors to peer acceptance. *Dev Psychol* 2004;**40**:691–702. <https://doi.org/10.1037/0012-1649.40.5.691>
92. Bayrami L. *Roots of Empathy: A Brief Summary of Research*. Roots of Empathy; 2016. URL: www.rootsofempathy.org/wp-content/uploads/2016/01/ROE-Research-Summary_Dec-2016.pdf (accessed 5 March 2018).
93. Gordon M. *Changing the World Child by Child*. Toronto, ON: Thomas Allen Publishers; 2005.
94. Schonert-Reichl KA, Smith V, Hertzman C. *Promoting Emotional Competence in School-aged Children: An Experimental Trial of the Roots of Empathy Programme*. Poster presented at Biennial Meeting of the Society for Research in Child Development, Boston, MA, USA, 2007.
95. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–60. <https://doi.org/10.1136/bmj.327.7414.557>

Appendix 1 Meta-analysis of existing evaluations of Roots of Empathy

Introduction

This meta-analysis seeks to combine the results of the available and eligible evaluations of ROE that have been conducted since the first evaluation in 2000. The primary research question for the analysis is:

- Does ROE improve prosocial behaviour and decrease aggressive behaviour in primary school-aged children?

The analysis has two secondary research questions:

1. Does ROE improve empathy and emotional regulation in primary school-aged children?
2. Is the effect of ROE in improving prosocial behaviour and decreasing aggressive behaviour sustained 3 years post test?

Characteristics of studies included in the analysis

To date, it has been possible to locate the full reports for seven previous quantitative evaluations of ROE (*Table 30*).

Inclusion and exclusion criteria

As there are relatively few evaluations of ROE, it was decided that the inclusion criteria would not be as stringent as those for typical meta-analyses. Studies were considered eligible if they employed an experimental or quasi-experimental design, quantitatively measured (at least) teacher-rated prosocial and aggressive behaviour, and collected outcome data both pre and post test. Studies were excluded if they were not an effectiveness evaluation or if they collected only qualitative data. The characteristics of included studies, including the current trial, are described in *Table 30*, and excluded studies are described in *Characteristics of excluded studies*.

Study design and location

Apart from the current trial, there is only one other randomised controlled trial of the programme³⁶ and – similar to the current trial – this evaluation also followed up participants after the immediately post-test data collection for a further 3 years. The remaining six studies employed a quasi-experimental design with pre- and post-test data collection only (no follow-up). Four studies were conducted in Canada,^{36,45–47} two were conducted in Scotland^{48,49} and one was conducted in Australia.⁵⁰

Sample

In total, 4140 primary school-aged children from 145 primary schools took part in the seven studies. The sample sizes ranged between 132 and 785 children, with an average sample size of 591.

Outcomes

All of the evaluations to date – including the current trial – have measured teacher-rated prosocial and aggressive behaviour using valid and reliable instruments. A range of other teacher- and child-rated outcomes were also measured; however, this meta-analysis focuses only on synthesising the effects for the most commonly measured outcomes:

- teacher-rated **prosocial behaviour** immediately post test (all seven previous studies)
- teacher-rated **aggressive behaviour** immediately post test (all seven previous studies)
- child-reported **empathy** immediately post test (five studies)
- child-reported **emotional regulation** immediately post test (two studies).

TABLE 30 Characteristics of previous quantitative evaluations included in the meta-analysis

Study ID	Methods	Participants	Intervention	Outcomes ^a	Analysis
Kendall <i>et al.</i> , 2006 ²⁶	Quasi-experimental, pre- and post-test design	27 primary schools, 34 classes and 648 pupils (6–7 years old) in Western Australia. A total of 85% of the sample was Caucasian and the final sample (after attrition) comprised 240 boys and 193 girls	ROE was delivered to the intervention group (17 schools, 22 classes and 427 children) and the control group (10 schools, 12 classrooms and 221 children) did not receive the programme. The study was conducted between 2005 and 2006, and there was no follow-up	Teacher-rated social behaviour (prosocial behaviour and aggression, empathy, personal distress, peer acceptance and general social skills) and emotional regulation Child-rated empathy, relationships, friendships, feelings about school and knowledge of infant development	Multivariate mixed models, taking clustering into account. Regression coefficients and 95% CIs were reported (all variables were standardised before analysis). Reported estimated effects are adjusted for clustering, pre-test scores and other relevant covariates
Macdonald <i>et al.</i> , 2013 ²⁴	Quasi-experimental, pre- and post-test design	34 primary schools, 37 classes and 785 children (6–10 years old) in Scotland. The sample comprised 384 boys and 401 girls	ROE was delivered to the intervention group (17 schools, 19 classes and 419 children). The control group (17 schools, 18 classes and 366 children) did not receive the programme. The study was conducted between 2011 and 2012, and there was no follow-up	Teacher-rated prosocial and difficult behaviour and emotional regulation Child-rated empathy, emotional regulation, prosocial behaviour, well-being and class climate	T-tests comparing change scores for intervention and control groups. Clustering was not taken into account. Effect sizes (SMD) were estimated (by SM) on the basis of post-test-only means, SDs and sample sizes
Santos <i>et al.</i> , 2011 ²⁰	Cluster randomised controlled trial	Eight school divisions, 27 elementary schools, 36 classrooms, 760 children (5–13 years old) in Canada. The unit of randomisation was school division. The numbers of boys and girls were not reported	ROE was delivered to the intervention group (five school divisions, 17 schools, 24 classrooms and 445 pupils) and the control group was a waiting list condition (three school divisions, 10 schools, 12 classrooms and 315 pupils). The trial was conducted between 2002 and 2006 and included a 3-year follow-up	Teacher-rated prosocial behaviour and aggression Child-rated prosocial behaviour and aggression	Multilevel modelling. Adjusted effect sizes and 95% CIs were reported. Reported estimated effects were adjusted for clustering, pre-test scores and other covariates

Study ID	Methods	Participants	Intervention	Outcomes ^a	Analysis
Schonert-Reichl <i>et al.</i> , 2002 ²¹	Quasi-experimental, pre- and post-test design	Five elementary schools, 10 classes and 132 children (6–7 years old) in Canada. The majority of the sample reported that English was their first language. The sample comprised 78 boys and 54 girls	ROE was delivered to the intervention group (five classes and 74 children) and the control group (five classes and 58 children) did not receive the programme. Both intervention and control classes were drawn from each participating school. The date of the study was 2000–1 and there was no follow-up	Teacher-rated social behaviour, including prosocial and aggressive behaviour Child-rated social and emotional understanding and understanding of infant crying	ANOVA comparing change scores for intervention and control groups. Clustering was not taken into account. Effect sizes (SMD) were estimated (by SJM) on the basis of the reported <i>F</i> -test statistic and sample size, taking into account pre-test scores and other relevant covariates
Schonert-Reichl <i>et al.</i> , 2012 ²³	Quasi-experimental, pre- and post-test design	28 elementary schools, 28 classes and 638 children (10–11 years old) in Canada. The final sample (<i>n</i> = 585) comprised 305 males and 280 females and 60% reported that English was their first language	ROE was delivered to the intervention group (14 schools, 14 classes and 306 children) and the control group (14 schools, 14 classes and 279 children) did not receive the programme. The study was conducted between 2001 and 2002 and there was no follow-up	Teacher-rated prosocial and aggressive behaviour Child-rated empathy, prosocial and anti-social behaviour and understanding of infant crying	ANOVA comparing change scores for intervention and control groups. Clustering was not taken into account. Effect sizes (SMD) were estimated (by SJM) on the basis of the reported of statistic and sample sizes, which controlled for pre-test scores and other relevant covariates
Smith 2008 ²²	Quasi-experimental, pre- and post-test design	Seven schools, 30 classrooms, 569 children (9–12 years old) in Canada. The sample included 271 males and 298 females, and 95% were Caucasian	ROE was delivered to the intervention group (four schools, 14 classes and 280 children). The control group (three schools, 16 classes and 289 children) was a waiting list condition. The study was conducted between 2006 and 2007, and there was no follow-up	Teacher-rated prosocial and aggressive behaviour Child-rated emotional regulation, interpersonal understanding, parenting efficacy and class climate; peer-rated prosocial and aggressive behaviour	Adjusted post-test means and SDs are reported that controlled for pre-test scores and other relevant covariates. Clustering was not taken into account. Effect sizes (SMD) were estimated (by SJM) on the basis of the reported <i>F</i> -statistic and sample sizes

continued

TABLE 30 Characteristics of previous quantitative evaluations included in the meta-analysis (*continued*)

Study ID	Methods	Participants	Intervention	Outcomes ^a	Analysis
Wrigley <i>et al.</i> , 2016 ²⁵	Quasi-experimental, pre-test and post-test design	17 primary schools, 31 classes and 695 children (7–8 years old) in Scotland. The final sample ($n = 661$) comprised 357 males and 304 females	ROE was delivered to the intervention group ($n = 352$) and the control group ($n = 309$) did not receive the programme. It is not clear how many classes were in the intervention and control groups. Both intervention and control classes were drawn from within each participating school. The study was conducted between 2014 and 2015 and there was no follow-up	Teacher-rated prosocial behaviour, aggressive behaviour and empathy Child-rated empathy	Raw pre- and post-test means and SDs are reported. Clustering was not taken into account. Effect sizes (SMD) were estimated (by SJM) on the basis of the reported <i>F</i> -statistic and sample sizes

ANOVA, analysis of variance; ID, identification.

^a Outcomes in bold denote those used in the meta-analysis.

In addition, the one study that measured outcomes for 3 years after the immediately post-test time point³⁶ will be summarised to provide an initial insight into the potential longer-term impacts of the programme.

Analysis

All of the studies were conducted in the school and classroom setting; however, only two of the seven studies took account of the clustered nature of the data in their estimation of the effect size and associated standard error. Ignoring clustering in the analysis can lead to artificially small standard errors and, consequently, smaller *p*-values. The implications of this are twofold:

1. Studies with smaller standard errors are deemed to be more precise estimates of the effect and are given more weight in the analysis (even though they might be methodologically weak), thereby inflating their relative contribution to the calculation of the weighted SMD.
2. A small standard error is more likely to yield a statistically significant result and lead to the incorrect rejection of the null hypothesis [i.e. concluding there is an effect when in fact there is not (a false positive)].

Publication status

Only two of the studies included in the analysis have been published in peer-reviewed journals.^{36,47} The remaining studies are unpublished reports and are considered grey literature.

Risk of bias

The risk of bias with respect to the methodological quality of each included study was assessed by the authors. It includes judgements about each of the following criteria:

- random sequence generation (selection bias)
- blinding of outcome assessment (detection bias)
- incomplete outcome data (attrition bias)
- other bias [defined in this instance as taking (or not) the clustered nature of the data into account at the analysis stage].

Table 31 shows the judgements about the risk of bias for each included study.

TABLE 31 Risk-of-bias summary: authors' judgements about each risk-of-bias item for each included study

Study	Risk-of-bias item			
	Selection	Detection	Attrition	Other
Kendall <i>et al.</i> , 2006 ²⁶	X	?	X	✓
MacDonald <i>et al.</i> , 2006 ²⁴	X	?	?	X
Santos <i>et al.</i> , 2011 ²⁰	✓	?	X	✓
Schonert-Reichl <i>et al.</i> , 2002 ²¹	X	?	?	X
Schonert-Reichl <i>et al.</i> , 2012 ²³	X	?	?	X
Smith 2008 ²²	X	?	?	X
Wrigley <i>et al.</i> , 2016 ²⁵	X	?	?	X

✓, low risk of bias; X, high risk of bias; ?, unclear risk of bias.

Characteristics of excluded studies

Although much research has been conducted on the ROE programme, not all of it can, or should, be included in the current synthesis. Studies that have employed non-experimental methodologies, collected qualitative data or answered research questions that are not specifically related to effectiveness are not eligible for inclusion. This does not mean that these studies are not valuable or important in and of themselves.

Three evaluations were excluded from the current synthesis and meta-analysis.

1. Rolheiser and Wallace²⁷ evaluated the extent to which the methods and approaches of the ROE curriculum aligned with CASEL (Collaborative for Academic, Social and Emotional Learning) guidelines and the teaching of character education more generally. In addition, the study employed a case study approach to inform best practice standards and, as such, this was not an effectiveness evaluation in which outcomes were measured pre and post intervention. Consequently, the data generated by this study are not eligible to be included in the meta-analysis.
2. The study by da Costa and Shultz²⁸ was conducted in two inner-city schools in Alberta, Canada. It evaluated an adapted version of the ROE programme that included the following changes to delivery: the programme was implemented in every classroom in the school; teachers delivered the programme; babies that were outside the recommended developmental age range were recruited; and some books were centralised so that they were available for all classes to access. Quantitative pre-and post-data were collected for two outcomes: empathy and student misconduct referrals. However, these data were collected only for the intervention children and the study did not include a control group (or attempt to measure the counterfactual). For this reason, the study was excluded from the current synthesis.
3. Cain and Carnellor²⁹ adopted a phenomenological approach to explore the impact of ROE on teachers in Western Australia. No quantitative data on outcomes were collected and for this reason the study was excluded from the current synthesis.

Other evaluations of Roots of Empathy

There were a number of additional evaluations that were identified during this process, but it has not been possible to locate a full report (or the data) for these. Authors were contacted directly, but at the time of writing no reply had been received. These studies include:

- A rural–urban evaluation conducted between 2002 and 2003. Participating children ($n = 419$) were in Grades 4–7 and the study was conducted by the University of British Columbia. The study is referenced in two places.^{92,93}
- A randomised controlled trial and 2-year follow-up conducted between 2003 and 2007. Participating children ($n = 456$) were in Grades 4–7 (20 classrooms) and the study was conducted in Vancouver by the University of British Columbia. The study is referenced in two places^{92,93} and it was presented at the Biennial Meeting of the Society for Research in Child Development, Boston, in 2007.⁹⁴
- Multiyear evaluation, 2008. It is not clear whether this is a separate evaluation or a summary of the evaluation work up until 2008. It is referenced in the ROE research summary.⁹²
- Alberta evaluation (year unknown). Participating children ($n = 221$) were in Grade 1 (14 classrooms). This evaluation also included academic outcomes. It is referenced in the ROE research summary.⁹²
- A 2-year study in New Zealand, with results to be reported in 2009. No other information is available and the study is referenced in the ROE research summary.⁹²
- An evaluation in the Isle of Man was conducted between 2009 and 2010 by the University of British Columbia and the Institute of Psychiatry, King's College London. Participating children ($n = 301$) were in Year 2 (19 classrooms). A poster depicting a summary of the findings for this study exists but the full report is not available.

It may be the case that some of these references are linked to studies already included in the current analysis but, without more information, it is not possible to be certain.

Main findings of meta-analysis

A random-effects meta-analysis was undertaken for each outcome specified using inverse variance. The measure of effect used is the SMD.

Prosocial behaviour

The results of the meta-analysis for teacher-rated prosocial behaviour are reported in *Figure 6*. All seven studies were included in the analysis and the overall SMD was 0.13 (95% CI 0.06 to 0.19) in favour of the intervention group. This means that ROE improves prosocial behaviour by, on average, 0.13 of a SD and this improvement is statistically significant ($p < 0.001$).

The studies with the smallest standard errors are making the largest relative contribution to the calculation of the total SMD. It should be borne in mind, however, that some studies are rated as being at a high risk of bias despite having narrow CIs (e.g. Smith²² and, to a lesser extent, Kendall *et al.*²⁶) and ordinarily might not have been included in this type of analysis, simply because the methodology is not sufficiently robust to allow confidence that the results are accurate or trustworthy. As the evidence for ROE grows and more rigorous studies are conducted, it will become possible to restrict the analysis to only those most rigorous studies, which will ultimately yield a more precise, less biased and more trustworthy result. At the moment, however, these analyses should be interpreted with a degree of caution.

The forest plot depicted in *Figure 6* provides an alternative representation of the table in *Figure 6*. Each green square represents the effect size associated with that study (the larger the green square, the larger the effect size) and the bars on either side of the square represent the 95% CI. The wider the CI, the less precise the estimated effect size. If the bars touch or cross the zero line, this means that the effect of the intervention was not statistically significant in that study. The black diamond depicts the overall effect size for all the studies combined and the tips of the diamond on either side represent the precision of the estimate. The wider the diamond, the less precise the estimate. If the diamond touches or crosses the zero line, this means that the overall effect is not statistically significant. τ^2 (0.00, $\chi^2 = 8.39$, degrees of freedom = 6; $p = 0.21$) represents an estimate of the between-study variance in a random-effects meta-analysis. The square root of τ^2 is the estimated SD of the underlying effects across studies. If τ^2 is not statistically significant, this suggests that there is no statistical evidence for differences between studies and it is valid to pool the results from these studies into a single estimate. When τ^2 is statistically significant, this suggests that there are important differences between studies (statistical heterogeneity) and that it might be questionable or invalid to pool the results. τ^2 tells us whether or not the heterogeneity is significant, although it should be interpreted with caution as it typically has low power. It does not, however, indicate the magnitude of the heterogeneity, which is represented by I^2 . Although the following rubric should be carefully applied, in general $I^2 = 0\%$ (no heterogeneity), $I^2 = 25\%$ (low heterogeneity), $I^2 = 50\%$ (moderate heterogeneity) and $I^2 = 75\%$ (high heterogeneity).⁹⁵ Thus, for the meta-analysis reported above, there is low statistical heterogeneity between studies.

Aggressive behaviour

The meta-analysis for teacher-rated aggressive behaviour also utilised all seven included studies (*Figure 7*). The overall SMD was -0.18 (95% CI -0.33 to -0.03) in favour of the intervention group. This means that ROE decreases aggressive behaviour for participating children by, on average, 0.18 of a SD and this improvement is statistically significant ($p = 0.02$).

In relation to the first research question, therefore, this meta-analysis of the results from all seven previous evaluations of ROE provides evidence that the programme is effective in improving prosocial behaviour [effect size (ES) = 0.13; $p < 0.001$] and decreasing aggressive behaviour (ES = -0.18 ; $p = 0.02$). These

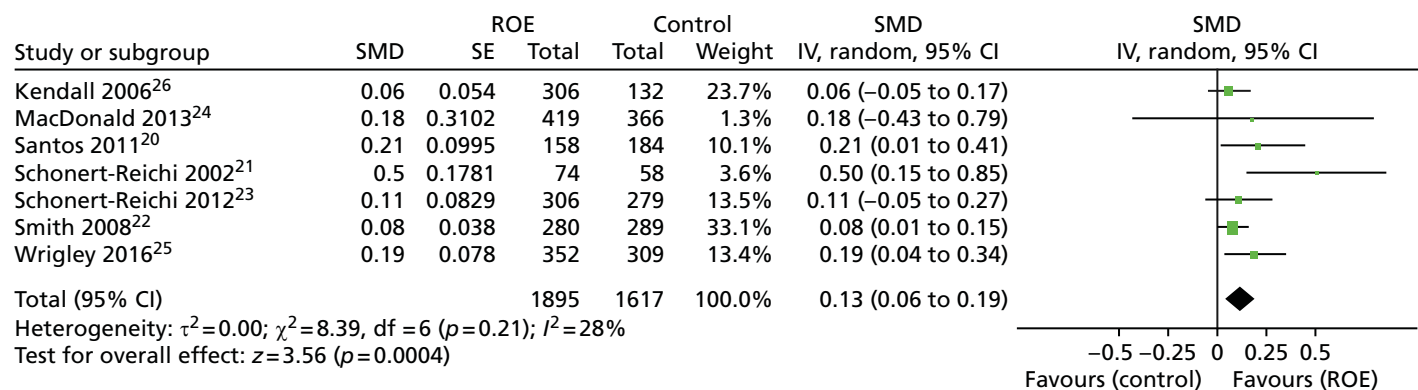


FIGURE 6 Forest plot of SMDs between intervention and control for prosocial behaviour. SE, standard error.

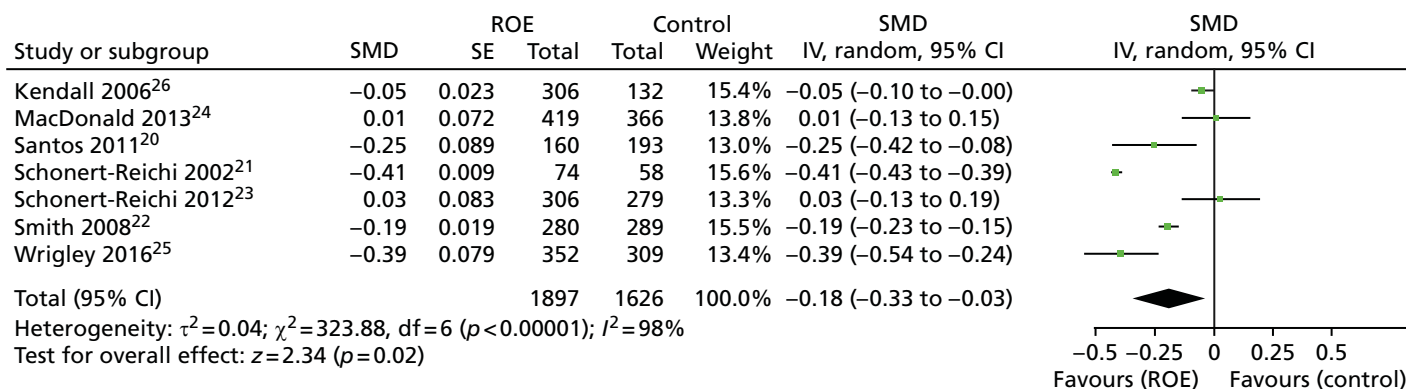


FIGURE 7 Forest plot of SMDs between intervention and control for aggressive behaviour. SE, standard error.

effects are statistically significant but there is considerable heterogeneity between the studies that generated these data ($\tau^2 = 0.04$; $p < 0.001$; $I^2 = 98\%$) and so caution should be exercised when interpreting these results. The CI associated with the effect size for aggressive behaviour is particularly wide (95% CI -0.33 to -0.03), indicating that this is not a very precise estimate of the true effect.

Empathy

Five of the seven studies measured child-reported empathy. As shown in *Figure 8*, the estimated SMD was 0.1 (95% CI -0.05 to 0.25) in favour of the intervention group, but this was not statistically significant ($p = 0.17$). This means that there is no evidence from the previous evaluations to suggest that ROE is effective at improving child-reported empathy. As above, there is moderate heterogeneity between the combined studies ($\tau^2 = 0.02$; $p < 0.001$; $I^2 = 72\%$).

Emotional regulation

Even fewer studies measured child-reported emotional regulation. Of the two that did, the SMD was 0.03 in favour of the intervention group, but this is an extremely small effect size and – as the forest plot in *Figure 9* depicts – is not statistically significant ($p = 0.60$).

With respect to the second research question, it seems that there is no (or, at best, insufficient) evidence to indicate that ROE improves child-reported empathy or emotional regulation.

Long-term effects of Roots of Empathy on (teacher-rated) prosocial and aggressive behaviour

Only one evaluation²⁰ studied the longer-term impact of the programme. This is the only other randomised controlled trial we have data for and it appears that, after 3 years, the intervention group had *poorer* prosocial behaviour than the control group (SMD -0.12 , 95% CI -0.17 to -0.07). This is in contrast to the positive impact of ROE on prosocial behaviour reported by this study immediately post test (ES 0.21). With respect to aggressive behaviour 3 years post intervention, the intervention group was displaying only slightly less aggressive behaviour than the control group (SMD -0.06 , 95% -0.09 to -0.03) and, although statistically significant, this effect was much reduced compared with the effect observed immediately post test (ES -0.25).

Conclusions

The meta-analyses reported above should be treated with a degree of caution owing to the unclear/high risk of bias of many of the included studies. Should the data from the other evaluations for which we have no information become available, then these will be incorporated into the analyses and the magnitude and precision of the estimated effect associated with each outcome may well change as a consequence.

Overall, and based on existing evidence, it seems that ROE consistently results in improvements in prosocial behaviour and reductions in aggressive behaviour. Interestingly, however, there appears to be no (or, at best, insufficient) evidence that the programme improves child-reported empathy or emotional regulation. Although this last outcome was measured by only two studies, neither study found that it was positively affected by ROE.

Similarly, the results from the one study that did follow-up children who participated in ROE found that the effects were not sustained long term. More evidence and more data are required to better understand if this is the case with different samples and in different contexts.

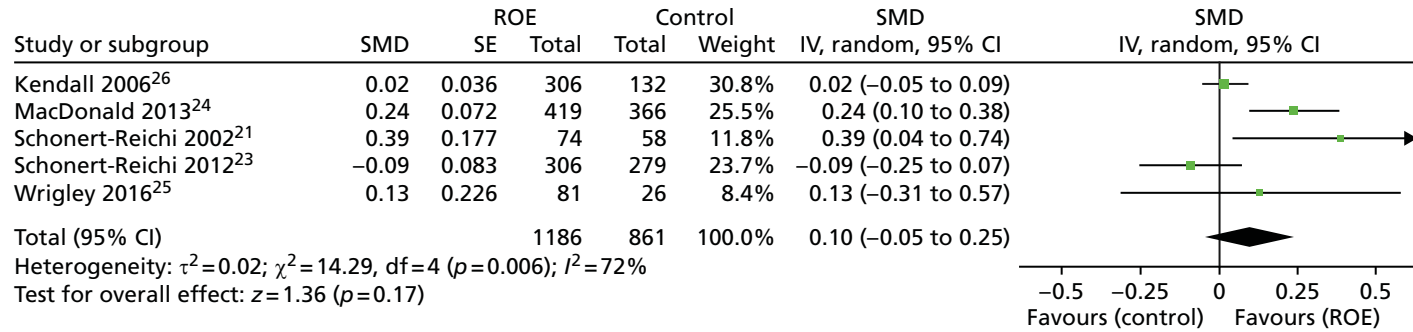


FIGURE 8 Forest plot of SMDs between intervention and control for child-reported empathy. SE, standard error.

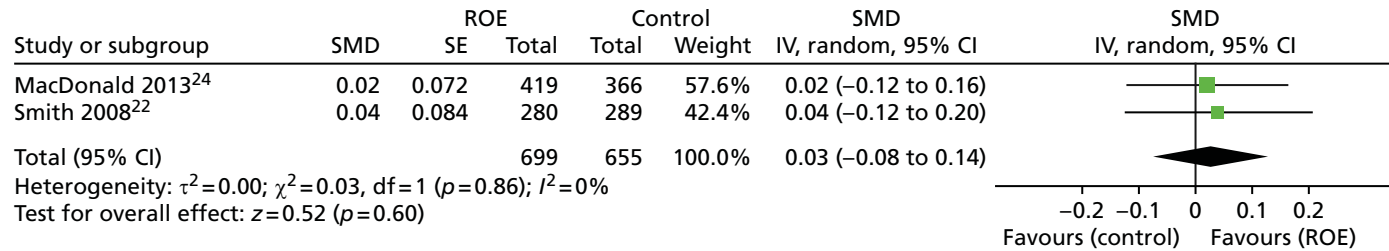


FIGURE 9 Forest plot of SMDs between intervention and control for child-reported emotional regulation. SE, standard error.

Appendix 2 Statistical models for immediately post test (time 1), time 2, time 3 and time 4

TABLE 32 Multilevel models fitted for the teacher-rated prosocial behaviour (SDQ) variable at each time point

Teacher-rated prosocial behaviour (SDQ)	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	0.199	0.099	-0.002	0.108	0.048	0.138	0.122	0.095
Gender	0.027	0.025	0.141	0.032	0.164	0.030	0.077	0.035
Deprivation	0.021	0.034	-0.028	0.042	0.050	0.041	0.027	0.045
Prosocial SDQ teacher T0	0.487	0.035	0.298	0.045	0.295	0.043	0.244	0.047
Difficulties SDQ teacher T0	-0.165	0.033	-0.169	0.042	-0.081	0.041	-0.177	0.046
Prosocial CBS teacher T0	0.040	0.036	-0.037	0.047	0.052	0.046	0.027	0.051
Aggressive CBS teacher T0	0.035	0.034	-0.065	0.044	0.013	0.044	-0.009	0.049
Reasons a baby cries T0	0.026	0.030	-0.086	0.039	-0.017	0.036	0.001	0.042
Ways to help a baby T0	0.020	0.031	0.060	0.041	0.022	0.039	0.047	0.046
Emotional recognition T0	0.027	0.025	0.017	0.033	0.081	0.031	0.021	0.036
Empathy T0	0.059	0.026	0.043	0.034	0.002	0.033	0.062	0.037
Emotional regulation T0	0.009	0.024	0.104	0.032	0.056	0.031	0.022	0.035
Bullying (victim) T0	0.019	0.025	-0.002	0.033	-0.001	0.032	-0.065	0.038
Quality of life T0	0.039	0.024	0.030	0.031	0.042	0.030	0.036	0.035
South Eastern HSCT dummy	0.000	0.062	0.030	0.069	0.018	0.086	-0.077	0.064
Southern HSCT dummy	0.045	0.062	0.019	0.065	0.067	0.086	0.122	0.061
Western HSCT dummy	-0.004	0.062	0.177	0.070	0.033	0.085	0.028	0.062
Constant	-0.121	0.071	-0.025	0.079	-0.101	0.100	-0.101	0.070
Log-likelihood	-1005.4		-1045.0		-1031.6		-920.4	
Ω_u	0.113	0.027	0.106	0.031	0.219	0.050	0.065	0.024
Ω_e	0.437	0.021	0.660	0.033	0.590	0.030	0.704	0.039

SE, standard error.

TABLE 33 Multilevel models fitted for the teacher-rated difficult behaviour (SDQ) variable at each time point

Teacher-rated difficult behaviour (SDQ)	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	-0.161	0.086	-0.144	0.118	-0.132	0.116	-0.142	0.096
Gender	0.019	0.020	-0.051	0.027	-0.064	0.028	-0.027	0.034
Deprivation	-0.028	0.028	-0.047	0.038	-0.077	0.038	-0.105	0.044
Prosocial SDQ teacher T1	-0.029	0.029	-0.094	0.038	-0.041	0.040	-0.080	0.045
Difficulties SDQ teacher T1	0.754	0.027	0.473	0.036	0.435	0.038	0.398	0.044
Prosocial CBS teacher T1	-0.030	0.030	0.034	0.040	-0.007	0.042	-0.102	0.049
Aggressive CBS teacher T1	-0.038	0.028	0.037	0.038	0.053	0.040	-0.088	0.047
Reasons a baby cries T1	-0.012	0.024	0.000	0.033	-0.009	0.034	0.051	0.040
Ways to help a baby T1	-0.028	0.026	-0.062	0.035	-0.017	0.036	-0.041	0.044
Emotional recognition T1	-0.014	0.021	-0.077	0.028	-0.090	0.029	-0.071	0.034
Empathy T1	-0.018	0.021	0.028	0.029	0.018	0.030	-0.045	0.035
Emotional regulation T1	-0.022	0.020	-0.053	0.027	-0.061	0.028	-0.045	0.034
Bullying (victim) T1	0.004	0.021	0.034	0.028	0.006	0.029	0.115	0.036
Quality of life T1	-0.018	0.020	-0.014	0.027	-0.058	0.028	-0.041	0.034
South Eastern HSCT dummy	0.039	0.053	0.007	0.073	-0.074	0.073	0.013	0.064
Southern HSCT dummy	0.003	0.053	-0.053	0.071	-0.169	0.072	-0.082	0.062
Western HSCT dummy	0.001	0.053	-0.131	0.077	-0.002	0.071	0.008	0.063
Constant	0.098	0.061	0.096	0.085	0.119	0.085	0.123	0.071
Log-likelihood	-824.1		-917.0		-961.0		-890.0	
Ω_u	0.086	0.020	0.151	0.037	0.147	0.036	0.074	0.028
Ω_e	0.297	0.014	0.470	0.024	0.506	0.026	0.637	0.035

SE, standard error.

TABLE 34 Multilevel models fitted for the child-rated 'reasons why a baby cries' variable at each time point

Child-rated 'reasons why a baby cries'	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	0.234	0.095	0.030	0.120	0.156	0.102	0.033	0.075
Gender	0.126	0.030	0.108	0.031	0.178	0.033	0.160	0.037
Deprivation	0.017	0.039	-0.001	0.042	0.026	0.043	0.094	0.044
Prosocial SDQ teacher T1	-0.007	0.041	-0.118	0.044	-0.017	0.046	0.040	0.050
Difficulties SDQ teacher T1	-0.160	0.039	-0.188	0.041	-0.100	0.043	-0.072	0.049
Prosocial CBS teacher T1	-0.036	0.043	0.039	0.047	-0.022	0.048	-0.030	0.054
Aggressive CBS teacher T1	0.020	0.040	0.019	0.043	-0.077	0.045	-0.037	0.052
Reasons a baby cries T1	0.178	0.036	0.213	0.038	0.111	0.040	0.147	0.044
Ways to help a baby T1	0.152	0.038	0.103	0.040	0.079	0.042	0.063	0.047
Emotional recognition T1	0.096	0.030	0.056	0.033	0.079	0.034	0.074	0.039
Empathy T1	0.039	0.031	0.024	0.033	0.027	0.035	-0.016	0.039
Emotional regulation T1	0.016	0.029	0.079	0.031	0.049	0.033	0.043	0.038
Bullying (victim) T1	-0.066	0.030	0.008	0.032	-0.075	0.034	0.002	0.040
Quality of life T1	-0.023	0.029	0.018	0.031	-0.001	0.032	-0.004	0.037
South Eastern HSCT dummy	0.006	0.061	0.055	0.075	0.118	0.066	-0.071	0.054
Southern HSCT dummy	-0.007	0.060	-0.007	0.074	-0.019	0.064	0.046	0.051
Western HSCT dummy	-0.051	0.059	-0.036	0.074	0.032	0.064	-0.041	0.049
Constant	-0.130	0.068	-0.023	0.086	-0.100	0.073	-0.059	0.057
Log-likelihood	-1196.5		-1139.4		-1196.4		-985.4	
Ω_u	0.088	0.025	0.158	0.039	0.096	0.029	0.009	0.014
Ω_e	0.659	0.031	0.668	0.033	0.765	0.037	0.844	0.046

SE, standard error.

TABLE 35 Multilevel models fitted for the child-rated 'ways to help a crying baby' variable at each time point

Child-rated 'ways to help a crying baby'	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	0.171	0.093	-0.025	0.126	0.103	0.104	-0.053	0.080
Gender	0.138	0.030	0.205	0.031	0.265	0.031	0.215	0.037
Deprivation	-0.019	0.039	0.027	0.043	-0.010	0.042	0.071	0.044
Prosocial SDQ teacher T1	-0.014	0.041	-0.023	0.044	-0.006	0.044	-0.074	0.050
Difficulties SDQ teacher T1	-0.121	0.040	-0.121	0.041	-0.086	0.041	-0.110	0.049
Prosocial CBS teacher T1	-0.056	0.044	-0.056	0.047	0.003	0.046	0.076	0.054
Aggressive CBS teacher T1	-0.050	0.041	-0.032	0.043	-0.070	0.043	-0.003	0.052
Reasons a baby cries T1	0.115	0.037	0.176	0.037	0.114	0.038	0.157	0.044
Ways to help a baby T1	0.225	0.038	0.130	0.040	0.105	0.040	0.027	0.047
Emotional recognition T1	0.078	0.031	0.072	0.032	0.042	0.033	0.039	0.039
Empathy T1	0.069	0.032	0.022	0.032	0.060	0.034	-0.001	0.039
Emotional regulation T1	0.028	0.029	0.058	0.031	0.026	0.032	0.016	0.037
Bullying (victim) T1	-0.021	0.031	0.015	0.032	-0.060	0.033	-0.004	0.040
Quality of life T1	-0.043	0.030	0.009	0.030	-0.008	0.031	0.028	0.037
South Eastern HSCT dummy	0.011	0.060	0.041	0.078	0.072	0.066	0.004	0.057
Southern HSCT dummy	0.023	0.058	0.045	0.077	-0.056	0.065	0.054	0.054
Western HSCT dummy	-0.014	0.058	-0.022	0.077	-0.004	0.064	-0.025	0.052
Constant	-0.096	0.067	0.008	0.090	-0.050	0.074	0.000	0.060
Log-likelihood	-1214.0		-1125.8		-1155.4		-978.3	
Ω_u	0.080	0.023	0.179	0.044	0.106	0.030	0.021	0.016
Ω_e	0.687	0.032	0.650	0.032	0.695	0.034	0.818	0.044

SE, standard error.

TABLE 36 Multilevel models fitted for the child-rated emotional recognition variable at each time point

Child-rated emotional recognition	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	0.094	0.075	-0.002	0.069	-0.036	0.072	-0.078	0.086
Gender	0.044	0.029	0.073	0.030	-0.044	0.036	-0.031	0.039
Deprivation	-0.038	0.035	0.000	0.036	0.030	0.040	-0.005	0.047
Prosocial SDQ teacher T1	-0.058	0.039	-0.025	0.042	-0.028	0.049	0.043	0.053
Difficulties SDQ teacher T1	-0.110	0.037	-0.121	0.039	-0.010	0.046	-0.019	0.051
Prosocial CBS teacher T1	0.020	0.041	0.057	0.044	0.061	0.051	0.030	0.057
Aggressive CBS teacher T1	-0.002	0.038	0.074	0.041	0.000	0.048	-0.001	0.055
Reasons a baby cries T1	-0.030	0.034	0.007	0.037	-0.049	0.044	-0.043	0.046
Ways to help a baby T1	0.110	0.036	0.008	0.038	0.009	0.045	0.017	0.050
Emotional recognition T1	0.276	0.029	0.219	0.032	0.007	0.038	0.061	0.041
Empathy T1	0.022	0.030	-0.015	0.031	-0.035	0.038	0.028	0.041
Emotional regulation T1	0.034	0.028	0.036	0.031	0.007	0.036	0.020	0.039
Bullying (victim) T1	-0.039	0.029	-0.032	0.031	-0.006	0.037	-0.007	0.042
Quality of life T1	0.084	0.028	0.035	0.030	0.024	0.035	0.016	0.039
South Eastern HSCT dummy	0.026	0.049	-0.093	0.046	-0.008	0.049	0.039	0.061
Southern HSCT dummy	0.028	0.047	-0.043	0.043	-0.073	0.045	0.023	0.057
Western HSCT dummy	-0.100	0.047	-0.056	0.043	0.019	0.045	-0.013	0.056
Constant	-0.006	0.054	0.063	0.051	0.043	0.053	0.014	0.064
Log-likelihood	-1152.8		-1111.2		-1275.4		-1016.7	
Ω_u	0.040	0.014	0.019	0.011	0.006	0.014	0.025	0.022
Ω_e	0.617	0.029	0.670	0.033	0.972	0.047	0.910	0.050

SE, standard error.

TABLE 37 Multilevel models fitted for the child-rated empathy variable at each time point

Child-rated empathy	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	-0.064	0.078	-0.045	0.099	0.088	0.028	-0.059	0.072
Gender	0.129	0.031	0.227	0.032	0.707	0.035	0.233	0.037
Deprivation	0.009	0.038	0.058	0.041	0.088	0.028	0.072	0.043
Prosocial SDQ teacher T1	0.036	0.041	0.028	0.045	0.707	0.035	0.067	0.050
Difficulties SDQ teacher T1	0.005	0.040	0.030	0.042	0.088	0.028	0.006	0.048
Prosocial CBS teacher T1	-0.020	0.044	0.066	0.047	0.707	0.035	-0.013	0.053
Aggressive CBS teacher T1	-0.016	0.041	0.004	0.044	0.088	0.028	0.042	0.051
Reasons a baby cries T1	-0.017	0.037	-0.040	0.039	0.707	0.035	-0.008	0.044
Ways to help a baby T1	0.019	0.039	0.046	0.040	0.088	0.028	0.047	0.047
Emotional recognition T1	0.037	0.032	-0.065	0.033	0.707	0.035	-0.027	0.039
Empathy T1	0.372	0.032	0.258	0.033	0.088	0.028	0.203	0.039
Emotional regulation T1	0.024	0.030	0.029	0.032	0.707	0.035	-0.021	0.038
Bullying (victim) T1	0.090	0.031	0.058	0.033	0.088	0.028	0.041	0.040
Quality of life T1	-0.020	0.030	-0.058	0.031	0.707	0.035	0.011	0.037
South Eastern HSCT dummy	0.077	0.051	0.010	0.064	0.088	0.028	-0.079	0.053
Southern HSCT dummy	0.053	0.049	0.057	0.061	0.707	0.035	-0.039	0.049
Western HSCT dummy	-0.017	0.049	-0.043	0.061	0.088	0.028	-0.139	0.047
Constant	0.012	0.056	0.015	0.072	0.707	0.035	0.048	0.055
Log-likelihood	-1224.0		-1156.8		-1162.1		-986.1	
Ω_u	0.039	0.017	0.088	0.028	0.043	0.017	0.003	0.013
Ω_e	0.720	0.034	0.707	0.035	0.731	0.035	0.845	0.046

SE, standard error.

TABLE 38 Multilevel models fitted for the child-rated emotional regulation variable at each time point

Child-rated emotional regulation	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	0.084	0.069	0.018	0.090	0.084	0.077	0.023	0.073
Gender	0.042	0.032	0.042	0.033	0.040	0.033	0.094	0.038
Deprivation	0.111	0.037	0.122	0.041	0.157	0.040	0.102	0.044
Prosocial SDQ teacher T1	0.045	0.042	0.068	0.046	0.046	0.046	0.074	0.051
Difficulties SDQ teacher T1	-0.116	0.041	-0.117	0.043	-0.078	0.043	-0.113	0.049
Prosocial CBS teacher T1	-0.093	0.045	0.032	0.048	0.042	0.048	-0.066	0.055
Aggressive CBS teacher T1	-0.039	0.041	0.014	0.045	0.024	0.045	-0.040	0.053
Reasons a baby cries T1	-0.046	0.038	-0.017	0.040	0.019	0.041	-0.069	0.046
Ways to help a baby T1	0.018	0.039	0.029	0.042	-0.031	0.042	0.061	0.049
Emotional recognition T1	0.018	0.032	0.037	0.034	0.004	0.035	0.019	0.041
Empathy T1	0.083	0.033	-0.013	0.034	0.031	0.035	-0.032	0.040
Emotional regulation T1	0.325	0.031	0.336	0.033	0.210	0.034	0.175	0.039
Bullying (victim) T1	0.001	0.032	0.023	0.034	-0.050	0.034	-0.005	0.041
Quality of life T1	0.057	0.031	0.021	0.032	0.067	0.033	0.051	0.038
South Eastern HSCT dummy	-0.002	0.046	-0.011	0.059	-0.052	0.052	-0.074	0.054
Southern HSCT dummy	0.085	0.044	0.080	0.056	0.103	0.049	0.044	0.050
Western HSCT dummy	-0.066	0.044	-0.036	0.056	-0.057	0.048	-0.016	0.048
Constant	-0.049	0.051	-0.054	0.066	-0.065	0.056	-0.030	0.056
Log-likelihood	-1248.7		-1185.5		-1202.6		-1012.5	
Ω_u	0.018	0.014	0.060	0.022	0.028	0.016	0.000	0.000
Ω_e	0.772	0.037	0.767	0.037	0.808	0.039	0.910	0.047

SE, standard error.

TABLE 39 Multilevel models fitted for the child-rated bullying (victim) variable at each time point

Child-rated bullying (victim)	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	0.122	0.080	0.004	0.076	-0.047	0.069	-0.028	0.076
Gender	0.043	0.032	0.055	0.032	0.072	0.036	0.014	0.039
Deprivation	-0.004	0.039	-0.005	0.039	-0.058	0.039	-0.068	0.045
Prosocial SDQ teacher T1	-0.011	0.043	0.034	0.045	0.057	0.048	-0.022	0.052
Difficulties SDQ teacher T1	0.105	0.041	0.210	0.042	0.130	0.045	0.060	0.051
Prosocial CBS teacher T1	0.028	0.045	0.023	0.047	0.028	0.051	0.012	0.056
Aggressive CBS teacher T1	-0.044	0.042	-0.016	0.044	0.023	0.047	0.042	0.054
Reasons a baby cries T1	-0.053	0.038	0.032	0.040	0.042	0.043	0.109	0.046
Ways to help a baby T1	0.000	0.040	-0.034	0.041	-0.062	0.044	-0.071	0.050
Emotional recognition T1	-0.013	0.032	-0.051	0.034	0.006	0.037	-0.090	0.042
Empathy T1	-0.014	0.033	0.070	0.034	-0.020	0.037	0.025	0.041
Emotional regulation T1	-0.018	0.031	-0.010	0.033	-0.038	0.036	-0.026	0.040
Bullying (victim) T1	0.298	0.032	0.250	0.034	0.254	0.037	0.192	0.042
Quality of life T1	-0.075	0.031	-0.031	0.032	-0.084	0.035	-0.043	0.039
South Eastern HSCT dummy	0.036	0.053	0.091	0.051	0.042	0.047	0.047	0.056
Southern HSCT dummy	-0.015	0.051	0.058	0.047	-0.081	0.043	0.046	0.052
Western HSCT dummy	0.049	0.050	0.062	0.047	0.018	0.043	0.049	0.050
Constant	-0.083	0.058	0.003	0.056	0.063	0.051	0.052	0.058
Log-likelihood	-1248.7		-1172.5		-1262.0		-1014.6	
Ω_u	0.042	0.018	0.027	0.016	0.002	0.013	0.004	0.013
Ω_e	0.759	0.036	0.765	0.037	0.948	0.046	0.933	0.050

SE, standard error.

TABLE 40 Multilevel models fitted for the child-rated bullying (bully) variable at each time point

Child-rated bullying (bully)	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	–	–	–	–	–0.005	0.070	–0.004	0.079
Gender	–	–	–	–	–0.047	0.037	–0.087	0.041
Deprivation	–	–	–	–	–0.084	0.040	–0.044	0.047
Prosocial SDQ teacher T1	–	–	–	–	–0.032	0.049	–0.045	0.055
Difficulties SDQ teacher T1	–	–	–	–	0.024	0.046	0.015	0.054
Prosocial CBS teacher T1	–	–	–	–	0.060	0.052	0.142	0.060
Aggressive CBS teacher T1	–	–	–	–	0.005	0.048	0.098	0.058
Reasons a baby cries T1	–	–	–	–	0.007	0.045	0.048	0.049
Ways to help a baby T1	–	–	–	–	–0.036	0.046	–0.036	0.052
Emotional recognition T1	–	–	–	–	–0.027	0.038	–0.102	0.044
Empathy T1	–	–	–	–	–0.004	0.038	–0.052	0.044
Emotional regulation T1	–	–	–	–	–0.031	0.037	–0.006	0.042
Bullying (victim) T1	–	–	–	–	0.144	0.038	0.161	0.044
Quality of life T1	–	–	–	–	–0.057	0.036	0.053	0.041
South Eastern HSCT dummy	–	–	–	–	–0.040	0.048	0.027	0.058
Southern HSCT dummy	–	–	–	–	–0.138	0.044	–0.034	0.054
Western HSCT dummy	–	–	–	–	–0.053	0.044	–0.054	0.052
Constant	–	–	–	–	0.038	0.052	0.037	0.060
Log-likelihood	–	–	–	–	–1285.5		–1042.2	
Ω_u	–	–	–	–	0.000	0.012	0.000	0.000
Ω_e	–	–	–	–	1.003	0.049	1.038	0.055

SE, standard error.

TABLE 41 Multilevel models fitted for the child-rated quality-of-life variable at each time point

Child-rated quality of life	Time point							
	Main model T1		Main model T2		Main model T3		Main model T4	
	β	SE	β	SE	β	SE	β	SE
Group	0.017	0.069	0.186	0.070	0.099	0.067	0.111	0.083
Gender	-0.037	0.033	0.013	0.034	0.019	0.035	-0.069	0.040
Deprivation	-0.074	0.037	0.025	0.038	0.035	0.038	0.014	0.047
Prosocial SDQ teacher T1	-0.009	0.044	-0.008	0.046	-0.075	0.048	0.045	0.054
Difficulties SDQ teacher T1	-0.029	0.042	0.029	0.043	-0.140	0.044	-0.047	0.053
Prosocial CBS teacher T1	-0.005	0.046	-0.020	0.049	-0.064	0.050	-0.051	0.058
Aggressive CBS teacher T1	0.048	0.043	-0.010	0.045	0.009	0.047	-0.050	0.056
Reasons a baby cries T1	0.023	0.039	-0.083	0.042	-0.058	0.043	-0.045	0.047
Ways to help a baby T1	0.033	0.040	0.015	0.042	0.010	0.044	0.020	0.051
Emotional recognition T1	0.034	0.034	0.005	0.036	0.062	0.037	0.023	0.042
Empathy T1	0.013	0.034	0.003	0.035	0.095	0.037	-0.015	0.042
Emotional regulation T1	0.038	0.032	0.052	0.034	0.056	0.035	-0.044	0.040
Bullying (victim) T1	-0.077	0.034	-0.058	0.035	-0.064	0.037	-0.062	0.043
Quality of life T1	0.311	0.032	0.258	0.033	0.206	0.034	0.112	0.040
South Eastern HSCT dummy	0.005	0.046	-0.126	0.047	-0.008	0.046	-0.010	0.060
Southern HSCT dummy	0.006	0.044	-0.099	0.043	0.078	0.042	0.075	0.056
Western HSCT dummy	0.062	0.043	-0.014	0.043	0.056	0.042	0.072	0.054
Constant	-0.005	0.051	-0.098	0.052	-0.083	0.050	-0.109	0.063
Log-likelihood	-1234.4		-1189.4		-1219.8		-1021.8	
Ω_u	0.014	0.013	0.009	0.012	0.000	0.000	0.016	0.019
Ω_e	0.806	0.039	0.835	0.041	0.911	0.043	0.955	0.052

SE, standard error.

Appendix 3 Sensitivity analysis to assess impact of not including three child-level covariates in the statistical models

The original protocol for this trial specified that a series of covariates would be added to the statistical models used to estimate the effects of the ROE programme. These were to include pre-test scores for all of the primary and secondary outcomes listed in the study, together with a series of child characteristics collected for the study. The data for five of the variables representing child characteristics depended on the completion and return of questionnaires by parents. These were:

- number of siblings in the family
- mother's highest educational qualification
- father's highest educational qualification
- mother's employment status
- father's employment status.

Unfortunately, owing to lower response rates, data on these measures were collected for only approximately half of the sample. As such, it was decided not to include these as covariates in the main models. This sensitivity analysis assesses whether or not the omission of these five variables has had an impact on the overall findings.

To do this, the main models were refitted with the inclusion of these five variables. The estimated effects for each primary and secondary outcome immediately post test from these models are compared with the estimated effects from the main analysis in *Table 42*. As can be seen, the overall pattern of findings is similar whether or not these additional covariates are added. However, it can also be seen that the

TABLE 42 Comparison of effect sizes immediately post test between the models used in the main analysis, these models extended with additional covariates and then analysed with multiple imputation of missing data^a

Outcomes	Effect sizes (with statistical significance)		
	Main analysis	Additional covariates	Additional covariates and imputed data
Primary outcomes			
Prosocial behaviour (SDQ)	0.199 ($p = 0.045$)	0.296 ($p = 0.006$)	0.209 ($p = 0.017$)
Difficult behaviour (SDQ)	-0.162 ($p = 0.060$)	-0.208 ($p = 0.013$)	-0.163 ($p = 0.022$)
Secondary outcomes			
Reasons why a baby cries	0.235 ($p = 0.014$)	0.264 ($p = 0.027$)	0.241 ($p = 0.005$)
Ways to help a crying baby	0.171 ($p = 0.066$)	0.094 ($p = 0.434$)	0.173 ($p = 0.051$)
Emotional recognition	0.094 ($p = 0.211$)	0.090 ($p = 0.241$)	0.128 ($p = 0.078$)
Empathy	-0.064 ($p = 0.410$)	-0.033 ($p = 0.653$)	-0.012 ($p = 0.863$)
Emotional regulation	0.084 ($p = 0.229$)	0.012 ($p = 0.906$)	0.107 ($p = 0.146$)
Bullying (victim)	0.122 ($p = 0.129$)	0.072 ($p = 0.447$)	0.090 ($p = 0.234$)
Quality of life	-0.017 ($p = 0.806$)	-0.004 ($p = 0.966$)	-0.010 ($p = 0.887$)

^a The five additional covariates are number of siblings, mother's highest qualification, father's highest qualification, mother's employment status and father's employment status.

addition of these covariates is associated with slightly higher estimated effects for the two primary outcomes and the one secondary outcome that was statistically significant in the main analysis.

Given that there is some evidence to suggest that the non-response of parents might not be random (see *Tables 11 and 12*), it was decided to extend this analysis further by rerunning the models with the five additional covariates using fully imputed data sets. For this, multiple imputation using chained equations was used to create 20 data sets ($m = 20$). The variables used to generate the data sets included all of the pre-test variables and all of the variables representing immediate post-test scores for all of the primary and secondary outcomes identified for the study. In addition, the five new covariates were also included together with the fully observed variables of gender and trust location.

As can also be seen from *Table 42*, the further analysis using imputed data sets has tended to reduce the effect sizes back to those found in the main analysis. As such, it can be concluded that the omission of these five covariates has not had a notable impact on the core findings of this trial.

Appendix 4 Resource use questionnaire

We are interested in finding out what health and social care services your child has had access to since they started Primary 5.

Please indicate below how many times (if at all) your child has used any of the services listed below. If your child has not been in contact with a particular service, please enter '0' rather than leaving it blank.

Note: please enter '0' if service has not been used

Service	Total number of contacts
General Practitioner (GP)	
School Nurse	
Accident and Emergency (A&E) Visit	
Social Worker	
Speech therapist	
Occupational therapist	
Educational Psychologist	
Counselling/therapy	
Dentist	
Optician	
Police	
Hospital stay	Number of nights:
Hospital outpatient visit	

Has your child used any other services that are not listed in the table above? If so please let us know:

Other service 1: _____

No. of contacts: _____

Other service 2: _____

No. of contacts: _____

Please list below your child's use of any medication taken since he/she started P5.

Name of medication	How long did your child take this medication for? (e.g. 1 week)	Daily Dosage
1.		
2.		
3.		
4.		

Since your child started P5, have you had to take time off work or your usual daily activities due to your child being off school? (For example, time off due to child's illness, behavioural problems, attending appointments etc.) Yes No

If yes, please state how many days: _____

THANK YOU

A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and depth.

**EME
HS&DR
HTA
PGfAR
PHR**

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library