



Incredible Years Pilot Study

Evaluation Report



Published April 2013
Ministry of Social Development
PO Box 1556
Wellington 6140
New Zealand

Telephone: +64 4 916 3300
Facsimile: +64 4 918 0099

Email: info@msd.govt.nz
Web: www.msd.govt.nz

ISBN: 978-0-478-33567-5 (print)
ISBN: 978-0-478-33568-2 (online)

Incredible Years Pilot Study

Evaluation Report

Fiona Sturrock and Dorian Gray
Centre for Research and Evaluation
Ministry of Social Development

In collaboration with

Academic Advisors:

Dr John Church

Professor David Fergusson

Associate Professor John Horwood

Professor Richie Poulton

and

Government Agencies:

Dr Sarah Hayward

John McGovern

APRIL 2013

Contents

LIST OF TABLES/ LIST OF FIGURES	VI
ACKNOWLEDGEMENTS	VII
EXECUTIVE SUMMARY	1
CHAPTER 1: INTRODUCTION	4
1.1 Background to the Incredible Years Pilot Study.....	4
1.2 Childhood conduct problems	5
1.3 Early intervention	6
1.4 Parent training programmes	6
1.5 Cost effectiveness.....	7
1.6 The New Zealand context.....	7
1.7 The report structure.....	8
CHAPTER 2: THE INCREDIBLE YEARS PARENTING PROGRAMME	9
2.1 An overview of the Incredible Years Parenting programme.....	9
2.2 Evaluations of the Incredible Years Parenting programme	11
CHAPTER 3: THE NEW ZEALAND INCREDIBLE YEARS PILOT STUDY.....	13
3.1 Research design.....	13
3.2 Fieldwork issues.....	19
3.3 Statistical analyses	20
CHAPTER 4: RECRUITMENT, RETENTION AND PARTICIPANT PROFILES.....	22
4.1 The referral process to Special Education Incredible Years programmes.....	22
4.2 Research participant recruitment and retention.....	23
4.3 Participant profiles.....	24
CHAPTER 5: FINDINGS RELATED TO CHILD BEHAVIOUR, PARENTING PRACTICE AND RELATIONSHIPS.....	27
5.1 Child behaviour.....	28
5.2 Parenting practices and relationships	30
5.3 Teacher pre/post measures of child behaviour	36
CHAPTER 6: OTHER OUTCOMES.....	39
6.1 Outcomes for health and stressful life events.....	39
6.2 Outcomes for children in clinical and sub-clinical groups.....	42
6.3 Outcomes for phase 1 and phase 2.....	46
6.4 Outcomes for the three sites.....	48
CHAPTER 7: PROGRAMME OUTCOMES FOR MAORI AND NON-MAORI.....	50
7.1 The Incredible Years Pilot Study	50
7.2 The kaupapa Māori study	58
7.3 Incredible Years for whānau Māori.....	59
CHAPTER 8: SINGLE CASE STUDIES.....	60
8.1 Parent Reports.....	60

CHAPTER 9: PROGRAMME FIDELITY, GROUP LEADER FEEDBACK AND PARENT SATISFACTION.....	68
9.1 Programme fidelity: Unified protocols.....	68
9.2 Group Leader feedback.....	69
9.3 Parent satisfaction with the Incredible Years Parenting programme.....	69
CHAPTER 10: POLICY IMPLICATIONS.....	72
10.1 The efficacy of the Incredible Years Parenting programme in New Zealand.....	73
10.2 Extensions of the Incredible Years Pilot Study.....	73
10.3 The selection of Incredible Years Parenting programme client families.....	74
10.4 The effects of context.....	74
10.5 Māori/non-Māori differences.....	75
10.6 Participants who do not respond to the Incredible Years Parenting programme.....	75
10.7 Site differences.....	75
10.8 The need for an effectiveness study.....	76
10.9 Organisational implications.....	76
10.10 Summary of Evaluation Advisory Group recommendations.....	77
CHAPTER 11: SUMMARY AND CONCLUSIONS.....	78
1.1 Summary of findings related to research questions.....	78
11.2 Conclusions.....	80
11.3 Future research.....	81
APPENDIX 1. THE INCREDIBLE YEARS PILOT STUDY EVALUATION ADVISORY GROUP.....	82
APPENDIX 2. SOURCE INSTRUMENTS FOR THE INCREDIBLE YEARS PILOT STUDY.....	83
APPENDIX 3. STATISTICAL METHODS.....	87
REFERENCES.....	91

List of Tables

Table 1.	A sample of RCT studies investigating the efficacy of IYP programmes, with post-treatment child behaviour effect sizes (ES).....	11
Table 2.	A sample of RCT studies investigating the efficacy of IYP programmes, with post-treatment parenting practices effect sizes (ES).....	12
Table 3.	Topics covered in the Family Interview.....	17
Table 4.	Child behaviours: Mean parent reports before, during and after IYP programme.....	29
Table 5.	Child behaviours: Mean parent reports before IYP programme and at follow-up.....	30
Table 6.	Parenting practices: Mean parent reports before, during and after IYP programme.....	31
Table 7.	Parenting practices: Mean parent reports before IYP programme and at follow-up.....	32
Table 8.	Parental conflict and relationships: Mean parent reports before, during and after IYP programme.....	34
Table 9.	Parental conflict and relationships: Mean parent reports before IYP programme and at follow-up.....	35
Table 10.	Child behaviours: Mean teacher reports before and after IYP programme.....	36
Table 11.	Child behaviours: Mean teacher reports before IYP programme and at follow-up.....	37
Table 12.	Child behaviours: Mean teacher reports before and after IYP programme and at follow-up, by referral source.....	37
Table 13.	Child behaviour: Effect sizes of mean teacher reports of change from baseline to post-course and follow-up, by referral source.....	38
Table 14.	Health and stressful life events: Mean parent reports before, during and after IYP programme.....	40
Table 15.	Health and stressful life events: Mean parent reports before IYP programme and at follow-up.....	41
Table 16.	Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by sub-clinical and clinical ECBI Intensity.....	42
Table 17.	Child behaviours at follow-up: Effect sizes of mean parent reports of change, by sub-clinical and clinical ECBI Intensity.....	43
Table 18.	Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by sub-clinical and clinical Social Competence Scale (SCS).....	44
Table 19.	Child behaviours at follow-up: Effect sizes of mean parent reports of change, by sub-clinical and clinical Social Competence Scale (SCS).....	45
Table 20.	Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by study phase.....	46
Table 21.	Child behaviours at follow-up: Effect sizes of mean parent reports of change, by study phase.....	47
Table 22.	Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by study site.....	48
Table 23.	Child behaviours at follow-up: Mean parent reports, by study site.....	49
Table 24.	Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by Māori and non-Māori children.....	51
Table 25.	Child behaviours at follow-up: Effect sizes of mean parent reports of change, by Māori and non-Māori children.....	52
Table 26.	Parenting practices after IYP programme: Effect sizes of mean parent reports of change, by Māori and non-Māori children.....	53
Table 27.	Parenting practices at follow-up: Effect sizes of mean parent reports of change, by Māori and non-Māori children.....	54

Table 28.	Conflict and relationships after IYP programme: Effect sizes of mean parent reports of change, by Māori and non-Māori children.....	55
Table 29.	Conflict and relationships at follow-up: Mean parent reports of change, by Māori and non-Māori children.....	56
Table 30.	Summary of the multivariate analyses for Māori and non-Māori children at IYP programme completion and follow-up.....	57
Table 31.	Satisfaction with course organisation: opinions before and after the course by ethnicity.....	70
Table 32.	Satisfaction with the course overall at course completion by ethnicity.....	71
Table 33.	Effect sizes for child behaviour and parenting practices reported in Larsson et al (2009) and the Incredible Years Pilot Study.....	80
Table 34.	Bonferroni adjustments applied.....	89
Table 35.	Summary of analysis methods used.....	90

List of Figures

Figure 1.	Incredible Years Parenting programme pyramid.....	10
Figure 2.	The research design.....	16
Figure 3.	Incredible Years Pilot Study participant recruitment and retention.....	23
Figure 4.	Example of no behaviour change (Child 101).....	62
Figure 5.	Example of behaviour change towards improvement (Child 311).....	62
Figure 6.	Predicted positive behaviour scores for each child and for the IYP programme overall.....	63
Figure 7.	Predicted negative behaviour scores for each child and for the IYP programme overall.....	64
Figure 8.	Predicted combined behaviour scores for each child and for the IYP programme overall.....	65
Figure 9.	Predicted positive behaviour scores for each child through the programme to the follow-up.....	66
Figure 10.	Predicted negative behaviour scores for each child through the programme to the follow-up.....	66
Figure 11.	Predicted combined behaviour scores for each child through the programme to the follow-up.....	67

Acknowledgements

The Incredible Years Pilot Study was a substantial cross-agency study conducted over 2 years. The Ministry of Education Special Education delivered the Incredible Years Parenting courses that were evaluated in the study. The Ministry of Health provided the operational funding for the study. The evaluation team is within the Ministry of Social Development.

We thank the following for their contribution to the Incredible Years Pilot Study:

- The Incredible Years Evaluation Advisory Group provided ongoing advice and support for the study: Dr John Church, Professor David Fergusson, Professor Richie Poulton, Dr Sarah Hayward, Brian Coffey, John McGovern and Ross Mackay. In particular: Professor Fergusson provided invaluable advice and guidance throughout the study; Dr Church designed, oversaw and reported on the Direct Observations trial; and John McGovern was the liaison between Special Education and the evaluation team.
- Associate Professor John Horwood of the Christchurch Health and Development Study provided essential statistical advice and review.
- The Group Leaders of the Incredible Years Parenting courses and their Supervisors co-operated with the evaluation team, supplying data additional to their regular administrative requirements and answering endless questions.
- Christina Smits helped with the administration of the Incredible Years Pilot Study and the editing of this report.
- The hard work and dedication of Robyn Deighton, Kathryn Handley and Trish Ross, the Interviewers in the three sites, resulted in quality data and high retention rates for the study.
- We are indebted to the 166 families who agreed to participate in the study, almost all of whom remained engaged throughout the four interviews.

Fiona Sturrock

Dorian Gray

Executive summary

Introduction

- A growing body of research provides evidence of the prevalence of childhood conduct problems and the long-term negative consequences that result. It also identifies parent management training such as the Incredible Years Parenting (IYP) programme as an effective evidence-based treatment.
- As part of the Drivers of Crime work programme the Ministries of Education, Health and Social Development established a pilot study of the IYP programme to assess the effectiveness of this programme in reducing conduct problems in a New Zealand context.
- The project was influenced by the recommendations of the Government Advisory Group on Conduct Problems, the Ministry of Education Positive Behaviour for Learning strategy and the desire to develop a new collaborative model to evaluate government-funded programmes.
- The higher rates of conduct problems in Māori children mean that, to reduce conduct problems in New Zealand, it is critical parenting interventions are effective for and acceptable to Māori parents.
- The IYP programme is based on social learning theory. It is an intensive programme consisting of weekly 2-hour sessions delivered over 12–18 weeks by two trained Group Leaders to a group of up to 16 participants.
- The Incredible Years Pilot Study was a substantial 2-year, multiple-informant study. It included mixed measurement methods, single case studies and a 6-month follow-up. The study was designed to understand the profile of the families referred to the IYP programme, to ensure programme fidelity, to measure both programme effectiveness and parent satisfaction, and to assess programme responsiveness to Māori.
- A total of 166 primary caregivers of children aged 3–8 years attending an IYP programme at three sites (Bay of Plenty, Canterbury and Mid-Central) participated in the study.
- The main study used a repeated measures design in which assessments of the research participants were made at baseline, mid-programme, post-programme and 6 months following programme completion.
- Family Interviews were supplemented by a postal Teacher Questionnaire which was administered at baseline, post-programme and the 6-month follow-up.
- The Family Interview incorporated items from a number of previously validated instruments to measure outcome variables describing changes over the course of the study. These changes related to child behaviour, parenting practices, family relationships and other relevant outcomes such as health, parental depression and stressful life events.
- The study also included Single Case Studies with a sample of the participants using weekly Parent Reports.

Key findings

- There was clear evidence of child behaviour change, with effect sizes as expected on the basis of the international literature. Effect size, as measured by Cohen's d , provides an indication of the extent of reported change, whereby an effect size of $d=.20$ is considered small, an effect size of $d=.50$ is medium and an effect size of $d=.80$ is large.
- There was clear evidence of parenting behaviour change, with effect sizes as expected on the basis of the international literature.
- The benefits of the IYP training were broadly similar for Māori and non-Māori families. Nevertheless the evidence suggests the need for further work on maximising gains for Māori families, particularly in the maintenance of behaviour change.
- Both Māori and non-Māori parents expressed high to moderate satisfaction with the programme.
- These results suggest the IYP programme can be successfully implemented in New Zealand and retain its general level of effectiveness.

- The findings provide a series of performance benchmarks against which the future roll out of the IYP programme in New Zealand can be assessed.

Detailed findings

- Seventy-two percent of parents visited by Group Leaders to enrol in the IYP programme participated in the research. The recruitment rate rose to 85 percent when taken as a proportion of parents who consented to be contacted for the research at the time of enrolment. Retention in the study was 98 percent of parents for whom there were baseline data.
- The study reached the target group, ie children aged 3–8 years, a third of whom identified as Māori.
- The programme was delivered with fidelity to the stipulated IYP programme delivery, processes and techniques.
- Baseline and post-course comparisons of outcome measures for families enrolled in the IYP programme produced effect size estimates for change in child behaviour, parenting practice and family relationships consistent with previous IYP programme evaluations using randomised control trials.
- Children and parents receiving IYP training showed significant improvements in the following areas: child behaviour median effect size of $d=.65$ (ranging from $d=.51$ to $d=.96$), parenting practices median effect size of $d=.54$ (ranging from $d=.26$ to $d=.83$) and relationships median effect size of $d=.48$ (ranging from $d=.21$ to $d=.60$).
- The improvements evident at the completion of the IYP course were mostly sustained at the 6-month follow-up. Significant improvements were maintained in the following areas: child behaviour median $d=.71$ (range $d=.56$ to $d=1.0$), parenting practices median $d=.52$ (range $d=.25$ to $d=.79$) and relationships median $d=.43$ (range $d=.15$ to $d=.59$).
- Effect sizes for parenting practices and family relationships were typically lower than effect sizes for child behaviour. This indicated small changes in parenting practice may produce substantial improvements in child behaviour.
- Before-and-after comparisons of scores on the Social Development Scale completed by teachers produced modest effect sizes ($d=.17$ at post-course, $d=.29$ at follow-up).
- Children with scores in the clinical range on the pre-course Eyberg Child Behaviour Inventory (ECBI) Scaled Intensity measure improved to a greater extent at post-course on all child behaviour measures except Social Competence than children with scores below the clinical cut-off. Significant differences in reported Conduct Disorder (CD) and Oppositional Defiant Disorder (ODD) remained at follow-up, whereas differences in Attention Deficit Hyperactivity Disorder (ADHD), Self-control and Anxiety/Withdrawal were no longer evident, perhaps reflecting a continued improvement for children in the sub-clinical group.
- The greater improvement reported for children in the clinical range on the pre-course Social Competence Scale (SCS) on measures of CD and ODD at post-course were sustained at follow-up. Self-control and Social Competence had also improved to a greater extent for children in the SCS clinical group than for children in the sub-clinical group 6 months following course completion.
- Small to medium improvements in caregiver stress and anxiety were reported at post-course and follow-up, as were fewer child visits to health services and fewer stressful life events. Parents also reported a decrease in depressive symptoms at follow-up.
- Parent ratings in both the Family Interview and the in-course Client Satisfaction Questionnaire showed high to moderate satisfaction with the IYP programme.
- Although improvements were evident at all three sites, larger effect sizes were found in Mid-Central than in the Bay of Plenty or Canterbury.
- The evidence suggests the IYP programme was both effective for and accepted by Māori families. However, at post-course and follow-up, parents of Māori children reported slightly smaller improvements in child behaviour and slightly smaller reductions in negative parenting practices on the individual outcome measures than did parents of non-Māori children. These differences were not significant.

- On the combined outcome measures at programme completion the differences between Māori and non-Māori children were not significant. However, at follow-up a small but statistically significant difference between responses by parents of Māori and non-Māori children was detected for the overall child behaviour outcome measure ($p=.025$). This suggests there may be a particular challenge in maintaining the benefits of the IYP programme for Māori families.
- The data from the weekly Parent Report Single Case Studies replicated the improvement in child behaviour evident in the main study for the majority of children. It also highlighted the variability in child behaviour change.

Evaluation Advisory Group policy recommendations

- The Incredible Years Pilot Study should be viewed as providing adequate evidence of the efficacy of the IYP programme in New Zealand.
- Investments should be made into the longer-term follow-up of the families participating in this evaluation, with further assessments being made at 1 year and 2 years post-treatment.
- Consideration should be given to providing parent management training programmes, such as Triple P Level 4, that are less intensive than the IYP programme for client families with children whose behaviour scores on the ECBI scales place them in the sub-clinical range.
- Consideration should be given to developing parallel research to evaluate the Incredible Years Teacher programme to determine the benefits of this programme for behaviours in the classroom and for behaviours at home.
- Further investment is needed to increase the efficacy of the IYP programme for Māori families.
- There is a need to develop further programmes and interventions to meet the needs of families whose children do not show sufficient behaviour change at the end of the IYP programme.
- Investment should be made into a national effectiveness study using the Incredible Years National Register data held by the Ministry of Education.
- The present consortium structure for evaluating programmes targeted at childhood conduct problems should be maintained.

Recommended future research

- A longer-term follow-up of the children and parents in the present trial to examine the extent to which IYP programme benefits are sustained and the extent to which these benefits may dissipate over time.
- A national effectiveness study based on the Incredible Years National Register data held by the Ministry of Education to investigate whether or not the effect sizes for the changes in the ECBI and SCS measures obtained in the Incredible Years Pilot Study can be replicated on national aggregated data and for individual courses at all sites.
- The performance benchmarks reported here provide measures against which the future roll out of IYP programmes should be judged. Consistent and rigorous quality control monitoring of provider outcomes will ensure the programme is effective and the funding is well spent.
- Further investigation is required to understand the circumstances that produced site variation in outcomes to inform best practice principles and to ensure quality control.
- The suggestions that the benefits of IYP training may be slightly larger for non-Māori than for Māori families and that there appears to be a particular challenge for Māori in maintaining behaviour change point to the need for further research to examine Māori/non-Māori differences in greater detail.
- An evaluation of Ngā Tau Miharo ō Aotearoa, the Incredible Years resources developed by the Werry Centre for Māori Group Leaders, is required to ascertain whether or not differences in outcomes for Māori and non-Māori families are evident in courses using these resources.
- The evaluation of the Incredible Years Teacher programme is needed to assess a school-based programme.
- Further research is required to compare the cost benefit of providing the IYP programme against the cost benefit of existing parenting programme.

Chapter 1:

Introduction



1.1 Background to the Incredible Years Pilot Study

This report describes an evaluation of the Incredible Years Parenting (IYP) Programme in New Zealand. IYP is a parent management training programme developed for parents of children with conduct problems. The effective treatment and management of conduct problems are a high priority for New Zealand's health, education, justice and welfare sectors and form part of a broader cross-government priority work programme to address the Drivers of Crime.

The Ministries of Education, Health and Social Development established a pilot study of the IYP programme under the Drivers of Crime Conduct Problems work stream to assess the programme in the New Zealand context. The results of the evaluation of this pilot will inform the ongoing development of services across government for conduct problems and guide decisions about whether or not to invest significant funding in rolling out the IYP programme on a population basis in New Zealand. If the programme is proved to be effective, this investment has the potential to make a significant difference in reducing the rates of conduct problems and the long-term social distress and high levels of fiscal costs that conduct problems incur.

The Incredible Years Pilot Study evolved from the following influences:

- The Advisory Group on Conduct Problems (AGCP) was set up in 2007 in response to a growing awareness of the research findings on the adverse outcomes associated with childhood conduct problems. The AGCP was charged with the responsibility of identifying the issues raised by childhood conduct problems and of recommending effective programmes and policies for the prevention, treatment and management of these conditions.
- The advice provided by the AGCP formed the foundations of the second influence on the Incredible Years Pilot Study, the Positive Behaviour for Learning (PB4L) strategy developed by the Ministry of Education. This strategy includes the implementation of three evidence-based programmes for addressing childhood conduct problems, one being the IYP programme.
- The third element was the desire to develop a new model to evaluate government-funded programmes. This collaborative model was based on a consortium approach that brought together the skills of several groups: i) the service providers who delivered the programme; ii) the government researchers who evaluated the programme; iii) academic advisors with expertise in research design and analysis; and iv) Māori researchers.

1.2 Childhood conduct problems

A growing body of research evidence, both in New Zealand and overseas, has built a clear picture of the prevalence, consequences and treatment of childhood conduct problems:

- Conduct problems are considered to be one of the most commonly occurring mental health issues among children and adolescents. Based on multiple-informant data, conduct problems are estimated to afflict up to 10 percent of children in New Zealand (Fergusson et al 2011) and internationally (Scott 2007). Māori children have higher rates (15% to 20%) of conduct problems than non-Māori (Ministry of Social Development 2009a).
- Conduct problems during early and middle childhood are associated with negative outcomes in adolescence and adulthood, including antisocial behaviour, mental health difficulties, suicidal behaviours, substance abuse, teenage pregnancy, inter-partner violence and physical health (Fergusson et al 2011; Lindsay et al 2011). The difficulties and stresses caused by conduct problems affect both the young people and those associated with them such as parents, teachers and peers (Kazdin 2007). A recent review of the evidence on conduct problems by the AGCP (Ministry of Social Development 2009a) concluded there is probably no other commonly occurring childhood and adolescent condition that has such far-reaching consequences for individual health, development and wellbeing. Moreover, the wide range of negative outcomes associated with conduct problems has high social and fiscal costs (Fergusson et al 2005a; Feinstein and Sabates 2006; Odgers et al 2007).
- Research into effective interventions for the prevention, treatment and management of conduct problems is the subject of a number of reviews and meta-analyses that have identified effective programmes (Brestan and Eyberg 1998; Kazdin 2007; Ministry of Social Development 2009b). Among these are parent management training programmes based on social learning theory, such as the IYP programmes, Triple P and programmes developed by the Oregon Social Learning Centre.

1.3 Early intervention

Early intervention is recognised as a crucial element in reducing the onset of behaviour problems that often start in early childhood. A body of robust evidence shows early intervention can have a significant impact on child development and later life outcomes (Fergusson et al 2005a; Karoly et al 2005; Odgers et al 2007). O'Neill (2009) notes that early childhood interventions, particularly at the family level, have proven effective in overcoming long-term disadvantage.

The enduring influence of parenting during the early years in a child's life is demonstrated in a longitudinal study that followed a sample of American children from birth to 12 years of age examining how parenting in the infant, toddler and preschool years might influence children when they are older (Belsky et al 2007). The quality of parenting experienced across the first 4.5 years of life consistently predicted many aspects of children's functioning even after many alternative sources of influence were accounted for, such as family economic and educational resources. Children who experienced parenting that was warm, sensitive, cognitively stimulating and not intrusive or overcontrolling early in life showed better cognitive functioning, academic achievement and social adjustment when in middle primary school. The opposite was true when children experienced care that could not be characterised this way.

Recognition that the early years lay the foundation for future development has led to investment in evidence-based prevention and treatment programmes for young children and their families (Gluckman 2011). These early intervention programmes seek to mitigate risks for vulnerable children by improving parental capabilities, addressing risk factors and enriching children's experiences. In the New Zealand Families Commission review of early interventions, Gray (2011) points out that programmes and services providing support early have stronger effects than those that intervene later. In the case of antisocial behaviour, conduct problems that first become evident early in life are particularly difficult to correct if left untreated until adolescence. Moreover, remedial programmes in the adolescent and young adult years cost much more to produce the same level of attainment in adulthood (Heckman 2006).

1.4 Parent training programmes

Given the importance of family and parenting characteristics in the development of preschool children, parenting programmes are generally viewed as an essential component of early intervention. Parent-based intervention strategies focus on enhancing parental skills in the expectation these enhancements will translate into improved child outcomes. For example, ineffective parenting and poor disciplinary practices are major determinants of conduct problems (Scott 2007). The parenting behaviours that constitute risk factors for the development of conduct problems are the main targets of interventions that seek to improve parenting skills and thus to improve outcomes for children (Dadds et al 2003).

Parent management training is one of the most successful approaches to addressing conduct problems in early and middle childhood, particularly for children aged 3–7 years. There is strong evidence from rigorous efficacy trials that parenting programmes can improve parenting skills and reduce children's behavioural difficulties (Lindsay et al 2011). A range of manualised, well validated and widely used programmes are available. These programmes offer options for delivering parent behaviour management training, from universal programmes to highly intensive programmes, provided in a group or individually (Fergusson et al 2011).

Parenting programmes that focus on teaching specific parenting skills and/or changing parents' attitudes and beliefs can be effective at improving parenting and can lead to improved child behaviour (Hendricks and Balakrishnan 2005). Typically, these programmes aim to strengthen parent-child relationships and bonding, to strengthen parents' interpersonal skills and support networks and to promote effective limit setting, non-punitive discipline and systematic behaviour plans. Programme strategies include increasing parents' empathy towards their children and increasing parents' knowledge of child developmental needs.

1.5 Cost effectiveness

Childhood conduct problems are a major predictor of lifetime resource use which results in substantial costs in the education, health, justice and welfare sectors. A wider uptake of evidence-based interventions is likely to lead to considerable economic benefits in the short term, and probably even more in the long term (Romeo et al 2006). Heckman (2006) maintains that, given the dynamic process of life cycle skill formation in which early inputs strongly affect the productivity of later inputs, the economic return from early intervention is higher and the return from later intervention is lower.

Research shows the return from well-implemented and well-evaluated prevention, intervention, and treatment programmes for conduct problems is often very good, with programmes returning several times their costs as a result of reduced rates of crime, imprisonment and associated costs (Fergusson et al 2011). O'Neill's (2009) cost-benefit analysis of the IYP programme suggests the long-term rate of return from parenting programmes is likely to be relatively high, while Scott (2007) estimated the longer term return from IYP training to be 10 times higher than its cost.

Although there is no guarantee cost-benefit analyses conducted overseas will apply in the New Zealand context, there is a universal consensus in the literature that a long-term investment strategy is likely to be highly cost-effective, providing the investment is made in well-founded and well-implemented evidence-based programmes (Fergusson et al 2011).

1.6 The New Zealand context

Because of the unique socio-demographic and cultural profile of the New Zealand population it is important to ensure a treatment for conduct problems, such as parent management training, is effective in the New Zealand context. The higher rates of conduct problems in Māori children mean that, to reduce conduct problems in New Zealand, it is critical parenting interventions are effective for and acceptable to Māori. This imperative highlights the need for Māori to engage in programmes that build resilience and improve outcomes for whānau. This is more likely to be achieved if there is a range of intervention options available for Māori, including kaupapa Māori interventions, Māori-enhanced interventions, and generic interventions.

Ongoing concerns are expressed about the suitability for, and acceptability to, Māori of programmes based on western science methodologies (Altena and Herewini 2009; Cargo 2008; Durie 2004, 2006). Although some evidence-based parenting programmes, such as the IYP programme, have been culturally enhanced to reflect local Māori tikanga (eg Werry Centre 2012), these should be distinguished from kaupapa Māori interventions, which are designed, developed and delivered by Māori.

MacFarlane (2011) notes that, while the literature is growing on the development of kaupapa Māori research, less consideration has been given to the ways in which western science and kaupapa Māori research can be combined to produce consensual decisions about programme effectiveness. He sets out a conceptual model that attempts to integrate western science and kaupapa Māori models of programme development and evaluation. The analogy is a braided river (he awa whiria) in which there are two main streams, representing western science and kaupapa Māori models, which are interconnected by minor tributaries with the two streams reaching a point of convergence.

MacFarlane's model acknowledges the western science and kaupapa Māori streams as distinctive approaches to the development and evaluation of programmes but permits knowledge and evaluation methodologies from each to inform and to be applied to the development of programmes in the other. Furthermore, the model assumes the determination of programmes as being effective will rely on an acceptance of evidence from both streams.

1.7 The report structure

This report is structured as follows:

- **Chapter 2** gives an overview of the IYP programme and a brief summary of existing evaluations.
- **Chapter 3** provides a description of the research design.
- **Chapter 4** describes recruitment and retention of the study participants and presents the participant profile.
- **Chapter 5** summarises the findings related to child behaviour and parenting practice and relationships.
- **Chapter 6** summarises the findings related to other outcomes.
- **Chapter 7** summarises the findings related to Māori and non-Māori families.
- **Chapter 8** summarises the findings related to Single Case Studies.
- **Chapter 9** addresses programme fidelity, Group Leader feedback and parent satisfaction.
- **Chapter 10** discusses the policy implications of the findings and presents Evaluation Advisory Group (EAG)¹ recommendations.
- **Chapter 11** provides a summary of findings and a recommended future research programme.

A Technical Appendix to this report (Gray 2013) contains detailed descriptions of the study instruments and statistical analyses. This report is available on request from the author.

1 See Appendix 1 for a list of the members of the EAG.

Chapter 2:

The Incredible Years Parenting programme



2.1 An overview of the Incredible Years Parenting programme

The Incredible Years Parent, Teacher and Child Training series has been developed over the last 30 years at the University of Washington by Carolyn Webster-Stratton and her associates (Webster-Stratton 1994). It has been implemented widely within the United States, the United Kingdom, Canada, Ireland, Norway, Denmark, Sweden, Australia and New Zealand. The Incredible Years series received a proven rating for IYP BASIC from the RAND Corporation (RAND 2007), has been endorsed in a number of jurisdictions (MSD 2007) and has been identified as one of 11 Blueprint interventions by the Centre for Violence Prevention at the University of Colorado, having satisfied stringent scientific criteria (Jones et al 2007).

The Incredible Years series consists of a number of comprehensive, multi-faceted and developmentally-based training programmes: five for parents, two for children and one for teachers. The programmes are specifically designed to promote emotional and social competence in young children and to prevent, reduce and treat behavioural and emotional problems.

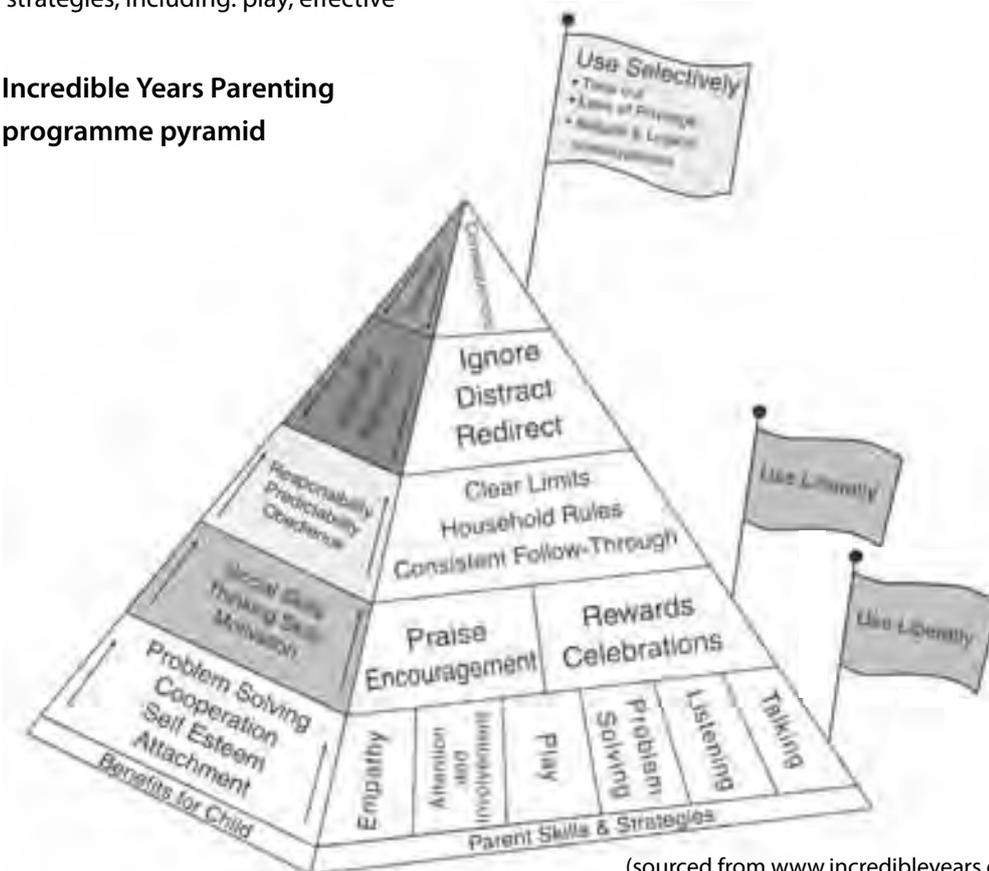
The IYP programme is based on social learning theory, particularly the coercion hypothesis of reinforcement developing and maintaining deviant behaviour and the notions of modelling and self-efficacy. It is an intensive programme, with two trained Group Leaders delivering a 2–2.5 hour session each week, typically to a group of up to 16 participants. Each programme lasts for 12–18 weeks. The programme uses a collaborative approach between Group Leaders and parents, including group discussions of video material showing both appropriate and inappropriate parent-child interactions. The IYP preschool programme is for parents of children aged 3–5 years; the school age IYP programme is targeted at parents of children aged 6–12 years.

The IYP training aims to increase positive and nurturing parenting styles, to decrease negative or harsh parenting, and to promote children’s social-emotional development. It covers a range of parenting strategies, including: play, effective

praise, motivating children through rewards, effective limit setting, and effective ways of handling misbehaviour. Parent competences are developed in areas such as communication, limit-setting, problem-solving and anger management. Parents are also encouraged to develop support networks. The programme covers all the basic aspects of parenting as well as offering help to individual parents who find their child’s behaviour challenging.

A central metaphor used in the IYP programme is a pyramid of parenting principles (see Figure 1). The pyramid is used to depict a hierarchy of parental strategies, some of which are the basic building blocks to help children develop to their potential and should be used very frequently (eg play, attention and involvement), and others which are also important for optimal child development but should be used less frequently (eg consequences for aggression).

Figure 1. Incredible Years Parenting programme pyramid



(sourced from www.incredibleyears.com)

2.2 Evaluations of the Incredible Years Parenting programme

There is a strong evidence base for the effectiveness of the IYP programme in enhancing parenting competencies and in reducing disruptive behaviours in children, and these gains have been maintained at long-term follow-up. Webster-Stratton and Reid (2010) reviewed the numerous published randomised control group trials (RCT) conducted by Webster-Stratton and her colleagues that have demonstrated the effectiveness of the IYP programme for young children with conduct problems (eg Webster-Stratton and Hammond 1997). These findings have been replicated in RCTs by independent investigators in England (Gardner et al 2006; Scott et al 2001), Wales (Hutchings et al 2007; Jones et al 2007), Ireland (O'Neill 2009) and Norway (Larsson et al 2009). Studies using a matched control group rather than a RCT (eg Kelleher and McGilloway 2006; Posthumus et al 2011) have also demonstrated the effectiveness of IYP programmes in reducing childhood conduct problems.

Table 1 summarises a sample of RCT studies that report significant effect sizes for changes in child behaviour at post-treatment following the parent's attendance at an IYP programme. Effect size², as measured by Cohen's 'd', provides an indication of the extent of reported change, whereby an effect size of $d=.20$ is considered small, an effect size of $d=.50$ is medium and an effect size of $d=.80$ is large. Two studies reported a moderate effect size and three studies reported large effects for child behaviour change. Effect sizes for measures of parenting practices are presented in Table 2 for those studies that reported these outcomes. Here, the effect size for Positive Parenting was small, while changes in the other parenting practices ranged from medium to large. The children in these studies were in the clinical range for conduct problems.

Table 1. A sample of RCT studies investigating the efficacy of IYP programmes, with post-treatment child behaviour effect sizes (ES)

Study	N	Child age (years)	Country	Measure	ES (d)
Gardner et al (2006)	76	2 to 9	England	ECBI	$d=.55$
Hutchings et al (2007)	153	3 to 5	Wales	ECBI	$d=.89$
Larsson et al (2009)	127	4 to 8	Norway	ECBI	$d=.65$
Scott et al (2001)	141	3 to 8	England	PR	$d=.89$
Webster-Stratton and Hammond (1997)	97	4 to 8	US	CBCL	$d=.89^*$

ECBI = Eyberg Child Behaviour Inventory, Intensity Score
 CBCL = Child Behaviour Checklist
 PR = Parent Report of Conduct Problems

*Effect size for Webster-Stratton and Hammond (1997) retrieved December 2012 from <http://www.incredibleyears.com/library>

2 Effect sizes are explained in Chapter 3, Section 3.3 and in Appendix 3.

Table 2. A sample of RCT studies investigating the efficacy of IYP programmes, with post-treatment parenting practices effect sizes (ES)

Study	N	Child age (years)	Country	Design	Measure	ES (d)
Gardner et al (2006) Positive Parenting Negative Parenting Parent Competencies	76	2 to 9	England	RCT	Observation Observation AOPS	d=.38 d=.74 d=.65
Hutchings et al (2007) Positive Parenting Parent Competencies	153	3 to 5	Wales	RCT	Observation AOPST	d=.57 d=.95
Larsson et al (2009) Positive Parenting Harsh Discipline Inconsistent Discipline	127	4 to 8	Norway	RCT	PPI PPI PPI	d=1.44 d=.61 d=1.05

AOPS = Arnold-O'Leary Parenting Scale Total
PPI = Parenting Practices Interview

There remains the question of whether evidence-based programmes developed outside of New Zealand are effective within a New Zealand context and, in particular, are effective with young Māori or Pacific people (Smith 1999). These debates highlight the importance of a thorough evaluation, including of cultural appropriateness, of programmes such as the IYP programme in New Zealand before they are widely implemented (Fergusson et al 2011). Further, as Aos et al (2006) notes, programme efficacy must be demonstrated to ensure public funding is being well spent.

A preliminary evaluation of IYP training in New Zealand (Fergusson et al 2009) found that parents attending the IYP courses reported

reductions in problem behaviour and increases in child social competence following the programme, and that client satisfaction with the programme was high. Similar findings were obtained for Māori and non-Māori parents. A second study, based on a qualitative evaluation of the responses of 10 Māori parents attending a marae-based version of the IYP courses (Altena and Herewini 2009), also reported that Māori enrolled in the IYP programme showed high rates of retention and satisfaction with the programme.

While these findings were promising, a more searching evaluation of the IYP programme in New Zealand and its applicability to Māori parents was required.

Chapter 3:

The New Zealand Incredible Years Pilot Study



On the recommendation of the AGCP, the Ministries of Education, Health and Social Development, with advice from academic experts, established the Incredible Years Pilot Study to assess the efficacy of the IYP programme in the New Zealand context. The Incredible Years EAG advising the study consisted of representatives of these Ministries, plus academic experts in the fields of conduct problems and research. The results of the evaluation detailed in this report will inform the further development of existing IYP programmes and guide the decisions about the investment of significant funding to roll out the IYP programme on a population basis in New Zealand.

3.1 Research design

To inform the future planning and delivery of the IYP programme, an evaluation must provide agencies with information on the profile of families referred to IYP training, demonstrate programme fidelity, measure both programme effectiveness and parent satisfaction, and assess programme responsiveness to Māori. To this end the Incredible Years Pilot Study was designed as a substantial 2-year, multiple-informant study that included mixed measurement methods, single case studies and a 6-month follow-up. The study received ethical approval from the Multi-region Ethics Committee, which reviews national and multi-regional studies (ethics reference MEC/10/08/075).

Research questions

The design of the Incredible Years Pilot Study was based around the following series of research questions relating to the delivery of the programme in New Zealand.

1. *Programme efficacy: To what extent was there evidence to suggest that the IYP programme was effective in a New Zealand context?*

This question was examined using three research approaches:

- a. The first approach was a before-and-after longitudinal design that compared the outcomes of children and families receiving the IYP programme at: baseline, mid-programme and programme completion. The main dependent variables were: parent reports of child behaviours; parent reports of parenting behaviours; measures of family functioning and parental mental health. Consistent with previous research it was hypothesised there would be improvements in these outcomes with parent participation in the IYP programme.
- b. The second approach involved collecting teacher report information on child behaviour at: baseline, mid-programme and programme completion.
- c. The third approach was a series of single case studies in which child behaviour outcomes were assessed at weekly intervals over the course of the programme. The main purpose of this component of the research was to examine the various trajectories followed by children whose parents were enrolled in the IYP programme.

It was proposed to supplement these three tests of programme efficacy with data from a direct observation study using a single case design but, for reasons discussed later, this approach proved to be non-informative.

2. *Sustainability of gains: To what extent were any benefits of IYP sustained over time?*

This question was addressed by extending the longitudinal, teacher and single case study design described above to include follow-up data at 6 or 7 months after course completion.

3. *Benefits for Māori: To what extent were the programme outcomes for Māori similar to or different from the outcomes for non-Māori?*

The extent to which programmes such as IYP that are based on western science methodology are effective for Māori has been the subject of substantial debate (eg Altena and Herewini 2009). The current consensus is that IYP training is effective for Māori providing the programme is delivered in a culturally appropriate way (eg Berryman et al 2009). To address these issues in the Incredible Years Pilot Study, analyses were conducted comparing:

- a. The outcomes for Māori and non-Māori families on the measures of IYP programme effectiveness described in 1a and 1b, and
- b. Measures of parent satisfaction gathered at programme entry and at programme completion.

4. *Programme entry measures: To what extent do the outcomes of the IYP programme vary with the child's behavioural adjustment at baseline?*

Special Education policy is to accept referrals from all sources without screening. They operate on the basis of IYP programme provision within a non-stigmatising community setting. However, this policy is not empirically based. To test this policy it was decided to examine the extent to which IYP programme outcomes as assessed in the design described in 1a and 1b varied with the child's scores on the ECBI at baseline.

5. *Between site differences: To what extent do the outcomes of IYP vary with the site at which the programme was delivered?*

As described below, the study was conducted at three Special Education sites. Comparisons of the outcomes at the three sites were investigated to examine the extent of between site variability in IYP programme outcomes.

6. *Client satisfaction: To what extent do the parents who received the IYP training see the programme as: a) helpful; b) appropriately organised; c) effective in dealing with their child's behavioural problems; d) culturally appropriate?*

To examine these issues parent satisfaction data were gathered at the baseline interview, the mid-course interview, the end-of-course interview and the 6-month follow-up.

The research-designated IYP programmes

The IYP programmes evaluated in this study were delivered by the Ministry of Education Special Education in 2011 at three sites: Bay of Plenty, Canterbury and Mid-Central. Four 18-week courses were offered at each site in two phases, two in the first half of the year (Phase 1) and two in the second half (Phase 2), yielding 12 courses in all. Parents were referred to IYP training because they had a child aged 3–8 years with conduct problems. Special Education did not make any changes to their usual practice in organising and delivering the research-designated courses. The IYP programmes evaluated in the Incredible Years Pilot Study were run in the same manner as Special Education IYP programmes not included in the research.

The research was conducted following a 3–5 year period in which Special Education was building capacity and capability for delivering IYP programmes throughout New Zealand. This entailed ensuring fidelity, procuring resources, building Group Leader skills and national data collection processes to evaluate outcomes and to establish ongoing monitoring and quality control as part of normal service delivery.

Participants

All parents enrolled in these IYP courses at the three sites were invited to take part in the evaluation; no parents were excluded from the research. The minimum sample size was set at 150 parents, 50 at each site. A total of 166 parents (Bay of Plenty=56, Canterbury=57 and Mid-Central=53) agreed to participate and completed the baseline interview. Written consent was obtained at each stage of the study. Participants received a grocery voucher for each completed interview as reimbursement for their time.

Preliminary research

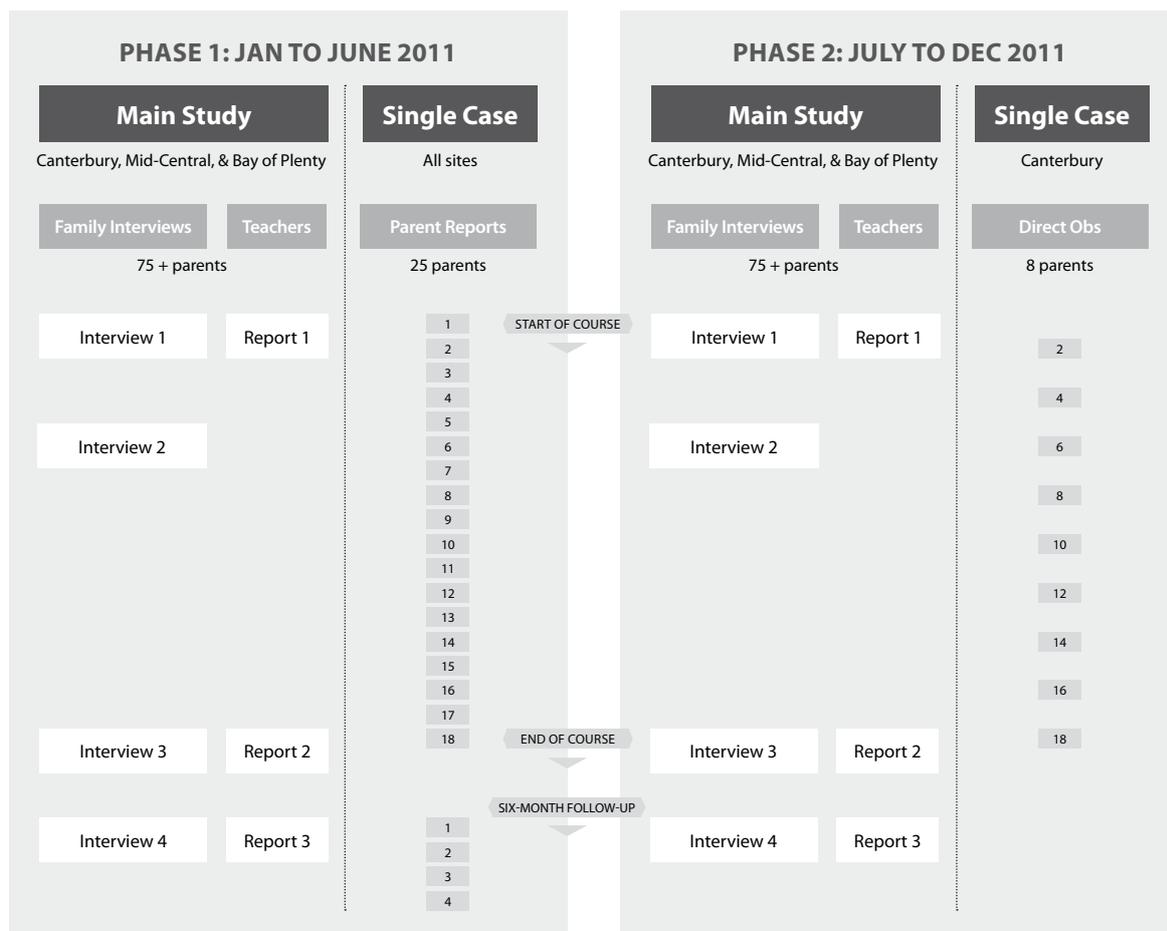
Preliminary research assessing client uptake and programme fidelity in the delivery of IYP training in Special Education courses was carried out from June to December 2010. Eleven courses across four sites took part: Bay of Plenty (3), Waikato (3), Central (3), and Canterbury (2). Programme fidelity documentation indicated the IYP programmes were delivered as specified by the Incredible Years manual. Although issues with the administrative data on the Special Education Incredible Years National Register meant historic client uptake data could not be ascertained, the problems were resolved so that reliable monitoring data were available for the main study.

Research design

The main study was a repeated measures design in which all research participants were interviewed four times: at baseline before the IYP course began, mid-course, post-course and 6-month follow-up. Family Interviews were supplemented by a postal Teacher Questionnaire administered at baseline, post-course and follow-up. Data for the Family Interviews and Teacher Reports from the two phases of the study were combined.³ The study also included Single Case Studies using weekly Parent Reports in Phase 1 and a trial of Direct Observations in Phase 2. Figure 2 contains a diagram of the research design.

3 Analyses of outcomes by phase are reported in Chapter 6, Section 6.3.

Figure 2. The research design



Family Interviews

Trained Interviewers employed for the Incredible Years Pilot Study conducted the four home-based interviews with the primary caregivers. Each interview lasted approximately 1–1.5 hours and covered a range of topics relating to child behaviour, parenting practices and relationships, and the family context (Table 3). The Family

Interview incorporated items from a number of previously validated measures to assess changes in outcome variables over the course of the study. Appendix 2 describes the source instruments that comprised the Family Interview; a Technical Appendix (Gray 2013) provides more details about the instruments.

Table 3. Topics covered in the Family Interview

Sections	Topic
A	Family composition
B	Child schooling
C to J	Child behaviour (aggression, defiance, hyperactivity, self-control, anxiety/withdrawal, friendships, social skills)
K	Child health
L & M	Parenting practices, dealing with misbehaviour, child management (primary caregiver)
N	Parenting practices, dealing with misbehaviour, child management (partner)
O	Partners and partner relationships
P	Primary carer's health/depression
Q	Stressful life events
R	Standard of living Family problems Work, education, living circumstances Ethnicity and cultural participation
S	Expectations of/satisfaction with Incredible Years

Note there was a process in place if any questions in the interview caused distress to the participants, or if the participants wanted help with an issue uncovered by the interview. Neither circumstance eventuated.

Teacher Questionnaires

Childcare, preschool or class teachers of the children whose parents gave permission were asked to complete a standardised behavioural assessment for that child. The assessment was a modified version of the Canterbury Social Development Scale (Church et al 2006). The questionnaires were posted to teachers at the beginning of the IYP course and again at course completion and the 6-month follow-up. Two forms were used: one with wording appropriate for preschoolers and one with wording appropriate for primary school children.

Single Case Studies

Parent Reports: Family Interviews were supplemented in Phase 1 at all three sites by Single Case Studies using weekly Parent Reports. Participants for this component of the study were randomly selected from those enrolled in the research programmes. A total of 25 primary caregivers completed weekly telephone Parent Reports, one for each of the 18 weeks of the IYP course, and again for 4 weeks at the 6-month follow-up. The Interviewers went through a checklist of 23 positive and 26 negative behaviours that may have occurred in the previous 24 hours.

Direct Observations: The measures described thus far are based on parent reports of behaviour. An attempt was made to supplement the self-report data with direct observations of parents and their children interacting. The intention was to trial the Direct Observations methodology in participants' homes in Canterbury in Phase 1 of the study, and to conduct these Single Case Studies at all three sites in Phase 2. Due to delays, including the Christchurch earthquake, the Phase 1 trial was not possible. Instead, the methodology and parent compliance were tested in a clinic setting in Phase 2 in Christchurch only. Because of the small sample of parents participating in this component of the study, and the fact it did not achieve what was intended, the findings are not included in this report. Church (2012a) summarises the Direct Observations and discusses lessons learned from the trial.

Programme fidelity

The Unified Protocol developed by Special Education assessed the fidelity of the IYP programme delivery from the point of parent recruitment through to the end of the delivery of the treatment. The following aspects were documented: recruitment, reducing barriers, programme delivery (number of sessions, order of activities, materials used, methods deployed, Group Leader training, and cultural enhancements), client retention, compliance, quality assurance strategies, and follow-up. Group Leaders completed the Unified Protocol checklists at three points throughout the set-up and delivery of the course: pre-programme, delivery and final. Supervisors and Group Leaders also provided reports of their IYP programme debriefs.

Administrative data

Parent referrals, uptake and retention were monitored using a Recruitment Form designed for the study and Special Education's administrative records captured in the Incredible Years National Register.

Special Education maintains a National Register for each IYP course containing the demographic details of the parents and the pre-course ECBI and Social Competency scores for the children. The Group Leaders recorded a variety of information on the Incredible Years National Register, including: how many, and which, sessions the parent attended; whether she/he withdrew and at what point in the programme and the reason; whether the Group Leaders did a make-up session for parents unable to attend a particular session; and any additional home visits made to support parents or any support parents were receiving to attend via a contribution to childcare or transport.

Māori in-depth study

A separate study examined Māori perspectives of the IYP programme using a kaupapa Māori approach that drew upon kaupapa Māori research methodology. The researchers obtained the views of four Māori Group Leaders and four Māori parents who had been involved with the IYP programmes that were part of the evaluation to assess whether or not this mainstream programme met their needs. Māori researchers contracted by the Ministry of Education carried out this study (Berryman et al 2012).

3.2 Fieldwork issues

Programme scheduling

The scheduling of two rounds of 18-week IYP research-designated courses in 1 year caused some logistical problems for the providers and the researchers, particularly in Phase 2 in the second half of the year.

The tight 2-week turnaround between Phase 1 and Phase 2, which also coincided with school holidays, meant there was not the 4–6 week lead-in for home visits in Phase 2. This resulted in some delayed parent referrals to the Interviewers. The tight timeframe also had an impact on the scheduling of the interviews, notably the third interview for Phase 1 and the baseline interview for Phase 2. With precedence given to the latter, there were delays in completing the former. This meant some interviews were completed several weeks after the course had finished. The Phase 2 courses finished at the start of the summer school holidays in mid-December, a difficult time to engage parents in the research. To be consistent with the timing of interviews in Phase 1, these Phase 2 end-of-course interviews were deferred to mid-January.

The earthquake in Christchurch in February 2011 meant the courses in Phase 1 were delayed 1 month and it was not possible to complete all of the baseline interviews before the IYP programme began. Finding suitable venues was also an issue. The disruption resulting from the earthquake created considerable problems both for the Group Leaders organising the IYP courses and for the parents attending. However, operating under these trying conditions has added to the ecological validity of the study findings for Canterbury.

The impact of these delays, if any, on the estimates of treatment gains is not known. They do, however, reflect the settings in which the programmes are usually delivered.

For the most part Interviewers reported little trouble in scheduling appointments and completing the interviews, although some second and third appointments were necessary. The shorter lead-in for the courses in Phase 2 meant some of the baseline interviews were more difficult to arrange than those in Phase 1. Most of the interviews took between 1–1.5 hours to complete. No incidents were reported.

Evaluation team location

The evaluation team was located in Wellington, some distance from the three research sites. This physical separation compounded the complexity of managing the fieldwork. A local project manager in each site would have been beneficial.

3.3 Statistical analyses

The main analyses are based on examining the outcomes of the IYP programme for the parents and their children. These outcomes are determined by comparing the measures of child behaviour, parenting practice and other measures observed at baseline with the same measures observed again at mid-course, course completion, and at a 6-month follow-up. All data are based on self-reports provided by parents and teachers. Data from the two phases of the Study were combined.⁴ The analyses were carried out on an intention-to-treat research design, which includes parents who had attended the parenting programme and those who had not or who had dropped out.

Any changes in the average scores of measures at each observation time were statistically tested under the hypothesis there was a straight-line or linear trend of change in the scores from baseline, through mid-course, to course completion. The maintenance or sustainability of behavioural change was also tested under the hypothesis there was a statistical difference in the mean scores observed between baseline and at the 6-month follow-up. A probability value (or p-value) is provided by the statistical tests for each measure of behaviour. If the p-value is small then the probability there was no linear trend of change or no difference in the mean scores is small. As a standard rule, p-values smaller than the $\alpha=.05$ level were accepted as evidence of a linear trend of change in average scores.

Effect sizes

In addition to the statistical test of changes in behaviour through the course, the size of those changes is also examined. It is the size of the outcome that is of key interest. The size of the changes in behaviour demonstrates the size of the programme outcomes, which is expressed using 'effect size' estimations. The effect size

estimate is the standardised difference between means or proportions, expressed as Cohen's 'd'. (Cohen 1977). Cohen suggests that an effect size of $d=.20$ is small, an effect size of $d=.50$ is medium, and an effect size of $d=.80$ is large. In real-world terms, the relationship between smoking and lung cancer is large, whereas the relationship between life stress and depression is small. These interpretations are arbitrary but provide an indication of the extent of the behaviour change. A positive effect size ($d>0$) indicates improved behaviours while a negative effect size ($d<0$) indicates worsened behaviours.

Testing for interactions between sub-group pairs

The programme outcomes or effects as demonstrated by behaviour change over time may be different between sub-groups of parents or children. Tables of programme outcomes by key dichotomous sub-groups (or pairs) were examined with the estimated effect sizes for each sub-group pair. These tables contain statistical tests for interactions based on a 2x2 factorial analysis, which compares the average scores at the two observations in time (eg baseline and course completion) by the two sub-group pairs of interest (eg Māori and non-Māori children).

The 2x2 factorial analyses are performed using Analysis of Variance models, which provide p-values representing the probability there was no interaction between the time of observation and the sub-groups. P-values smaller than the $\alpha=.05$ level suggest there was an interaction, which essentially suggests the two sub-groups of parents or children had different effect sizes.

4 Analyses of outcomes by phase are reported in Chapter 6, Section 6.3.

Adjustments to multiple statistical tests

Each programme outcome and outcome sustainability table examines behaviours with more than one measure. For instance, Child Behaviour was assessed with eight measures and a statistical test of significance for each measure. Multiple significance testing increases the chance of false-positive conclusions where stated significant differences in average scores occurred by chance. To minimise the chance of false-positive conclusions, the $\alpha=.05$ level for p-values was adjusted using a Bonferroni adjustment where the $\alpha=.05$ level was divided by the number of tests made in each table. For example, $\alpha=.05$ level was divided by eight for the Child Behaviour tables, which means only p-values smaller than .006 were considered significant.

The outcomes by sub-group tables contain multiple statistical tests for interactions for each individual measure. Each table was further supported by a 2x2 factorial analysis where all the measures in a table were analysed as one multivariate. These tests were done using a Multivariate Analysis of Variance, which provided a p-value that represents the probability there was no interaction between the observation times on the overall measure and the two sub-group pairs.

Profile tables

The differences between parent and child profiles discussed in this report were verified using chi-square statistics on contingency tables where p-values were less than $\alpha=.05$.

Details of the statistical models used

The statistical models used in this study are described in Appendix 3, with more details available in the Technical Appendix (Gray 2013).



Chapter 4:

Recruitment, retention and participant profiles



This chapter describes the referral process to Special Education IYP courses, the recruitment and retention of parents to the research and the characteristics of the children, primary caregivers and households of the participant families.

4.1 The referral process to Special Education Incredible Years programmes

In the Bay of Plenty the local district health board (DHB) handled the wait-list for the IYP programmes delivered by all providers in the region. Parents were referred to the DHB, who then sent letters to everyone on the wait-list when courses became available. The DHB forwarded a list to Special Education of only those parents interested in an IYP course delivered by Special Education. In Mid-Central and Canterbury the wait-lists did not go through the DHB, but went directly to Special Education.

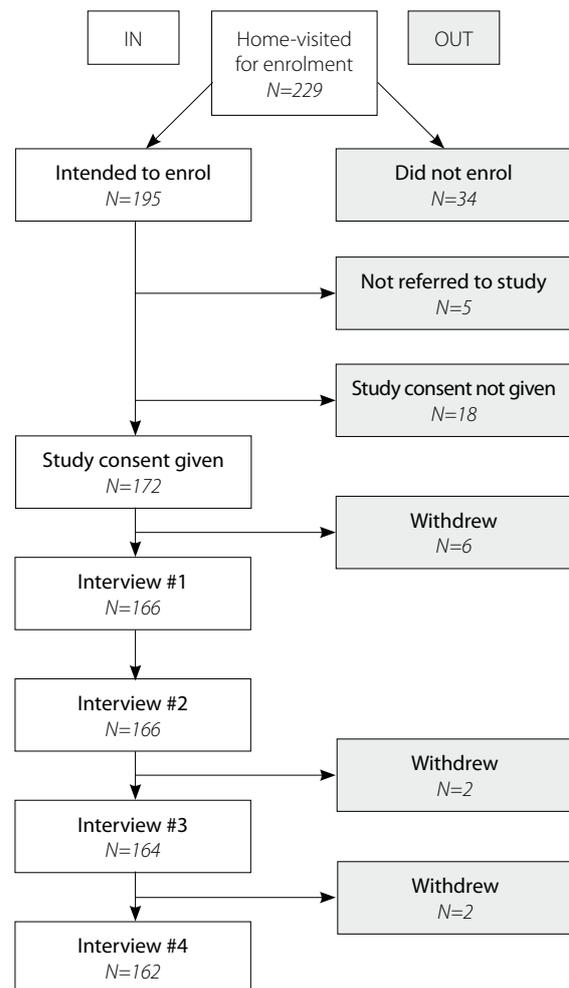
Special Education IYP Group Leaders contacted parents from the wait-list in chronological order when an IYP programme became available. They aimed to recruit 20 parents for each course to allow for drop-outs. Interested parents were visited at home. At this time the programme and time commitments were explained, the concerns for her/his child discussed and any barriers identified (eg transport or childcare). If, after this, the parent intended to enrol in the IYP programme she/he completed the consent forms and the pre-course measures. For the IYP programmes included in the Incredible Years Pilot Study, the research was explained and the consent to be contacted by the interviewer was sought.

4.2 Research participant recruitment and retention

Following the home visit, Group Leaders provided the names and contact details of those parents who gave consent to be contacted for the research to the Interviewers who then arranged an appointment to conduct the baseline interview. Figure 3 contains a flow diagram accounting for all parents the Group Leaders visited through to course enrolment and research participation. Research participation rates were calculated both as a proportion of families home-visited (Invitation to Treat) and as a proportion of families enrolled in an IYP programme (Intention to Treat). Group Leaders completed a Recruitment Form that documented the number of families contacted for the IYP course, those who did not enrol in the course, and those who did not consent to be contacted for the research.

The Group Leaders obtained parental consent to be contacted for the research from 75 percent of those home-visited to enrol in the IYP programme (Invitation to Treat) and 88 percent of those who actually did enrol at the time of the home visit (Intention to Treat). Seventy-two percent of the Invitation to Treat parents and 85 percent of the Intention to Treat parents completed the baseline interview. Of these 166 parents, 162 (98%) remained in the research through the follow-up interview. Recruitment and retention rates were similar at all three sites.

Figure 3. Incredible Years Pilot Study participant recruitment and retention



Parent loss from the Incredible Years Pilot Study up to the point of IYP course enrolment was relatively high (25%), whereas the loss was very low (4%) after enrolment. Of the 57 parents the Group Leaders visited who did not take part in the research, most (34) decided not to enrol in the programme while a further five were not offered the research due to an administrative error. Most of the parents who had expressed interest in the IYP programme when contacted but did not enrol stated that their family circumstances had changed or that they were no longer interested. Eighteen of the parents intended to enrol and were invited to participate in the research but declined to be involved with the study. Nearly all of these parents gave no reason for not taking part in the study although a few felt it would be too intrusive or they were too busy.

While the high retention rate of the 166 parents in the study indicates good internal validity for the Incredible Years Pilot Study, the extent to which the findings apply to the 229 parents contacted for the IYP programme is not known. This raises issues about external validity. Data were not available to assess whether or not the study sample was representative of all parents referred for IYP training, although it is not unreasonable to assume it would be similar.

4.3 Participant profiles

The participant profiles presented are as at the baseline interview for the 166 children and parents participating in the Incredible Years Pilot Study. Similar client profiles have been reported in other Incredible Years research (eg Hutchings et al 2007, Larsson et al 2009; Scott et al 2001; Webster-Stratton and Hammond 1997).

The profile of participants shows the research reached the desired target group: the majority of the children were aged 3–8 years and one-third were identified as Māori by their primary caregivers. A comparison of the participants in Phase 1 and Phase 2 indicates that splitting the study into two phases did not introduce bias into the sample as the only significant difference between the two groups was with the occurrence of one stressful life event. Most of the significant differences found in the profiles of the parents at each of the three sites were as expected, such as the highest proportion of Māori parents being located in the Mid-Central region.

The children

Almost all (94%) of the children in the study were within the target age range of the IYP programme, with 40 percent aged 3–4 years and 54 percent aged 5–8 years. Just over a third (36%) were identified by their primary caregivers as Māori, and three-quarters (73%) were male. The distribution of ethnicity varied between the sites, with the smallest proportion (19%) of Māori children in Canterbury, over one-third (38%) in the Bay of Plenty, and more than half (53%) in Mid-Central.

Only one of the children in the study had not attended preschool and very few (8%) had been refused entry to preschool or had been stood down or suspended from primary school. Kindergartens, play centres, and general childcare facilities accounted for most (84%) of the preschools attended. There were differences in the types of preschools attended by site, with Canterbury having the highest proportion attending mainstream preschools/schools (95%), followed by the Bay of Plenty (81%) and Mid-Central having the lowest proportion (77%).

One in five (20%) of the children had contact with Special Education within the 3 months before the baseline interview and slightly more (23%) had seen a psychologist. A higher proportion of the children in the Bay of Plenty had seen a psychologist (39%) than in either Canterbury (18%) or Mid-Central (11%). About a third (37%) of the children had been to health professionals such as a GP, paediatrician or public health nurse. A minority had visited specialist health services such as an audiologist (13%), a speech/language therapist (13%) or an optometrist (10%). Specialist mental health services (10%), social workers (11%) and Barnardos (11%) were also mentioned. Significantly fewer participants used these services in Canterbury compared with those at the Bay of Plenty and Mid-Central sites.

The Primary caregivers

The great majority of primary caregivers were women (95%) and the natural mother of the participant child (92%). About half (55%) of the primary caregivers were currently married or partnered and the others were either single (33%), separated or divorced (10%) or living apart from a partner (2%). Two-thirds (64%) were aged 25–39 years, one in 10 (11%) were aged 19–24 years and a quarter (25%) were older than 40. However, a majority (60%) of the caregivers had their first child when they were younger than 24 years of age.

Nearly a quarter (23%) of primary caregivers identified themselves as Māori. Mid-Central had the highest proportion of Māori caregivers (40%), with the Bay of Plenty next (20%) and Canterbury with the lowest proportion (12%).

Just over half (54%) of the parents had no qualifications or school qualifications only, while 31 percent had a certificate or diploma and 12 percent had a degree. The highest level of education differed by site, with the Bay of Plenty having proportionally more educated primary caregivers and Canterbury having the fewest. Canterbury had twice as many primary caregivers with no education qualifications (49%) as the Bay of Plenty (25%) and Mid-Central (25%), whereas the Bay of Plenty had the highest proportion of degree-qualified primary caregivers (20%) compared with Mid-Central (9%) and Canterbury (7%).

A quarter (24%) of the study parents reported eight or nine of the nine depression symptoms measured, and a further 22 percent reported six or seven of the nine symptoms. A third of primary caregivers (36%) showed signs of mid to high levels of stress, and 13 percent showed medium to high levels of anxiety.



The households

Where the primary caregiver had a resident partner, the data for the respondent and the partner were combined to describe the household.

Two-thirds of the participants (66%) lived in households of four or more people, including siblings. About a quarter (23%) of the children had no siblings, while 40 percent had one sibling and 21 percent had two. Private rental (54%) was the most common type of accommodation, followed by mortgaged housing (30%) and state housing (10%).

About a third (36%) of the caregivers were single and unemployed, with a further 30 percent being part of a one-income couple. Employment was the main source of household income for over half (56%) of the participants and the Domestic Purposes Benefit provided the income for a third (34%). The majority of households in the study lived in areas at the medium to high (poverty) end of the New Zealand Deprivation Index, which is based on current census data. Eighty percent were in quintiles 3 (25%), 4 (26%) and 5 (29%) of the Index, where 5 is the most deprived. The three sites differed slightly in terms of the Index with Mid-Central having the highest proportion of participants living in quintile 5 (42%). In comparison, the Bay of Plenty had 26 percent in quintile 5 and Canterbury had 20 percent.

Four in five (82%) of the households experienced at least one stressful life event in the 3 months before the baseline interview. Two-thirds (67%) reported three or more such events. Serious illness, accident or death affected almost half (45%) of the households, serious financial problems were experienced by 39 percent, almost a third (31%) reported prolonged arguments within their family/extended family, and a quarter (25%) of the primary caregivers experienced prolonged arguments with their partners.

Households in Canterbury reported significantly fewer stressful life events than did those in the Bay of Plenty or Mid-Central. Fifty-one percent of Canterbury households reported at least three stressful life events compared to Mid-Central households (72%) and Bay of Plenty households (79%). Arguments in the family were more common in the first phase than in the second phase of the study, with 42 percent of households reporting arguments in Phase 1 compared to 19 percent in Phase 2. This was the only significant difference between the Phase 1 and Phase 2 samples.

Chapter 5:

Findings related to child behaviour, parenting practice and relationships



Central to the Incredible Years Pilot Study was the assessment of change in child behaviour and parenting practice following the IYP training. This chapter presents the findings of these before-and-after comparisons. The comparisons were made on all study participants irrespective of the extent to which they had participated in IYP courses. The child behaviour and parenting practice measures were based on self-reported data collected in the Family Interview administered before, during and after the IPY course and again 6 months later.

Family Interview data made up the core of the research.⁵ The outcomes of the IYP programme were determined by comparing the first (baseline) interview scores with the scores of the third (post-course) interviews. The second (mid-course) interview was used in the test for a linear trend. The trend p-value provides the evidence of a linear trend of change in the mean scores throughout the duration of the programme. Data from the fourth (follow-up) interview assessed whether or not any changes evident at course completion were sustained over the medium term. All changes referred to in Table 3 onwards are changes in mean parent or teacher reports.

5 Data re-entry checks produced an infinitesimal error rate across all items tested.

5.1 Child behaviour

Pre-test/post-test comparisons: Child behaviour

The instruments⁶

In the Family Interview, parents were questioned about their child's behaviour during the past 4 weeks using 111 items based on four recognised scales: the Eyberg Child Behaviour Inventory (Eyberg and Ross 1978; Eyberg 1980), the Strengths and Difficulties Questionnaire (Goodman 1997), the Incredible Years Social Competence Scale (Corrigan 2002) and items from the 5-year evaluation of Early Start (Fergusson et al 2005b). For consistency, all scale items were rated on a 3-point rating scale: not at all, somewhat and a great deal.

Confirmatory factor analysis of this item set (Horwood et al 2011) showed these items measured six correlated dimensions of child behaviour⁷. These dimensions were:

- **Conduct Disorder:** This dimension was based on a sum of 18 items describing the extent to which the child displayed aggressive and antisocial behaviours. The reliability of the scale assessed by Cronbach's alpha (α) was .87.
- **Defiance:** This dimension was based on a sum of 15 items describing the extent to which the child showed oppositional, defiant or dishonest behaviours (α =.89).
- **ADHD:** This dimension was based on a sum of 16 items describing the extent to which the child showed hyperactive, impulsive or inattentive behaviours (α =.92).
- **Self-control:** This dimension was based on a sum of 15 items describing the extent to which the child showed self-regulatory, flexible or compliant behaviours (α =.87).
- **Anxiety/Withdrawal:** This dimension was based on a sum of 16 items describing the extent to which the child showed anxious, withdrawn or shy behaviours (α =.78).

- **Social Competence:** This dimension was based on a sum of 26 items describing the extent to which the child showed helpful, empathetic, respectful, diligent or likeable behaviours (α =.91).

The findings

As shown in Table 4, parents reported significant improvement in their children's behaviour on all six measures in the Family Interview following the completion of the IYP programme. All the behaviour scores show a linear trend of improving change through the programme (p <.001). The effect sizes of these changes were medium to large. The greatest improvement was evident for Oppositional Defiant Disorder (d =.96) and Self-control (d =.96), followed by Social Competence (d =.68) and Conduct Disorder (d =.62). ADHD (d =.55) and Anxiety/Withdrawal (d =.51) showed moderate improvement.

Significant, large improvements were also reported in the ECBI Intensity (d =1.0) and Social Competency (d =.87) scales administered by the Group Leaders before and after the IYP course. The international RCT studies quoted in Chapter 2 (Gardner et al 2006; Hutchings et al 2007; Larsson et al 2009; Scott et al 2001; Webster-Stratton and Hammond 1997) all reported post-treatment effect sizes for child behaviour change in this range.

Moreover, the proportion of children scoring above the clinical cut-off of 60 on the ECBI Intensity Scale dropped from 75 percent pre-course to 19 percent post-course. Twenty-five percent of the children were below the clinical cut-off of 17 on the Social Competency Scale pre-course compared with 58 percent post-course. Larsson et al (2009) also reported significantly fewer children in the clinical range after the IYP course.

6 See Appendix 2 for descriptions of source instruments.

7 Five of the 111 items were discarded as they did not belong obviously to any of the subscales.

Table 4. Child behaviours: Mean parent reports before, during and after IYP programme

	Baseline mean (std)	Mid-course mean (std)	Post-course mean (std)	Trend <i>p</i> ⁶	Post-course effect size <i>d</i> (95% CI)
Child behaviours¹					
Conduct Disorder (CD)	28.8 (6.9)	25.6 (6.3)	24.5 (5.6)	<.001	.62 (.40 to .84)
ODD ²	35.8 (7.3)	31.4 (7.5)	28.8 (6.8)	<.001	.96 (.73 to 1.2)
ADHD ³	34.2 (8.5)	31.2 (8.3)	29.6 (8.5)	<.001	.55 (.33 to .77)
Self-control	35.9 (7.8)	40.4 (8.0)	43.3 (8.2)	<.001	.96 (.73 to 1.2)
Anxiety/Withdrawal	27.4 (5.8)	25.4 (5.6)	24.5 (5.6)	<.001	.51 (.29 to .73)
Social Competence	54.6 (10.4)	58.5 (10.7)	61.5 (10.2)	<.001	.67 (.45 to .90)
In-course behavioural assessments					
ECBI Intensity ⁴	65.1 (10.2)		54.8 (9.5)	<.001	1.0 (.72 to 1.3)
Social Competence Scale ⁵	17.1 (7.9)		23.9 (8.5)	<.001	.87 (.59 to 1.2)

¹Sample sizes are N=166, 165, 163 (baseline, mid-course, post-course).

²Oppositional Defiant Disorder.

³Attention Deficit Hyperactivity Disorder.

⁴Sample sizes are N=151, 108 (baseline, post-course).

⁵Sample sizes are N=150, 107 (baseline, post-course).

⁶Trends with *p*-values <.006 are considered significant allowing for the multiple statistical tests of significance (Bonferroni adjustment to $\alpha=.05$ for eight tests).

Child behaviour outcomes at follow-up

Parents were interviewed again about half a year after course completion, which is approximately 1 year after the baseline interview before the start of the IYP programme. The results from the follow-up interviews were compared to the baseline interviews to measure the medium-term sustainability of any behaviour change evident at post-course.

All the behaviour scores in Table 5 showed a linear trend of improving change ($p<.001$) with the estimated effect sizes ranging from medium to large ($d=.56$ to $d=1.0$, median $d=.71$). Parents reported their children's improved behaviours following the IYP programme were sustained 6 months after programme completion. The ECBI Intensity scores and the Social Competence Scale were obtained by Group Leaders before the IYP course and at completion only. These measures were not collected at follow-up.

Table 5. Child behaviours: Mean parent reports before IYP programme and at follow-up

	Baseline mean (std)	Follow-up mean (std)	Trend <i>p</i> ⁴	Follow-up effect size <i>d</i> (95% CI)
Child behaviours¹				
Conduct Disorder (CD)	28.8 (6.9)	24.1 (6.0)	<.001	.68 (.46 to .91)
ODD ²	35.8 (7.3)	28.6 (7.6)	<.001	.98 (.75 to 1.2)
ADHD ³	34.2 (8.5)	29.0 (8.3)	<.001	.61 (.39 to .84)
Self-control	35.9 (7.8)	43.9 (8.7)	<.001	1.0 (.81 to 1.3)
Anxiety/Withdrawal	27.4 (5.8)	24.2 (5.6)	<.001	.56 (.34 to .78)
Social Competence	54.6 (10.4)	62.1 (10.5)	<.001	.73 (.50 to .95)

¹Sample sizes are N=166, 162 (baseline, follow-up).

²Oppositional Defiant Disorder.

³Attention Deficit Hyperactivity Disorder.

⁴Trends with *p*-values <.008 are considered significant allowing for the multiple statistical tests of significance (Bonferroni adjustment to $\alpha=.05$ for six tests).

5.2 Parenting practices and relationships

Pre-test/post-test comparisons: Parenting practice

The instruments⁸

The parenting practices measures in the Family Interview consisted of two recognised instruments: the Alabama Parenting Questionnaire (Shelton et al 1996) and the Arnold-O'Leary Parenting Scale (Arnold et al 1993). Intact instruments were used, but the scale items were all rated on a consistent 3-point scale: never, sometimes and often for the Alabama; never, less than once a month and once a week or more for the Arnold-O'Leary.

The Alabama Parenting Questionnaire is a 42-item scale designed to tap the parenting dimensions that are risk factors associated with child conduct

disorder. It loads onto five subscales: parental supervision, positive parenting, corporal punishment, parental involvement and inconsistent discipline ($\alpha=.44$ to $.79$, median $\alpha=.53$). An evaluation with a community sample of Australian children aged 4–9 years (Dadds et al 2003) showed good internal consistency, validity and test-retest reliability for the measure.

The Arnold-O'Leary Parenting Scale is a 30-item inventory of parenting competencies that measures dysfunctional and/or ineffective parenting practices of parents with younger children. The scale yields an overall score and three revised subscale scores of dysfunctional strategies used by parents tackling problem behaviour ($\alpha=.42$ to $.74$, median $\alpha=.65$). 'Laxness' refers to insufficient monitoring of the child

8 See Appendix 2 for descriptions of source instruments.

and her/his behaviour, allowing rules to go unenforced or providing positive reinforcement for misbehaviour. 'Over-reactivity' refers to displays of anger, meanness or irritability. 'Hostility' refers to the use of verbal or physical force. The scale has adequate internal consistency and has been found to have good test-retest reliability.

The findings

According to parent reports of their parenting practices, all eight measures obtained in the Family Interviews showed an improvement after the parents had completed the IYP course compared to their reported practices before the course (Table 6). All parenting factors showed significant linear trends of change throughout

the programme ($p < .001$). The sizes of the improvements were medium or medium/large ($d = .46$ to $d = .71$), with the exception of a small improvement for Poor supervision ($d = .26$). The overall median effect size was $d = .54$ (excluding the Total Scale). The Total Scale (Arnold-O'Leary Parenting Scale), which includes items additional to those loading on the specific factors, produced a large effect size of $d = .83$. The parents in the Hutchings et al (2007) and Larsson et al (2009) studies also reported the reduced use of harsh and inconsistent discipline as well as the increased use of positive parenting strategies, while Gardner et al (2006) reported a reduction in observed negative parenting and an improvement in observed positive parenting.

Table 6. Parenting practices: Mean parent reports before, during and after IYP programme

Parenting measures ¹	Baseline mean (std)	Mid-course mean (std)	Post-course mean (std)	Trend p^5	Post-course effect size d (95% CI)
Parenting practices²					
Poor supervision	6.6 (1.3)	6.3 (0.8)	6.3 (0.9)	<.001	.26 (.04 to .47)
Positive parenting	13.7 (1.5)	14.3 (1.0)	14.4 (0.9)	<.001	.46 (.24 to .68)
Corporal punishment	3.7 (0.8)	3.4 (0.7)	3.2 (0.6)	<.001	.59 (.37 to .81)
Parental involvement	11.8 (2.0)	12.4 (1.8)	12.9 (1.7)	<.001	.53 (.31 to .75)
Inconsistent discipline	7.4 (2.2)	6.6 (1.9)	6.2 (2.0)	<.001	.54 (.32 to .76)
Dealing with misbehaviour³					
Lax discipline	8.2 (2.1)	7.5 (2.0)	7.1 (1.9)	<.001	.53 (.31 to .75)
Over-reactive discipline	8.7 (2.2)	7.7 (1.6)	7.1 (1.7)	<.001	.71 (.49 to .94)
Hostile discipline	4.4 (1.1)	3.8 (1.0)	3.7 (0.9)	<.001	.68 (.46 to .91)
Total scale ⁴	1.6 (0.3)	1.5 (0.3)	1.4 (0.3)	<.001	.83 (.60 to 1.1)

¹Sample sizes are N=166, 166, 164 (baseline, mid-course, post-course).

²From the Alabama Parenting Questionnaire.

³From the Arnold-O'Leary Parenting Scale.

⁴The Total Scale includes all the questions from the Arnold-O'Leary Parenting Scale including additional questions that are not part of the Lax, Over-reactive, or Hostile discipline factors, and the Total Scale scores are standardised to an average score between 1 and 3.

⁵Trends with p -values $< .005$ are considered significant allowing for the multiple statistical tests of significance (Bonferroni adjustment to $\alpha = .05$ for nine tests).

Parenting practice outcomes at follow-up

The follow-up effect sizes indicated the improved parenting practices evident at programme completion were sustained 6 months later. As shown in Table 7, the significant linear trends of change in parenting practices remained at the follow-up interviews ($p < .001$). The effect sizes

were small to medium ($d = .33$ to $d = .68$) with Poor supervision remaining small ($d = .25$). The overall median effect size was $d = .54$, excluding the Total Scale. The Total Arnold-O’Leary Parenting Scale maintained a large effect size of $d = .79$ at the fourth interview.

Table 7. Parenting practices: Mean parent reports before IYP programme and at follow-up

Parenting measures ¹	Baseline mean (std)	Follow-up mean (std)	Trend p^5	Follow-up effect size d (95% CI)
Parenting practices²				
Poor supervision	6.6 (1.3)	6.3 (0.8)	<.001	.25 (.04 to .47)
Positive parenting	13.7 (1.5)	14.2 (1.1)	<.001	.33 (.11 to .54)
Corporal punishment	3.7 (0.8)	3.3 (0.6)	<.001	.48 (.26 to .70)
Parental involvement	11.8 (2.0)	12.8 (1.8)	<.001	.47 (.25 to .69)
Inconsistent discipline	7.4 (2.2)	6.2 (1.9)	<.001	.55 (.33 to .77)
Dealing with misbehaviour³				
Lax discipline	8.2 (2.1)	7.1 (2.0)	<.001	.52 (.29 to .74)
Over-reactive discipline	8.7 (2.2)	7.2 (1.6)	<.001	.68 (.46 to .90)
Hostile discipline	4.4 (1.1)	3.7 (1.0)	<.001	.60 (.38 to .82)
Total scale ⁴	1.6 (0.3)	1.4 (0.3)	<.001	.79 (.56 to 1.0)

¹Sample sizes are N=166 baseline, 162 follow-up.

²From the Alabama Parenting Questionnaire.

³From the Arnold-O’Leary Parenting Scale.

⁴The Total Scale includes all the questions from the Arnold-O’Leary Parenting Scale including additional questions that are not part of the Lax, Over-reactive, or Hostile discipline factors, and the Total Scale scores are standardised to an average score between 1 and 3.

⁵Trends with p -values $< .005$ are considered significant allowing for the multiple statistical tests of significance (Bonferroni adjustment to $\alpha = .05$ for nine tests).

Pre-test/post-test comparisons: Conflict and relationships

The instruments⁹

The revised Straus Parent/Child and Partner Conflict Tactics Scales (Straus et al 1996; Straus et al 1998), the Partner Attachment Scale (Braiker and Kelley 1979) and the Dadds Parent Problem Checklist (Dadds and Powell 1991) measured parents' relationships with their children and partners. The intact instruments were used with scale items all rated on a consistent 3-point scale: never, less than once a week and once a week or more (Parent/Child Conflict Tactics); never, sometimes and often (Partner Conflict Tactics); not at all, somewhat and a great deal (Parent Problem Checklist); doesn't apply, somewhat applies and definitely applies (Partner Attachment Scale).

The Conflict Tactics Scale Parent/Child (CTSPC) and the Revised Conflict Tactics Scale 2 (CTS2) assess violence within families and intimate relationships. The 24 CTSPC items related to the parents' use of verbal aggression and physical assault in their relationships with their children ($\alpha=.28$ to $.73$, median $\alpha=.50$). The CTS2 focused on violence between the parents ($\alpha=.67$ to $.87$, median $\alpha=.82$) across 42 items.

The Parent Problem Checklist (PPC) was developed as a measure of inter-parental conflict, especially as it relates to the parents' ability to co-operate and to act as a team in performing the executive parenting functions within the family. It contains 16 items measuring the presence or absence of parental disagreement over rules and discipline for child misbehaviour, the occurrence of open conflict over child-rearing issues and whether or not parents undermine each other's relationships with the children ($\alpha=.86$ to $.93$, median $\alpha=.89$). The PPC is a unidimensional measure with moderately high internal consistency and high test-retest reliability.

The Partner Attachment Scale measures the quality of the relationship between parents. The items used in the Incredible Years Pilot Study were based on a selected series of items from Braiker and Kelley (1979) as used in the Christchurch Health and Development Study (CHDS) 21-Year Interview, and included 12 items producing reliability coefficients of $\alpha=.80$ to $.93$, median $\alpha=.88$.

The findings

Table 8 summarises the findings on a range of inter-partner and parent/child relationships. Verbal aggression and Physical assault, for both the primary caregivers and their partners, showed a linear trend of change ($p<.001$) with medium effect sizes ($d=.46$ to $d=.60$, median $d=.50$). Reported Child-rearing disagreement between parents reduced ($p<.001$) with a medium estimated effect size of $d=.49$ but there was no discernible difference in the quality of the parents' relationship before and after the IYP course. There was a small reduction reported in violence from the partner towards the primary caregiver ($d=.27$) but no change in violence from the primary caregiver towards the partner.

9 See Appendix 2 for descriptions of source instruments.

Table 8. Parental conflict and relationships: Mean parent reports before, during and after IYP programme

Relationship types	Baseline mean (std)	Mid-course mean (std)	Post-course mean (std)	Trend p^5	Post-course effect size d (95% CI)
Primary caregiver and child¹					
Verbal aggression	8.4 (2.3)	7.3 (2.1)	7.0 (2.0)	<.001	.60 (.38 to .83)
Physical assault	11.9 (1.2)	11.5 (1.1)	11.3 (0.8)	<.001	.50 (.28 to .72)
Other parent² and child					
Verbal aggression	7.5 (2.3)	6.6 (1.8)	6.4 (1.5)	<.001	.50 (.20 to .81)
Physical assault	11.8 (1.3)	11.5 (1.1)	11.2 (0.7)	<.001	.46 (.15 to .77)
Inter-parental violence³					
Violence to partner	22.5 (2.3)	22.1 (2.4)	21.9 (1.4)	.182*	.23 (-.07 to .54)
Violence from partner	22.6 (2.5)	22.0 (1.7)	21.9 (1.5)	.003	.27 (-.03 to .58)
Inter-parental relationship⁴					
Child-rearing disagreement	25.0 (6.8)	22.3 (5.8)	21.6 (5.1)	<.001	.49 (.18 to .80)
Relationship quality	31.5 (4.5)	32.0 (4.1)	32.5 (3.4)	.108*	.21 (-.09 to .51)

¹Scores are based on the Parent/Child Conflict Tactics Scale (M. Straus); for the primary caregiver sample sizes were N=166, 165, 163 (baseline, mid-course, post-course).

²For the other parent and inter-parental factors (where applicable) samples sizes were N=91, 87, 84 (baseline, mid-course, post-course).

³Inter-parental violence is from the Revised Conflict Tactics Scale (CTS2, M. Straus) as found in the CHDS 21-Year Interviews.

⁴Child-rearing disagreement is from the Parent Problem Checklist (M. Dadds), and Relationship quality is from the CHDS 21-Year Interviews.

⁵Trends with p-values <.006 are considered significant allowing for the multiple statistical tests of significance (Bonferroni adjustment to $\alpha=.05$ for eight tests).

* No evidence of a linear trend of change.

Conflict and relationships outcomes at follow-up

As shown in Table 9, the reported changes in conflict and relationships at follow-up ($d=.40$ to $d=.59$, median $d=.47$) were similar to the

outcomes at post-programme. In addition, caregivers reported a reduction in their partners' violence towards them (median $d=.32$, $p<.001$) not evident at programme completion. Relationship quality remained unchanged.

Table 9. Parental conflict and relationships: Mean parent reports before IYP programme and at follow-up

Relationship types	Baseline mean (std)	Follow-up mean (std)	Trend p^5	Follow-up effect size d (95% CI)
Primary caregiver and child¹				
Verbal aggression	8.4 (2.3)	7.1 (2.0)	<.001	.59 (.37 to .82)
Physical assault	11.9 (1.2)	11.4 (0.8)	<.001	.46 (.24 to .68)
Other parent² and child				
Verbal aggression	7.5 (2.3)	6.4 (1.7)	<.001	.48 (.18 to .78)
Physical assault	11.8 (1.3)	11.3 (0.8)	<.001	.40 (.11 to .70)
Inter-parental violence³				
Violence to partner	22.5 (2.3)	21.8 (1.9)	<.001	.29 (-.01 to .58)
Violence from partner	22.6 (2.5)	21.7 (1.6)	<.001	.35 (.05 to .64)
Inter-parental relationship⁴				
Child-rearing disagreement	25.0 (6.8)	21.5 (6.3)	<.001	.51 (.21 to .80)
Relationship quality	31.5 (4.5)	32.2 (4.0)	.230*	.15 (-.14 to .44)

¹Scores are based on the Parent/Child Conflict Tactics Scale (M. Straus); for the primary caregiver sample sizes were $N=166$, 162 (baseline, follow-up).

²For the other parent and inter-parental factors (where applicable) samples sizes were $N=91$, 90 (baseline, follow-up).

³Inter-parental violence is from the Revised Conflict Tactics Scale (CTS2, M. Straus) as found in the CHDS 21-Year Interviews.

⁴Child-rearing disagreement is from the Parent Problem Checklist (M. Dadds), and Relationship quality is from the CHDS 21-Year Interviews.

⁵Trends with p -values $<.006$ are considered significant allowing for the multiple statistical tests of significance (Bonferroni adjustment to $\alpha=.05$ for eight tests).

* No evidence of a trend of improvement or change.

5.3 Teacher¹⁰ pre/post measures of child behaviour

The behaviour of children displaying conduct problems may be setting-specific, such as at home or at school. Webster-Stratton (2012) noted that only about half of the children with behaviour problems at home also display these behaviours in the school environment. To assess whether or not any improvements in child behaviour had generalised from home to school following the IYP training, behaviour in the classroom or early child centre was assessed using the Canterbury Social Development Scale¹¹ (Church et al 2006). This is a 30-item rating scale which consists of brief descriptions of 15 antisocial behaviours and 15 positive social behaviours which are likely to occur in the preschool/school environment ($\alpha=.96$). The original scale was adapted by reducing the 5-point rating scale to a 3-point rating scale.

Packets containing information about the Incredible Years Pilot Study, a copy of the Social Development Scale and a self-addressed pre-paid envelope for returns were couriered to the preschool/school principals of the 160 study children whose parents had given consent for the researchers to contact their children’s teachers or childcare staff. Principals were asked to pass the packets on to the relevant teachers. After two telephone follow-ups, the return rate was 86 percent.

Teachers of the study children reported on 30 specific positive and negative behaviours in the school, preschool, or care centre setting before the parents attended the IYP programme. This was repeated at post-programme and follow-up for the children for whom there was teacher baseline data. Lower scores indicated better behaviour.

Outcomes post-course and follow-up in the school setting

Table 10 shows teachers reported small but significant improvements in child behaviours, on average, in the school setting at post-course (71.2) compared to the baseline score (75.7), albeit with a very modest effect size of $d=.17$ ($p=.008$). The follow-up score of 67.9, with the small effect size of $d=.29$ ($p=.005$), demonstrated the behaviour change was sustained and slightly improved (Table 11). The effect sizes on teacher reports of behaviour change were much smaller than those based on parent reports.

Table 10. Child behaviours: Mean teacher reports before and after IYP programme

	Baseline mean (std)	Post-course mean (std)	Trend <i>p</i>	Post-course effect size <i>d</i> (95% CI)
Teacher reports¹				
Total scores	75.7 (27.1)	71.2 (25.5)	.008	.17 (-.07 to .40)

¹Sample sizes are N=150 baseline, 139 post-course.

10 The term ‘teacher’ refers to preschool/school teacher or childcare worker; ‘school’ refers to preschool/school or childcare facility.

11 See Appendix 2 for descriptions of source instruments.

Table 11. Child behaviours: Mean teacher reports before IYP programme and at follow-up

	Baseline mean (std)	Follow-up mean (std)	Trend <i>p</i>	Follow-up effect size <i>d</i> (95% CI)
Teacher reports¹				
Total scores	75.7 (27.1)	67.9 (25.4)	.005	.29 (.04 to .54)

¹Sample sizes are N=150 baseline, 124 follow-up.

Referrals from schools

It is possible that referrals to IYP training from a school-based source (early childhood education teacher, Resource Teacher Behaviour and Learning, classroom teacher, or Special Education) indicated poorer child behaviour in the school setting. When teacher reports were grouped into

those who were school-referred and other, the children in the former group appeared to engage in higher rates of misbehaviour, on average, than children in the latter. This difference was significant ($p=.001$ and $p<.001$) and occurred at each stage of the study (Table 12).

Table 12. Child behaviours: Mean teacher reports before and after IYP programme and at follow-up, by referral source

	Non-school referred mean (std)	School referred mean (std)	Difference between referral source mean (std)	Test ¹ for no difference between referral source ² <i>p</i>
Teacher reports				
Baseline scores ³	67.3 (25.6)	87.2 (25.8)	19.9 (25.7)	<.001
Post-course scores ⁴	64.4 (23.7)	80.0 (26.6)	15.6 (24.8)	.001
Follow-up scores ⁵	61.4 (22.8)	78.6 (26.6)	17.2 (24.2)	<.001

¹Satterthwaite test.

²Sample sizes were slightly smaller for this analysis because referral source was not available for all participants; unclassified were N=13 baseline, 12 post-course, 9 follow-up.

³Baseline sample sizes are N=84, 53 (non-school referred, school referred).

⁴Post-course sample sizes are N=79, 48 (non-school referred, school referred).

⁵Follow-up sample sizes are N=74, 41 (non-school referred, school referred).

Although teachers reported higher levels of misbehaviour for children referred by a school source compared with those referred by another source, improved behaviour was reported for both groups of children following their parents' attendance at the IYP programme. As seen in Table 13, the effect sizes at post-programme for non-school referred children ($d=.11$) and for

school-referred children ($d=.28$) were small and consistent ($p=.242$). The follow-up stage also showed small effects for non-school referred children ($d=.23$) and school referred children ($d=.33$), which were consistent ($p=.746$). Regardless of whether the children were school-referred or not, the improvement in behaviour was broadly similar.

Table 13. Child behaviour: Effect sizes of mean teacher reports of change from baseline to post-course and follow-up, by referral source

	Non-school referred ¹ <i>d (95% CI)</i>		School referred ² <i>d (95% CI)</i>		Behaviour change test between source of referral <i>p</i>
Child behaviours					
Baseline to post-course	.11	(-.20 to .43)	.28	(-.13 to .68)	.242
Baseline to follow-up	.23	(-.09 to .56)	.33	(-.10 to .77)	.746

¹Sample sizes for children who were referred to the IYP programme by non-school sources are N=84 baseline, 79 post-course, 74 follow-up.

²Sample sizes for children who were referred to the IYP programme by a school are N=53, 48, 41 (baseline, post-course, follow-up). Sample sizes for children not classifiable are N=13 baseline, 12 post-course, 9 follow-up.

The lower effect size on the teacher measure compared with the parent measure may suggest the major benefit of the IYP programme was in the home, not in the school. However, given the informant and the measuring instrument differed as well as the setting, it is not possible to conclude that the different effect sizes can be accounted for by a change in setting. Nevertheless, it is likely that, although parent management training like

the IYP courses may be beneficial in reducing behaviour problems in school, they need to be used in conjunction with school-based programmes for maximum return in the school environment. Webster-Stratton et al (2004) reported that adding Teacher Training to Parent Training or Child Training improved treatment outcomes in terms of teacher behaviour management in the classroom and in reports of behaviour problems.

Chapter 6:

Other outcomes



The previous chapter presented the findings relating to the child behaviour, parenting practice and relationship outcomes for families who attended the IYP programme. This chapter examines other relevant outcomes.

In addition to the outcomes already reported, parent management training might be expected to have broader benefits for the family in terms of improved health or a decrease in stressful life events. Also, it is possible that the presenting level of behaviour problems exhibited by the child might affect the degree of behaviour change observed. Because the study was conducted over two phases at three sites, it was necessary to examine outcomes by phase and site to assess the possibility of bias in the study sample.

6.1 Outcomes for health and stressful life events

Parental depression, stress and anxiety and poor child health are known correlates of conduct problems (Odgers et al 2007), while a family's exposure to stressful and adverse life events may exacerbate existing problems.

An inventory of child health and child-related services derived from the Early Start Field Trial 5-Year Follow-up questionnaire (Fergusson et al 2005b) was used to assess child health ($\alpha=.48$ to $.52$, median $\alpha=.51$). The Composite International Diagnostic Interview version of the Diagnostic and Statistical Manual of Mental Disorders (World Health Organisation 1993) assessed depression symptoms ($\alpha=.87$ to $.89$, median $\alpha=.89$) and the anxiety and stress scales of the Depression

Anxiety Stress Scale (Lovibond and Lovibond 1995) measured anxiety and stress ($\alpha=.63$ to $.91$, median $\alpha=.86$). The occurrence of stressful life events was measured using items from the Early Start Field Trial 5-Year Follow-up questionnaire (Fergusson et al 2005b), based on Holmes and Rahe (1967).

Table 14 presents parents' reported changes in life and health factors following their attendance at the IYP programme. Stress decreased ($p<.001$) with a small effect size of $d=.30$. Anxiety showed no change, but the initial scores were very low. Depression symptoms were not measured at

post-course because the 12-month recall period for the measure was outside the post-course timeframe. Parental reports of decreased stress were found in other studies (eg Larsson et al 2009).

There was a suggestion of a trend towards a decrease in the occurrence of stressful life events ($p<.001$) post-course, albeit with a small effect size of $d=.25$. There was also evidence the children in the Incredible Years Pilot Study had fewer visits to health and other child-related agencies at programme completion ($p=.007$). Again, the effect size of this improvement was small ($d=.21$).

Table 14. Health and stressful life events: Mean parent reports before, during and after IYP programme

	Baseline mean (std)	Mid-course mean (std)	Post-course mean (std)	Trend p^5	Post-course effect size d (95% CI)
Measures of child health					
Health and services ¹	1.7 (1.7)	1.5 (1.6)	1.4 (1.5)	.007	.21 (-.01 to .42)
Measures of parental health²					
Stress	10.8 (3.6)	10.3 (3.8)	9.7 (3.5)	<.001	.30 (.08 to .52)
Anxiety	7.4 (1.9)	7.2 (2.2)	7.0 (1.8)	.065*	.19 (-.03 to .40)
Depression symptoms ³	4.7 (3.0)	NA	NA		
Stressful life events⁴					
Score of life events	6.9 (7.1)	5.4 (5.9)	5.2 (5.7)	<.001	.25 (.03 to .47)

¹Average number of visits to health agencies and/or child-related services in a 3-month period; sample sizes are N=166, 165, 163 (baseline, mid-course, post-course).

²Scores based on the Depression Anxiety Stress Scale (DASS); sample sizes are N=166, 164, 164 (baseline, mid-course, post-course).

³Depression symptoms were not calculated for mid-course or post-course because the recall period was 12 months (see Table 15).

⁴The life events score is the sum of events/intensity of event; sample sizes are N=166, 166, 164 (baseline, mid-course, post-course).

⁵Trends with p -values $<.010$ are considered significant allowing for the multiple statistical tests of significance (Bonferroni adjustment to $\alpha=.05$ for five tests).

* No evidence of a linear trend of change.

At the follow-up interview parents reported the use of fewer health services for their children than at baseline ($p=.006$). The number of stressful life events had also decreased between the first interview and the follow-up ($p<.001$).

Although stress symptoms were the only health measure that showed an improvement at the post-course measure, 6 months later parents

reported the number of stress, anxiety and depression symptoms had all shown improvements compared to the baseline, with small to medium effect sizes ($d=.29$ to $d=.54$, median $d=.39$). Depression symptoms had decreased the most compared to the baseline, with an effect size of $d=.54$ ($p<.001$). See Table 15.

Table 15. Health and stressful life events: Mean parent reports before IYP programme and at follow-up

	Baseline mean (std)	Follow-up mean (std)	Trend p^5	Follow-up effect size d (95% CI)
Measures of child health				
Health and services ¹	1.7 (1.7)	1.3 (1.5)	.006	.21 (-.01 to .43)
Measures of parental health				
Stress ²	10.8 (3.6)	9.4 (3.2)	<.001	.39 (.17 to .61)
Anxiety	7.4 (1.9)	6.8 (1.9)	.001	.29 (.07 to .51)
Depression symptoms ³	4.7 (3.0)	3.1 (3.0)	<.001	.54 (.32 to .76)
Stressful life events⁴				
Score of life events	6.9 (7.1)	4.4 (4.5)	<.001	.36 (.14 to .58)

¹Average number of visits to health agencies and/or child-related services in a 3-month period; sample sizes are N=166, 161 (baseline, follow-up).

²Stress and anxiety scores are based on the Depression Anxiety Stress Scale (DASS); sample sizes are N=166, 162 (baseline, follow-up).

³Depression symptoms are based on the Composite International Depression Index (CID) and are a count of symptoms only; samples sizes are N=166, 162 (baseline, follow-up).

⁴The life events score is the sum of events/intensity of event; sample sizes are N=166, 162 (baseline, follow-up).

⁵Trends with p -values $<.010$ are considered significant allowing for the multiple statistical tests of significance (Bonferroni adjustment to $\alpha=.05$ for five tests).

6.2 Outcomes for children in clinical and sub-clinical groups

Group Leaders asked parents to complete the ECBI Intensity and Problem Scales and the Social Competency Scale as standard practice before the start of the IYP programme, and again on completion. These scales measure the level of Conduct Problem and Social Competency in the children.

The ECBI Scaled Intensity measure

About one-quarter (28%) of the children in the Incredible Years Pilot Study had a pre-course score of less than 60 on the ECBI Scaled Intensity measure (N=42 of 151 matched records), suggesting a sizeable proportion of the children had sub-clinical levels of Conduct Problem at baseline. The size of the sub-clinical sample made it possible to test whether or not IYP programme efficacy differed for children in the clinical and sub-clinical groups.

As shown in Table 16, for Conduct Disorder, ODD, and ADHD (all $p < .001$) and Self-control ($p = .009$), there was a significant interaction between the change in behaviour and the clinical/sub-clinical sub-group. The interaction was marginal for Anxiety/Withdrawal ($p = .046$). This suggests the children with clinical level ECBI Intensity scores at baseline benefitted more from the IYP programme on most of the behaviour measures. The interaction for Social Competence was not significant.

The multivariate analysis of IYP programme outcomes on overall Child Behaviour confirmed that the effect sizes for children in the sub-clinical and clinical Conduct Problem groups were different ($p < .001$), where the median effect sizes were $d = .51$ for sub-clinical children (ranging from $d = .28$ to $d = .90$) compared to $d = .83$ for the clinical children (ranging from $d = .59$ to $d = 1.4$).

Table 16. Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by sub-clinical and clinical ECBI Intensity

	Sub-clinical Conduct Problem <i>d</i> (95% CI)	Clinical ¹ Conduct Problem <i>d</i> (95% CI)	Test for interactions ⁵ <i>p</i>
Child behaviours^{2,6}			
Conduct Disorder (CD)	.43 (-.01 to .86)	.84 (.56 to 1.1)	<.001*
ODD ³	.90 (.45 to 1.3)	1.4 (1.1 to 1.7)	<.001*
ADHD ⁴	.28 (-.15 to .71)	.80 (.52 to 1.1)	<.001*
Self-control	.86 (.41 to 1.3)	1.4 (1.1 to 1.6)	.009*
Anxiety/Withdrawal	.38 (-.06 to .81)	.59 (.32 to .87)	.046*
Social Competence	.59 (.15 to 1.0)	.82 (.54 to 1.1)	.120

¹Clinical Conduct Problem is defined by the pre-course ECBI Intensity Score being 60 or greater.

²N (baseline, post-course) for sub-clinical Conduct Problem=42, 42; for clinical Conduct Problem=109, 107.

³Oppositional Defiant Disorder.

⁴Attention Deficit Hyperactivity Disorder.

⁵The results of tests for interactions between the sub-groups and the observations at baseline and course completion (outcomes) for each measure are represented by the *p*-values, where $p < .05$ is evidence the change over time for a measure was not similar between the sub-groups of sub-clinical and clinical Conduct Problem.

⁶A test for an interaction between the sub-groups and the observations at baseline and course completion for all six Child Behaviour measures showed a significant interaction between the overall outcome and the clinical/sub-clinical Conduct Problem sub-groups: $F(6, 142) = 4.10, p < .001$.

*The change in scores over time between clinical and sub-clinical Conduct Problem children was not similar.

The evidence presented in Table 17 confirmed the differences for some of the measures detected post-programme between the clinical and sub-clinical groups of children at the 6-month follow-up. The sub-clinical group typically had smaller effect sizes for the child behaviour measures than did the clinical group. There was a significant interaction between the change in behaviour and the clinical/sub-clinical sub-groups on the Conduct Disorder ($p < .001$) and ODD ($p = .003$) measures, as was found at the end of the IYP course. However, the ADHD, Self-control and Anxiety/Withdrawal measures were no longer significantly different, perhaps reflecting the continued improvement reported by

parents of the children in the sub-clinical group. Improvement in Social Competence remained similar for both groups of children at the follow-up interviews.

Effect sizes for the sub-clinical children were small to large, ranging from $d = .40$ to $d = 1.3$ (median $d = .64$), while the clinical children had medium to large effect sizes, ranging from $d = .61$ to $d = 1.4$ (median $d = .94$). These differences in the median effect sizes were confirmed by the multivariate analysis of the overall Child Behaviour outcome interaction between the two observation points and the clinical/sub-clinical sub-groups ($p < .001$).

Table 17. Child behaviours at follow-up: Effect sizes of mean parent reports of change, by sub-clinical and clinical ECBI Intensity

	Sub-clinical Conduct Problem <i>d</i> (95% CI)		Clinical ¹ Conduct Problem <i>d</i> (95% CI)		Test for interactions ⁵ <i>p</i>
Child behaviours^{2,6}					
Conduct Disorder (CD)	.40	(-.04 to .83)	.96	(.68 to 1.2)	<.001*
ODD ³	1.0	(.54 to 1.5)	1.4	(1.1 to 1.7)	.003*
ADHD ⁴	.52	(.08 to .96)	.83	(.55 to 1.1)	.053
Self-control	1.3	(.78 to 1.7)	1.4	(1.1 to 1.7)	.459
Anxiety/Withdrawal	.58	(.14 to 1.0)	.61	(.34 to .89)	.432
Social Competence	.69	(.24 to 1.1)	.92	(.64 to 1.2)	.199

¹Clinical Conduct Problem is defined by the pre-course ECBI Intensity Score being 60 or greater.

²N (baseline, follow-up) for sub-clinical Conduct Problem=42, 41; for clinical Conduct Problem=109, 107.

³Oppositional Defiant Disorder.

⁴Attention Deficit Hyperactivity Disorder.

⁵The results of tests for interactions between the sub-groups and the observations at baseline and follow-up for each measure are represented by the *p*-values, where $p < .05$ is evidence the change over time for a measure was not similar between the sub-groups of sub-clinical and clinical Conduct Problem.

⁶A test for an interaction between the sub-groups and the observations at baseline and follow-up for all six Child Behaviour measures showed a significant interaction between the overall outcome and the clinical/sub-clinical Conduct Problem sub-groups: $F(6, 141) = 4.50, p < .001$.

*The change in the scores over time between clinical and sub-clinical Conduct Problem children was not similar.

The Social Competence Scale measure

The Social Competence Scale (SCS) measures pro-social behaviours, communication skills and self-control. Table 18 examines the behaviour outcomes for children scoring above and below the clinical cut-off score of 17 on the Social Competence Scale at baseline.

The reported changes in the Conduct Disorder and ODD measures were significantly different ($p=.003$ and $p=.007$ respectively) for the clinical and sub-clinical SCS children, with larger effect sizes evident for children in the clinical group.

This implies a greater improvement on these two measures for the clinical SCS children. However, the multivariate analysis on a combined measure of Child Behaviour indicates the effect sizes were broadly similar for the two groups of children ($p=.100$) at the end of the IYP course. The median effect sizes were $d=.62$ for sub-clinical children ($d=.41$ to $d=.96$) compared to $d=.87$ for clinical children ($d=.63$ to $d=1.4$).

Table 18. Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by sub-clinical and clinical Social Competence Scale (SCS)

	Sub-clinical ¹ SCS <i>d</i> (95% CI)		Clinical SCS <i>d</i> (95% CI)		Test for interactions ⁵ <i>p</i>
Child behaviours^{2,6}					
Conduct Disorder (CD)	.53	(.18 to .87)	.86	(.54 to 1.2)	.003*
ODD ³	.86	(.50 to 1.2)	1.3	(.96 to 1.6)	.007*
ADHD ⁴	.49	(.14 to .83)	.78	(.46 to 1.1)	.058
Self-control	.96	(.60 to 1.3)	1.4	(1.1 to 1.8)	.079
Anxiety/Withdrawal	.41	(.07 to .75)	.63	(.31 to .95)	.100
Social Competence	.70	(.36 to 1.0)	.87	(.54 to 1.2)	.182

¹Sub-clinical Social Competence is defined by the pre-course Social Competence Scale scoring 17 or more.

²For sub-clinical Social Competence Scale scores $N=68, 68$ (baseline, post-course); for clinical Social Competence Scale scores $N=82, 80$ (baseline, post-course).

³Oppositional Defiant Disorder.

⁴Attention Deficit Hyperactivity Disorder.

⁵The results of tests for interactions between the sub-groups and the observations at baseline and course completion (outcomes) for each measure are represented by the p -values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of sub-clinical and clinical SCS.

⁶A test for an interaction between the sub-groups and the observations at baseline and course completion for all six Child Behaviour measures showed no significant interaction between the overall outcome and the clinical/sub-clinical SCS sub-groups: $F(6, 141)=1.82, p=.100$.

*The change in scores over time between clinical and sub-clinical SCS children was not similar.

Six months after the completion of the IYP course parent reports of Conduct Disorder and ODD remained significantly different for children with clinical and sub-clinical baseline SCS scores (Table 19). For Conduct Disorder the effect size was $d=.51$ for those in the sub-clinical group compared to the effect size of $d=1.0$ for those in the clinical group ($p<.001$). For ODD the sub-clinical SCS effect size was $d=.84$ compared to the clinical SCS effect size of $d=1.4$ ($p=.008$). In addition, effect sizes for the clinical and sub-clinical SCS children were significantly different at follow-up for both Self-control ($p=.039$) and Social Competence ($p=.009$), a difference not evident immediately after finishing the IYP programme.

The multivariate analysis for the combined Child Behaviour measure confirmed these findings. The effect sizes for the children in the sub-clinical SCS group were medium to large ($d=.51$ to $d=1.0$, median $d=.61$), whereas for the clinical SCS children the effect sizes were large ($d=.63$ to $d=1.6$, median $d=1.1$) ($p=.004$). The implication is that, in the medium term, improvements in child behaviour may be greater for children with a clinical level SCS pre-course score than for those with a sub-clinical level pre-course score.

Table 19. Child behaviours at follow-up: Effect sizes of mean parent reports of change, by sub-clinical and clinical Social Competence Scale (SCS)

	Sub-clinical ¹ SCS <i>d</i> (95% CI)		Clinical SCS <i>d</i> (95% CI)		Test for interactions ⁵ <i>p</i>
Child behaviours^{2,6}					
Conduct Disorder (CD)	.51	(.16 to .85)	1.0	(.72 to 1.4)	<.001*
ODD ³	.84	(.48 to 1.2)	1.4	(1.0 to 1.7)	.008*
ADHD ⁴	.55	(.21 to .90)	.88	(.56 to 1.2)	.057
Self-control	1.0	(.66 to 1.4)	1.6	(1.2 to 2.0)	.039*
Anxiety/Withdrawal	.55	(.21 to .90)	.63	(.31 to .95)	.677
Social Competence	.66	(.31 to 1.0)	1.1	(.75 to 1.4)	.009*

¹Sub-clinical Social Competence is defined by the pre-course Social Competence Scale scoring 17 or more.

²For sub-clinical Social Competence Scale scores $N=68, 67$ (baseline, follow-up); for clinical Social Competence Scale scores $N=82, 80$ (baseline, follow-up).

³Oppositional Defiant Disorder.

⁴Attention Deficit Hyperactivity Disorder.

⁵The results of tests for interactions between the sub-groups and the observations at baseline and follow-up for each measure are represented by the p -values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of sub-clinical and clinical SCS.

⁶A test for an interaction between the sub-groups and the observations at baseline and follow-up for all six Child Behaviour measures showed a significant interaction between the overall outcome and the clinical/sub-clinical SCS sub-groups: $F(6, 140)=3.32, p=.004$.

*The change in the scores over time between clinical and sub-clinical Conduct Problem children was not similar.

6.3 Outcomes for phase 1 and phase 2

The Incredible Years Pilot Study was carried out over two time periods in 2011 to ensure an adequate sample size for the research: Phase 1 was February to July (N=88); Phase 2 was August to December (N=78). Table 20 examines the interactions between the phase samples and the change in outcomes across the Child Behaviour measures to test for possible differences in outcomes by phase at the end of the IYP programme. The test for differences in outcomes at programme completion yielded marginal p-values for ODD (p=.041) and Social Competence (p=.048). Phase 1 had a median effect size of d=.77 compared to Phase 2 with a median effect size of d=.55.

These results may suggest the programmes delivered in Phase 2 were slightly less effective than those delivered in Phase 1. The multivariate analysis of the combined Child Behaviour measure indicated the Phase 1 programmes achieved marginally higher effect sizes than the Phase 2 ones as assessed at the post-course interview (p=.033). The in-course ECBI Intensity score also showed a difference between Phase 1 and Phase 2 outcomes, whereby the Phase 1 effect size was d=1.2 compared to the Phase 2 effect size of d=.76 (p=.041). The reasons for these differences are not known.

Table 20. Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by study phase

	Phase 1 <i>d</i> (95% CI)		Phase 2 <i>d</i> (95% CI)		Test for interactions ⁶ <i>p</i>
Child behaviours^{1,7}					
Conduct Disorder (CD)	.71	(.41 to 1.0)	.51	(.19 to .83)	.130
ODD ²	1.1	(.78 to 1.4)	.81	(.48 to 1.1)	.041*
ADHD ³	.57	(.26 to .87)	.53	(.21 to .85)	.664
Self-control	1.0	(.72 to 1.4)	.88	(.54 to 1.2)	.631
Anxiety/Withdrawal	.45	(.15 to .76)	.56	(.24 to .89)	.410
Social Competence	.83	(.52 to 1.1)	.51	(.19 to .83)	.048*
In-course behavioural assessments					
ECBI Intensity ⁴	1.2	(.83 to 1.6)	.76	(.35 to 1.2)	.041*
Social Competence Scale ⁵	.81	(.43 to 1.2)	.94	(.53 to 1.4)	.853

¹Sample sizes for Phase 1 are N=88, 86; for Phase 2 N=78, 77 (baseline, post-course).

²Oppositional Defiant Disorder.

³Attention Deficit Hyperactivity Disorder.

⁴Sample sizes for Phase 1 on the ECBI Intensity scores are N=80, 50; for Phase 2 N=70, 51 (baseline, post-course).

⁵Sample sizes for Phase 1 on the Social Competence Scale are N=80, 57; for Phase 2 N=70, 50 (baseline, post-course).

⁶The results of tests for interactions between the sub-groups and the observations at baseline and course completion (outcomes) for each measure are represented by the p-values, where p<.05 is evidence the change over time for a measure was not similar between the sub-groups of the study phase.

⁷A test for an interaction between the sub-groups and the observations at baseline and course completion for all six Child Behaviour measures showed a significant interaction between the overall outcome and the study phase sub-groups: F(6, 156)=2.36, p=.033. This statistic does not include the in-course ECBI and SCS assessments.

*The change in score over time between Phase 1 and Phase 2 samples was not similar.

However, the observed differences between the study phases at post-course were not sustained to the follow-up interviews (Table 21). Results from the multivariate analysis showed Phase 1 (median effect size $d=.79$) and Phase 2 (median effect size

$d=.61$) were not significantly different ($p=.492$). The lack of evidence of a sustained difference in outcomes between Phase 1 and Phase 2 children at follow-up justifies combining the results for these two groups.

Table 21. Child behaviours at follow-up: Effect sizes of mean parent reports of change, by study phase

	Phase 1 <i>d (95% CI)</i>		Phase 2 <i>d (95% CI)</i>		Test for interactions ⁴ <i>p</i>
Child behaviours^{1,5}					
Conduct Disorder (CD)	.76	(.45 to 1.1)	.59	(.26 to .91)	.238
ODD ²	1.1	(.79 to 1.4)	.84	(.50 to 1.2)	.118
ADHD ³	.66	(.35 to .96)	.57	(.24 to .89)	.948
Self-control	1.1	(.78 to 1.4)	.97	(.54 to 1.2)	.883
Anxiety/Withdrawal	.56	(.25 to .86)	.56	(.24 to .89)	.902
Social Competence	.82	(.51 to 1.1)	.63	(.30 to .95)	.477

¹Sample sizes for Phase 1 are N=88, 87; for Phase 2 are N=78, 75 (baseline, follow-up).

²Oppositional Defiant Disorder.

³Attention Deficit Hyperactivity Disorder.

⁴The results of tests for interactions between the sub-groups and the observations at baseline and follow-up for each measure are represented by the p-values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of the study phase.

⁵A test for an interaction between the sub-groups and the observations at baseline and follow-up for all six Child Behaviour measures showed no significant interaction between the overall outcome and the study phase sub-groups: $F(6, 155)=0.91, p=.492$.

6.4 Outcomes for the three sites

The study was conducted at three sites: Bay of Plenty, Canterbury and Mid-Central. Table 22 presents evidence of an interaction between the pilot site and the change in Child Behaviour at programme completion for the measures of ODD ($p=.047$), ADHD ($p=.005$), Anxiety/Withdrawal ($p=.021$), and the ECBI Intensity scores ($p<.001$). Mid-Central had consistently larger effect sizes across the behaviour measures ($d=.71$ to $d=1.4$,

median $d=.87$) compared to both the Bay of Plenty ($d=.46$ to $d=1.0$, median $d=.63$) and Canterbury ($d=.35$ to $d=.85$, median $d=.60$). This finding was confirmed by the multivariate analysis on the combined Child Behaviour measure ($p=.002$). Mid-Central also had the largest difference in effect sizes on the ECBI Intensity ($d=1.6$) compared with the Bay of Plenty ($d=.69$) and Canterbury ($d=.95$).

Table 22. Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by study site

	Bay of Plenty <i>d</i> (95% CI)	Canterbury <i>d</i> (95% CI)	Mid-Central <i>d</i> (95% CI)	Test for interactions ⁶ <i>p</i>
Child behaviours^{1,7}				
Conduct Disorder (CD)	.59 (.21 to .97)	.56 (.17 to .94)	.83 (.43 to 1.2)	.234
ODD ²	.84 (.45 to 1.2)	.85 (.46 to 1.2)	1.4 (.93 to 1.8)	.047*
ADHD ³	.48 (.10 to .86)	.36 (-.02 to .74)	.90 (.50 to 1.3)	.005*
Self-control	1.0 (.63 to 1.4)	.74 (.35 to 1.1)	1.2 (.76 to 1.6)	.168
Anxiety/Withdrawal	.46 (.08 to .83)	.35 (-.03 to .73)	.71 (.31 to 1.1)	.021*
Social Competence	.66 (.28 to 1.0)	.63 (.24 to 1.0)	.73 (.33 to 1.1)	.615
In-course behavioural assessments				
ECBI Intensity ⁴	.69 (.24 to 1.1)	.95 (.46 to 1.4)	1.6 (1.0 to 2.2)	<.001*
Social Competence Scale ⁵	.56 (.12 to 1.0)	.82 (.33 to 1.3)	1.3 (.77 to 1.9)	.090

¹Sample sizes for Bay of Plenty are N=56, 56; for Canterbury N=57, 55; for Mid-Central N=53, 52 (baseline, post-course).

²Oppositional Defiant Disorder.

³Attention Deficit Hyperactivity Disorder.

⁴Sample sizes on ECBI Intensity for Bay of Plenty are N=52, 41; for Canterbury=53, 36; for Mid-Central=46, 31 (baseline, post- course).

⁵Sample sizes for Social Competence Scale for Bay of Plenty are N=52, 41; for Canterbury=52, 35; for Mid-Central=46, 31 (baseline, post-course).

⁶The results of tests for interactions between the sub-groups and the observations at baseline and course completion (outcomes) for each measure are represented by the p-values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of the study site.

⁷A test for an interaction between the sub-groups and the observations at baseline and course completion for all six Child Behaviour measures showed a significant interaction between the overall outcome and the study site sub-groups: $F(12, 310)=2.64, p=.002$. This statistic does not include the in-course ECBI and SCS assessments.

*The change in scores over time between the study sites was not similar.

Reported Child Behaviour outcomes at course completion were sustained at the follow-up interviews at the three sites (Table 23), with medium to large effect sizes in the Bay of Plenty and Canterbury and consistently higher effect sizes again observed in Mid-Central. The multivariate analysis revealed a significant difference among the sites ($p < .001$). Effect sizes

for all three sites were good, but Mid-Central ($d = .91$) was relatively better than Canterbury ($d = .71$) and the Bay of Plenty ($d = .66$).

The reason for this site difference is not known, but it does not appear to be an issue of programme fidelity (see Chapter 9, Section 9.1).

Table 23. Child behaviours at follow-up: Mean parent reports, by study site

	Bay of Plenty <i>d</i> (95% CI)	Canterbury <i>d</i> (95% CI)	Mid-Central <i>d</i> (95% CI)	Test for interactions ⁴ <i>p</i>
Child behaviours ^{1,5}				
Conduct Disorder (CD)	.63 (.25 to 1.0)	.67 (.28 to 1.0)	.87 (.46 to 1.3)	.211
ODD ²	.78 (.40 to 1.2)	.88 (.49 to 1.3)	1.5 (1.1 to 1.9)	.014*
ADHD ³	.55 (.18 to .93)	.43 (.05 to .81)	.95 (.54 to 1.4)	.032*
Self-control	1.1 (.74 to 1.5)	.81 (.42 to 1.2)	1.2 (.80 to 1.6)	.170
Anxiety/Withdrawal	.51 (.13 to .88)	.40 (.02 to .77)	.79 (.38 to 1.2)	.012*
Social Competence	.78 (.40 to 1.2)	.65 (.26 to 1.0)	.75 (.35 to 1.2)	.447

¹Sample sizes for Bay of Plenty are N=56, 56; for Canterbury N=57, 55; for Mid-Central N=53, 51 (baseline, follow-up).

²Oppositional Defiant Disorder.

³Attention Deficit Hyperactivity Disorder.

⁴The results of tests for interactions between the sub-groups and the observations at baseline and follow-up for each measure are represented by the *p*-values, where $p < .05$ is evidence the change over time for a measure was not similar between the sub-groups of the study site.

⁵A test for an interaction between the sub-groups and the observations at baseline and follow-up for all six Child Behaviour measures showed a significant interaction between the overall outcome and the study site sub-groups: $F(12, 308) = 2.80, p = .001$.

*The change in scores over time between the study sites was not similar.

Chapter 7:

Programme outcomes for Māori and non-Māori



An important issue raised by the preceding results concerns the extent to which the study findings held for Māori and non-Māori families. This section addresses this question by comparing the outcomes for Māori children in the study with the outcomes for non-Māori children on the pre-test/post-test and follow-up comparisons of child behaviours, parenting practices and relationships reported from Table 4 to Table 9 in Chapter 5. The independent kaupapa Māori study is also discussed in this chapter. Satisfaction ratings of the IYP programme for Māori and non-Māori parents are presented in Chapter 9.

7.1 The Incredible Years Pilot Study

Pre-test/post-test comparisons: Child behaviour

Table 24 contains the effect sizes for Māori and non-Māori children on the Child Behaviours measured in the Incredible Years Pilot Study as well as an assessment of the extent to which outcomes varied by ethnicity. If reported outcomes differed for Māori and non-Māori, tests of the interaction of change by ethnicity should be significant.

It is evident from the data that the effect sizes of the before-and-after scores for Māori and non-Māori children were similar. The child behaviour effect sizes for Māori children ranged from $d=.43$ to $d=.82$ (median $d=.56$) compared to the effect sizes for non-Māori, which ranged from $d=.46$ to $d=1.1$ (median $d=.77$). In all but one case the change by ethnicity interaction was non-significant, implying an absence of difference in the outcomes for Māori and non-Māori children. The exception was a before-and-after difference in the level of improvement in parent-reported Social Competence ($p=.012$). This interaction

reflected the fact that, although both Māori and non-Māori children showed improvement in Social Competence, the improvement for non-Māori was larger ($d=.79$) than the improvement for Māori ($d=.47$).

Although estimated effect sizes across the measures in Table 24 appear typically better for non-Māori children, the multivariate analysis demonstrated the overall programme outcome on Child Behaviours for Māori (median effect size $d=.56$) and non-Māori (median effect size $d=.73$) was similar ($p=.072$).

Table 24. Child behaviours after IYP programme: Effect sizes of mean parent reports of change, by Māori and non-Māori children

	Māori <i>d</i> (95% CI)		Non-Māori <i>d</i> (95% CI)		Test for interactions ⁶ <i>p</i>
Child behaviours^{1,7}					
Conduct Disorder (CD)	.53	(.16 to .90)	.67	(.39 to .95)	.186
ODD ²	.82	(.44 to 1.2)	1.1	(.76 to 1.3)	.087
ADHD ³	.43	(.06 to .79)	.62	(.34 to .90)	.131
Self-control	.81	(.43 to 1.2)	1.0	(.76 to 1.3)	.119
Anxiety/Withdrawal	.58	(.21 to .95)	.46	(.19 to .74)	.304
Social Competence	.47	(.10 to .83)	.79	(.51 to 1.1)	.012*
In-course behavioural assessments					
ECBI Intensity ⁴	.93	(.41 to 1.4)	1.1	(.73 to 1.4)	.980
Social Competence Scale ⁵	.96	(.43 to 1.5)	.88	(.55 to 1.2)	.786

¹Sample sizes for Māori are N=60, 59; for non-Māori are N=106, 104 (baseline, post-course).

²Oppositional Defiant Disorder.

³Attention Deficit Hyperactivity Disorder.

⁴Sample sizes for Māori are N=53, 32; for non-Māori are N=98, 76 (baseline, post-course).

⁵Sample sizes for Māori are N=52, 31; for non-Māori are N=98, 76 (baseline, post-course).

⁶The results of tests for interactions between the sub-groups and the observations at baseline and course completion (outcomes) for each measure are represented by the *p*-values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of Māori and non-Māori children.

⁷A test for an interaction between the sub-groups and the observations at baseline and course completion for all six Child Behaviour measures showed no significant interaction between the overall outcome and the Māori/non-Māori sub-groups: $F(6, 156)=1.98$, $p=.072$. This statistic does not include the in-course ECBI and SCS assessments.

*The change in scores over time between Māori and non-Māori was not similar.

Data from the follow-up interviews given in Table 25 show that, as with the post-course findings above, only the Social Competence measure showed a significantly ($p=.008$) smaller effect size ($d=.50$) for Māori than for non-Māori ($d=.86$) children. However, the effect sizes across all the Child Behaviour measures for non-Māori

(median $d=.80$) remained consistently better than for Māori (median $d=.63$). This was confirmed by the multivariate analysis for the combined Child Behaviour measure ($p=.025$). This significant difference was not evident at post-programme, but became apparent by the 6-month follow-up.

Table 25. Child behaviours at follow-up: Effect sizes of mean parent reports of change, by Māori and non-Māori children

	Māori <i>d</i> (95% CI)		Non-Māori <i>d</i> (95% CI)		Test for interactions ⁵ <i>p</i>
Child behaviours ^{1,4}					
Conduct Disorder (CD)	.58	(.20 to .95)	.74	(.46 to 1.0)	.102
ODD ²	.81	(.43 to 1.2)	1.1	(.79 to 1.4)	.068
ADHD ³	.57	(.20 to .94)	.64	(.36 to .91)	.671
Self-control	.89	(.51 to 1.3)	1.1	(.83 to 1.4)	.168
Anxiety/Withdrawal	.68	(.31 to 1.1)	.49	(.22 to .77)	.231
Social Competence	.50	(.13 to .87)	.86	(.58 to 1.1)	.008*

¹Sample sizes for Māori are N=60, 58; for non-Māori are N=106, 104 (baseline, follow-up).

²Oppositional Defiant Disorder.

³Attention Deficit Hyperactivity Disorder.

⁴The results of tests for interactions between the sub-groups and the observations at baseline and follow-up for each measure are represented by the *p*-values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of Māori and non-Māori children.

⁵A test for an interaction between the sub-groups and the observations at baseline and follow-up for all six Child Behaviour measures showed a significant interaction between the overall outcome and the Māori/non-Māori sub-groups: $F(6, 155)=2.49, p=.025$.

*The change in scores over time between Māori and non-Māori were not similar.

These findings show that, with the possible exception of Social Competence, the parents of both Māori and non-Māori children reported improvements in their children's behaviour that was sustained at the 6-month follow-up. It is possible that, given the number of comparisons made, the result for Social Competence reflects a chance variation rather than a genuine Māori/non-Māori difference in outcomes.

However, the effect sizes for Māori children were consistently smaller than those for non-Māori children, suggesting the IYP programme was less effective for Māori. Multivariate analyses on a combined Child Behaviour measure revealed that, although there was no significant difference between Māori and non-Māori children post-programme, a significant difference was evident at follow-up.

Pre-test/post-test comparisons:

Parenting practice

The effect sizes for parenting practices before and after the IYP programme for Māori and non-Māori shown in Table 26 were similar. The effect sizes for Māori parents ranged from $d=.28$ to $d=.72$ compared to the effect sizes for non-Māori parents, which ranged from $d=.24$ to $d=.90$. Hostile discipline on the Arnold-O'Leary Parenting Scale was the only significant change by ethnicity interaction ($p=.006$). The improvement appeared to be primarily on the item measuring bad language only, not on the other two items (slapping and insulting) underlying this factor. Both groups reported low levels of Hostile discipline before the beginning of the IYP course and

both Māori ($d=.44$) and non-Māori ($d=.83$) parents reported an improvement. However the change for non-Māori parents was larger.

The Arnold-O'Leary Total Scale of Dealing with Misbehaviour showed similar outcome scores for the parents of Māori and non-Māori children, with large effect sizes of $d=.72$ for Māori children and $d=.90$ for non-Māori children. Moreover, the multivariate analysis on a combined Parenting measure of outcomes between Māori (median effect size $d=.47$) and non-Māori (median effect size $d=.60$) indicated no significant difference ($p=.385$).

Table 26. Parenting practices after IYP programme: Effect sizes of mean parent reports of change, by Māori and non-Māori children

Parenting measures ^{1,3}	Māori <i>d</i> (95% CI)	Non-Māori <i>d</i> (95% CI)	Test for interactions ⁴ <i>p</i>
Parenting practices			
Poor supervision	.28 (-.08 to .64)	.24 (-.03 to .51)	.895
Positive parenting	.45 (.09 to .82)	.47 (.19 to .74)	.735
Corporal punishment	.56 (.19 to .93)	.60 (.33 to .88)	.932
Parental involvement	.54 (.17 to .90)	.52 (.25 to .80)	.954
Inconsistent discipline	.41 (.05 to .78)	.63 (.35 to .90)	.378
Dealing with misbehaviour			
Lax discipline	.47 (.10 to .83)	.57 (.30 to .85)	.749
Over-reactive discipline	.63 (.26 to 1.0)	.77 (.49 to 1.1)	.400
Hostile discipline	.44 (.07 to .80)	.83 (.55 to 1.1)	.006*
Total Scale ²	.72 (.34 to 1.1)	.90 (.61 to 1.2)	.406

¹Sample sizes for Māori are $N=60, 59$; for non-Māori $N=106, 105$ (baseline, post-course).

²The Total Scale includes all the questions from the Arnold-O'Leary Parenting Scale including additional questions that are not part of the Lax, Over-reactive, or Hostile discipline factors.

³A test for an interaction between the sub-groups and the observations at baseline and course completion for all nine Parenting Practice measures showed no significant interaction between the overall outcome and the Māori/non-Māori sub-groups: $F(9, 154)=1.07, p=.385$.

⁴The results of tests for interactions between the sub-groups and the observations at baseline and course completion (outcomes) for each measure are represented by the p -values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of Māori and non-Māori children.

*The change in scores over time between Māori and non-Māori was not similar.

Effect sizes for changes in parenting practices reported by Māori and non-Māori parents were again broadly similar across the measures at follow-up (Table 27). While the Dealing with Misbehaviour Total Scale showed similar changes in practices for both groups of parents, an examination of the subscales suggests an interaction between ethnicity and the change in the outcome scores for both the Hostile discipline ($p<.017$) and the Over-reactive discipline ($p<.001$)

measures whereby the change for non-Māori parents was greater than that for Māori parents.

On a combined Parenting measure, multivariate analysis suggested no differences in outcomes between the parents of Māori and non-Māori children. The effect sizes for Māori children ranged from $d=.22$ to $d=.62$ with a median $d=.41$, and for non-Māori from $d=.22$ to $d=.89$ with a median $d=.56$.

Table 27. Parenting practices at follow-up: Effect sizes of mean parent reports of change, by Māori and non-Māori children

Parenting measures ^{1,3}	Māori <i>d</i> (95% CI)	Non-Māori <i>d</i> (95% CI)	Test for interactions ⁴ <i>p</i>
Parenting practices			
Poor supervision	.32 (-.05 to .68)	.22 (-.06 to .49)	.483
Positive parenting	.22 (-.15 to .58)	.38 (.11 to .66)	.175
Corporal punishment	.38 (.01 to .75)	.54 (.26 to .82)	.480
Parental involvement	.44 (.07 to .81)	.48 (.21 to .76)	.573
Inconsistent discipline	.41 (.04 to .78)	.65 (.37 to .93)	.162
Dealing with misbehaviour			
Lax discipline	.45 (.08 to .82)	.56 (.29 to .84)	.639
Over-reactive discipline	.49 (.12 to .85)	.81 (.52 to 1.1)	.017*
Hostile discipline	.32 (-.05 to .68)	.77 (.49 to 1.1)	<.001*
Total Scale ²	.62 (.25 to .99)	.89 (.60 to 1.2)	.099

¹Sample sizes for Māori are N=60, 58; for non-Māori N=106, 104 (baseline, follow-up).

²The Total Scale includes all the questions from the Arnold-O'Leary Parenting Scale including additional questions that are not part of the Lax, Over-reactive, or Hostile discipline factors.

³A test for an interaction between the sub-groups and the observations at baseline and follow-up for all six Child Behaviour measures showed no significant interaction between the overall outcome and the Māori/non-Māori sub-groups: $F(9, 152)=1.51, p=.150$.

⁴The results of tests for interactions between the sub-groups and the observations at baseline and follow-up for each measure are represented by the *p*-values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of Māori and non-Māori children.

*The change in scores over time between Māori and non-Māori was not similar.

It appears both Māori and non-Māori parents reported broadly similar improvements in their parenting practice at the end of the IYP programme and 6 months beyond. It is possible, as with Social Competence, the difference found for Hostile discipline results from a chance variation due to the number of comparisons made rather than a genuine Māori/non-Māori difference.

Pre-test/post-test comparisons: Conflict and relationships

Reported improvements in inter-partner and parent/child relationships (Verbal aggression and Physical assault) were similar for Māori and non-Māori parents at post-course (Table 28). The effect sizes for Māori parents ranged from $d=.41$ to $d=.49$ (median $d=.46$) compared to the effect sizes for non-Māori parents, which ranged from $d=.48$ to $d=.67$ (median $d=.53$). There were no significant interactions for change by ethnicity, which suggests both Māori and non-Māori parents reported similar improvements in their relationships. The multivariate analysis produced no significant differences.

Table 28. Conflict and relationships after IYP programme: Effect sizes of mean parent reports of change, by Māori and non-Māori children

Relationship types ⁵	Māori <i>d</i> (95% CI)	Non-Māori <i>d</i> (95% CI)	Test for interactions ⁶ <i>p</i>
Primary caregiver and child¹			
Verbal aggression	.49 (.13 to .86)	.67 (.39 to .95)	.338
Physical assault	.45 (.08 to .81)	.53 (.26 to .81)	.946
Other parent² and child			
Verbal aggression	.46 (-.10 to 1.0)	.52 (.15 to .89)	.992
Physical assault	.41 (-.14 to .96)	.48 (.11 to .85)	.550
Inter-parental violence³			
Violence to partner	.30 (-.25 to .84)	.21 (-.16 to .57)	.168
Violence from partner	.40 (-.15 to .95)	.22 (-.15 to .58)	.143
Inter-parental relationship⁴			
Child-rearing disagreement	.45 (-.11 to 1.0)	.53 (.16 to .90)	.877
Relationship quality	.30 (-.25 to .84)	.18 (-.19 to .54)	.742

¹Scores are based on the Parent/Child Conflict Tactics Scale; for Māori primary caregivers sample sizes were N=60, 59; for non-Māori N=106, 105 (baseline, post-course).

²For the other parent and inter-parental factors (where applicable) samples sizes were for Māori N=27, 26 (baseline, post-course); for non-Māori N=64, 58.

³Inter-parental violence is from the Revised Conflict Tactics Scale (CTS2).

⁴Child-rearing disagreement is from the Parent Problem Checklist, and Relationship quality is from the Partner Attachment Scale.

⁵A test for an interaction between the sub-groups and the observations at baseline and course completion for all eight Relationship and Conflict measures showed no significant interaction between the overall outcome and the Māori/non-Māori sub-groups: $F(8, 72)=.59, p=.783$.

⁶The results of tests for interactions between the sub-groups and the observations at baseline and course completion (outcomes) for each measure are represented by the p-values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of Māori and non-Māori children.

As was the case at post-course, Table 29 shows consistent outcomes at the 6-month follow-up across the measures of Conflict, Inter-parental violence, Child-rearing disagreement and

Relationship quality for the parents of both the Māori and non-Māori study children. The multivariate analysis produced no significant differences.

Table 29. Conflict and relationships at follow-up: Mean parent reports of change, by Māori and non-Māori children

Relationship types ⁵	Māori <i>d</i> (95% CI)	Non-Māori <i>d</i> (95% CI)	Test for interactions ⁶ <i>p</i>
Primary caregiver and child¹			
Verbal aggression	.45 (.08 to .82)	.68 (.40 to .96)	.131
Physical assault	.30 (-.07 to .67)	.57 (.29 to .85)	.214
Other parent² and child			
Verbal aggression	.35 (-.18 to .87)	.54 (.18 to .90)	.753
Physical assault	.46 (-.07 to .99)	.38 (.02 to .73)	.670
Inter-parental violence³			
Violence to partner	.19 (-.34 to .71)	.40 (.05 to .76)	.939
Violence from partner	.37 (-.16 to .90)	.34 (-.02 to .69)	.885
Inter-parental relationship⁴			
Child-rearing disagreement	.41 (-.12 to .93)	.58 (.22 to .93)	.969
Relationship quality	.19 (-.33 to .72)	.14 (-.21 to .49)	.926

¹Scores are based on the Parent/Child Conflict Tactics Scale; for Māori primary caregivers sample sizes were N=60, 58; for non-Māori N=106, 104 (baseline, follow-up).

²For the other parent and inter-parental factors (where applicable) samples sizes were for Māori N=27, 28 (baseline, follow-up); for non-Māori N=64, 62.

³Inter-parental violence is from the Revised Conflict Tactics Scale (CTS2).

⁴Child-rearing disagreement is from the Parent Problem Checklist, and Relationship quality is from the Partner Attachment Scale.

⁵A test for an interaction between the sub-groups and the observations at baseline and the follow-up for all eight Relationship and Conflict measures showed no significant interaction between the overall outcome and the Māori/non-Māori sub-groups: $F(8, 68)=.56, p=.805$.

⁶The results of tests for interactions between the sub-groups and the observations at baseline and follow-up for each measure are represented by the *p*-values, where $p<.05$ is evidence the change over time for a measure was not similar between the sub-groups of Māori and non-Māori children.

Summary of the multivariate analyses for Māori and non-Māori children

Although on the individual outcome measures of child behaviour, parenting practice and relationships the differences between Māori and non-Māori children were not significant, the effect sizes were typically better for non-Māori children compared to Māori children. As reported in this chapter, the trend indicating possible differences in outcomes for Māori and non-Māori was examined by a doubly multivariate repeated measures Analysis of Variance, which tested for interactions between the combined outcome measures, the repeated interviews, and the ethnicity sub-groups, thereby verifying the multiple hypothesis tests in each table. Baseline differences in reported behaviour for Māori and non-Māori children were taken into account in the model fitted.

As shown in the summary of the doubly multivariate tests on the previous tables presented in Table 30, at IYP programme completion the differences between Māori and non-Māori children were not significant. However at the follow-up, significant differences were evident on the measures of Child Behaviours ($p=.025$), suggesting the outcomes on these measures for non-Māori children were stronger than those for Māori children, albeit Māori children still showed improvement. Differences in outcomes for Māori and non-Māori were not found by Fergusson et al (2009), which may reflect the more intensive measurement used in the present study or the lack of follow-up in the earlier study.

These findings indicate the need for culturally responsive refinements to IYP programmes to ensure the maximum effectiveness for all parents and children and the maintenance of behaviour improvements.

Table 30. Summary of the multivariate analyses for Māori and non-Māori children at IYP programme completion and follow-up

	Effect size range for Māori <i>d</i> (median)	Effect size range for non-Māori <i>d</i> (median)	Test for interactions ⁴ <i>p</i>
Programme outcomes			
Child behaviours ¹	.43 to .82 (.56)	.46 to 1.1 (.73)	.072
Parenting practices ²	.28 to .72 (.47)	.24 to .90 (.60)	.385
Relationships ³	.30 to .49 (.43)	.18 to .67 (.50)	.783
Follow-up outcomes			
Child behaviours ¹	.50 to .89 (.63)	.49 to 1.1 (.80)	.025*
Parenting practices ²	.22 to .62 (.41)	.22 to .89 (.56)	.150
Parent/Child relationships ³	.19 to .46 (.36)	.14 to .68 (.47)	.805

¹Child behaviour outcomes are from measures of: Conduct Disorder, ODD, ADHD, Self-control, Anxiety/Withdrawal, and Social Competence.

²Parenting practices include measures of: Poor supervision, Positive parenting, Corporal punishment, Parental involvement, and Inconsistent discipline, Lax discipline, Over-reactive discipline, Hostile discipline, and the Total Scale of Dealing with Misbehaviour.

³Relationships include measures of: Verbal aggression, Physical assault, Violence to/from partner, Child-rearing disagreement, and Relationship quality.

⁴The results of tests for interactions between the sub-groups and the observations over time for the groups of measures are represented by the *p*-values, where $p<.05$ is evidence the change over time for a group of measures was not similar between the sub-groups of Māori and non-Māori children.

*The change in scores over time across grouped measures between Māori and non-Māori was not similar.

7.2 The kaupapa Māori study

The independent kaupapa Māori study (Berryman et al 2012) explored the experiences of four Māori mothers who participated in one of the evaluated IYP programmes in 2011 in Canterbury or the Bay of Plenty, and the experiences of four Māori Special Education staff members who were Group Leaders of evaluated IYP programmes. The full report is available from the Ministry of Education.

The interviews are presented in the report as two collaborative stories, one the combined experiences of the four Māori mothers and the other the combined experiences of the four Māori Group Leaders. From these stories the authors conclude that:

- “Māori whānau wanted and needed to participate but they wanted to do so on their own terms. A culturally responsive programme should enable this to happen, however, it was clear that for at least two of these mothers this is not what had happened.” (page 56)
- “Māori facilitators understood that the most important element for them was to establish and consolidate the kaupapa, that is, what the delivery of Incredible Years required of management and of colleagues trained to facilitate Incredible Years and with whom they worked, to be most effective for whānau Māori. When this was firmly established, connecting with Māori parents, no matter their circumstances, was more likely to lead to the strengthening of collaborative learning that in turn led to whānau uptake of Incredible Years and as a consequence the further enhancement of the mana of the parent and the wellbeing of their whānau.” (page 56)

In their discussion the authors draw out three themes:

1. *Understanding culturally responsive adult learning*: “A better understanding of culturally responsive andragogy may be needed if the Incredible Years programme is going to work for more whānau Māori.... Berryman, SooHoo and Nevin (in print) identified the importance of the space within which the responsive dialogic rituals must proceed in order for a relationship of trust and respect to be nurtured amongst both parties. ... (They) suggest that listening to the other is more likely to occur when spaces to develop respectful relationships are given priority before engaging in any joint project.... Whakawhanaungatanga provides a cultural space for this to occur; where things are not done to people but where people are accorded the autonomy to contribute on their own terms.” (page 56)
2. *Programme fidelity (doing what we know works)*: “... the Māori facilitators explained that they modified and varied their delivery style of the Incredible Years programme so as to better engage with whānau Māori. Māori facilitators learned to make the programme as effective as possible for Māori families by drawing on the resources (knowledge and experiences) of their participants to find solutions that were a better fit with Māori values and practices.... (They) showed that when facilitators are able to make the ‘connects’, and when time and resources are available to facilitate the programme responsively, then the Incredible Years programme can be more effective for whānau and welcomed by them.” (page 57)

3. *A differentiated response:* "... a differentiated response ... should be available and doable within the Incredible Years programme. Beginning from a point of getting to 'really' know whānau by listening to them is the starting point. ... Incredible Years is a comprehensively researched and studied programme internationally and there is also an increasing body of literature pertaining to its delivery in New Zealand and its potential effectiveness for whānau Māori.... In New Zealand, Berryman, Woller and Glynn (2009) previously concluded that Incredible Years was effective for Māori when whānau were able to develop relationships of trust with facilitators to the extent that their own experiences were validated and they were able to contribute and to learn. Māori facilitators understood that they needed to both listen to whānau (responsive) and respond accordingly, and to bring a Māori perspective (appropriate) to the programme to help facilitate success for whānau." (page 58)

Berryman et al conclude that, although the findings of this kaupapa Māori study reiterate those of other studies, work must be done to provide a more responsive, cohesive (and at times differentiated) and aligned approach to meet the needs of Māori families.

7.3 Incredible Years for whānau Māori

The Incredible Years Pilot Study provides evidence of benefits from IYP training for whānau Māori. However, the benefits appear to be consistently lower than those reported for non-Māori families, particularly for ODD and Social Competence. The authors of the independent kaupapa Māori study, while acknowledging the IYP programme's potential effectiveness for whānau Māori, emphasise the importance of being more responsive to their needs (Berryman et al 2012). There are grounds for making further investment in developing culturally responsive refinements to the IYP programme to maximise the programme's effectiveness for whānau Māori.

The Werry Centre has hosted a series of annual hui (eg Werry Centre 2008, 2010) for Māori Group Leaders to share innovations in IYP programme delivery and to develop resources to ensure cultural responsiveness when delivering IYP courses to whānau Māori. At these hui Māori Group Leaders have consistently advocated for the need to implement, as standard practice, the many contributions they have naturally included in their delivery as culturally responsive practices to gain the best results for whānau Māori.

Recently the Werry Centre (2012) published *Ngā Tau Mīharo o Aotearoa*, a set of IYP programme resources for Māori Group Leaders. These resources recognise the importance of Te Ao Māori in delivering IYP courses to whānau Māori and were developed in collaboration with experienced Māori Group Leaders and Kaumātua. They include a DVD featuring Māori Group Leaders and whānau Māori, a workbook for Māori accreditees and practical resources for use in group delivery. The concept of weaving together the strands from two worlds derives from the metaphor of the weaving of harakeke (flax). In respecting the different meanings and understandings within both Te Ao Māori and Te Ao Pākehā, the Werry Centre acknowledges the partnership between these two worlds and how the skill sets may be aligned to maximise outcomes for whānau.

Chapter 8:

Single Case Studies



Single Case Studies were included in the Incredible Years Pilot Study to provide an alternative methodology to test the hypotheses of behaviour change and to investigate variability in individual child behaviour. Weekly Parent Reports were conducted at all three sites in Phase 1.

8.1 Parent Reports

A sub-sample of participants was randomly selected at each of the three sites for the Parent Report component in Phase 1 of the study. Ten parents each from Mid-Central and Canterbury consented to participate and five parents from the Bay of Plenty, making a total of 25 parents. The Parent Report sample consisted of nine Māori and 16 non-Māori children with a roughly even split of children aged under 5 years and 5 years or older. Of the 25 children, 20 had a clinical ECBI score (60 or more on the scaled Intensity score). Only two of the 25 parents did not complete the IYP course. Twenty parents completed all 18 weekly reports, a further three completed 16 and two completed 15. Twenty-three parents completed the four weekly reports at the follow-up stage, while two parents missed one interview each.

The weekly Parent Report¹² in the Incredible Years Pilot Study was a 10-minute telephone interview consisting of a checklist of 23 questions assessing positive behaviours and 26 questions assessing negative behaviours to which parents answered 'yes' or 'no' regarding their child's behaviour during the previous day. The interviews were conducted for every week of the 18-week IYP course and for 4 weeks at the 6-month follow-up. It was based on the Parent Daily Report (PDR) developed at the Oregon Learning Centre (Chamberlain and Reid 1987). This measure was chosen because it is well suited for repeated daily assessments and because it is especially effective in picking up low-frequency behaviour.

The data obtained from the Parent Reports provided two additional sources of information about the IYP programme: an alternative measure to the main interviews to assess programme efficacy and the ability to examine the individual progress for each child throughout the programme and at follow-up.

Initial data investigation

The Parent Reports recorded negative and positive behaviours throughout the IYP course and at the follow-up for each of the 25 children, thereby providing a week-by-week behaviour assessment at a single case level.

Initially, the positive and negative behaviour scores were plotted together on graphs for each individual child. By observing the plots on each graph, it was found that:

- The plots for many children contained unexplained variation, as demonstrated in the example graphs for Child 101 (Figure 4) and Child 311 (Figure 5), where behaviour changed considerably from week to week.
- The plots of positive behaviour started with high scores. These data were therefore less informative because the scope for behaviour change was restricted.

Addressing the unexplained variance

Given the amount of unexplained variance in the scores, trend lines through the weekly scores were calculated to determine whether the children demonstrated trends of change in their behaviour. These trend lines were produced using a mixed effects linear regression model. This method of statistical analysis provided a line for each child that represented her/his change in behaviour from the start of the course to course completion.

12 Descriptions of source instruments are in Appendix 2.

Figure 4. Example of no behaviour change (Child 101)

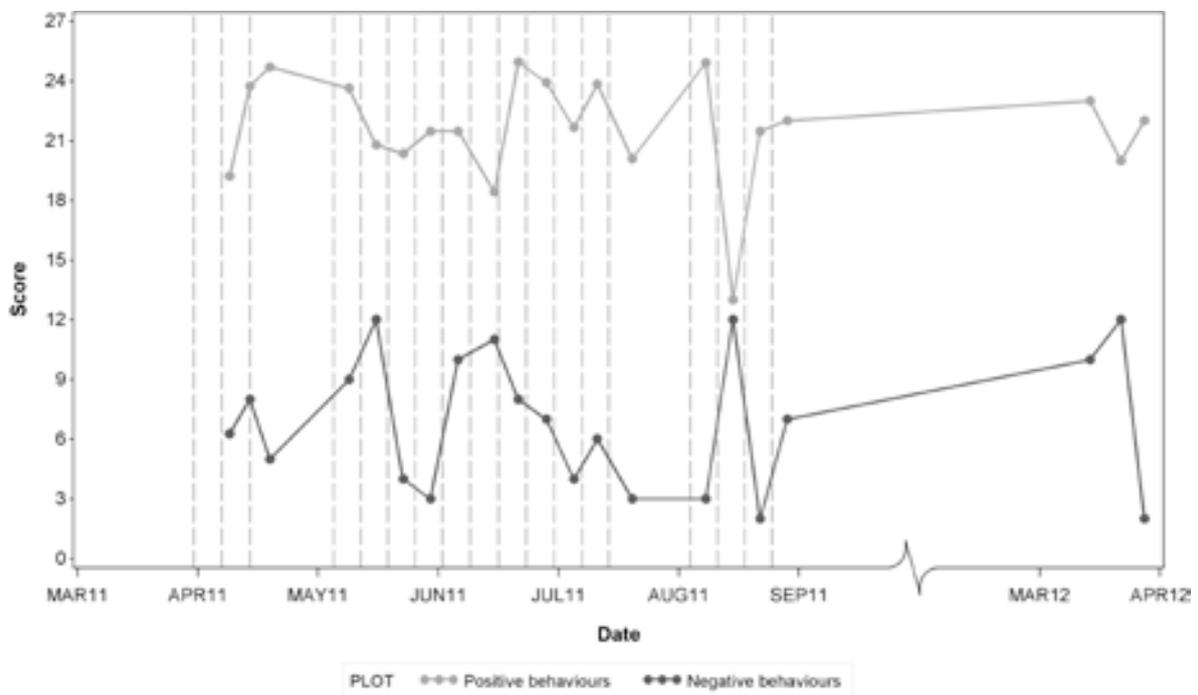
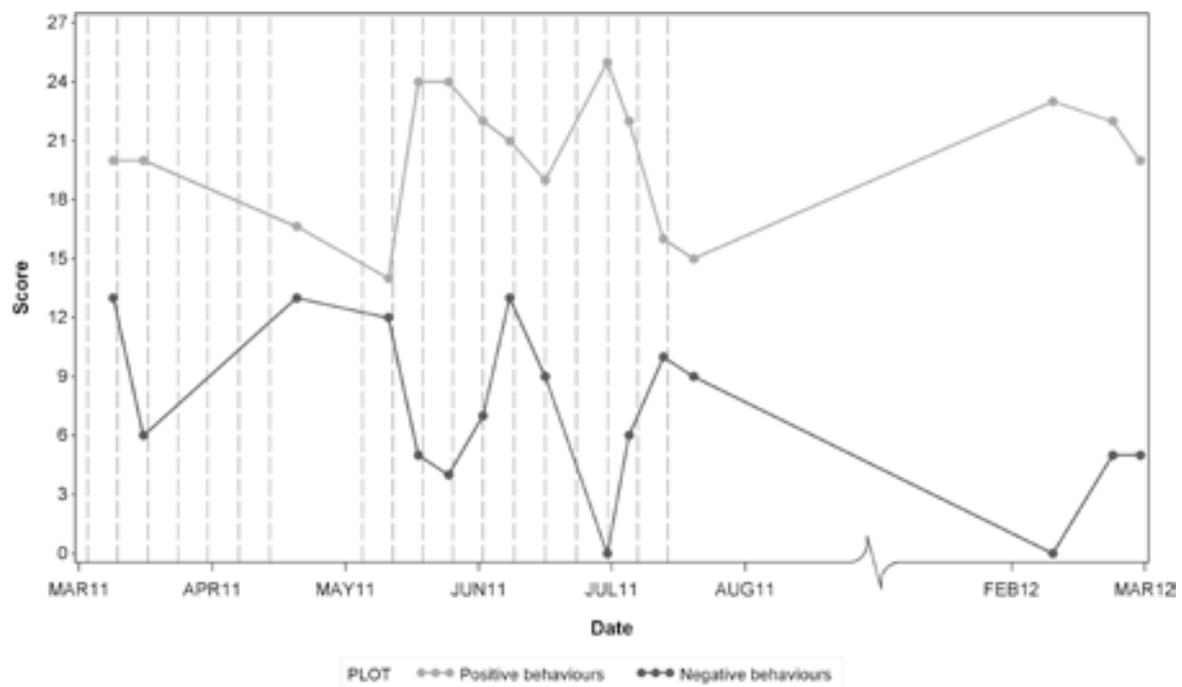


Figure 5. Example of behaviour change towards improvement (Child 311)



Note for both figures above: a) the vertical dashed lines represent the days on which IYP courses were run; b) the horizontal axis break represents the period before the follow-up where no data were collected.

Addressing the positive behaviour scores

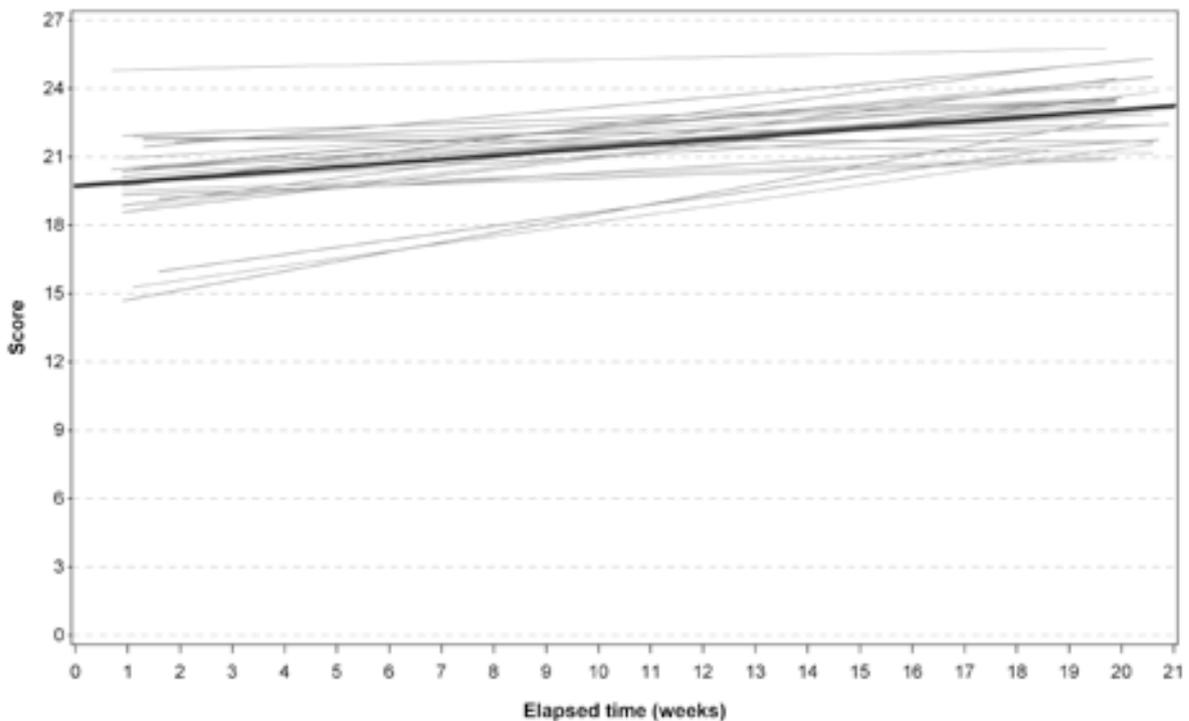
It was evident from Figure 6 that positive behaviour trended towards improvement but was limited by the maximum score assessed or ceiling. This ceiling effect on the scores was verified by the covariance parameters produced by the model, which showed a 0.81 correlation between the starting point of positive behaviour (the intercept) and the rate of improvement (the slope).

As shown in Figure 7 there were definite trends to the scores of negative behaviours. However, unlike the positive behaviour scores there was no effect of the scores being limited by the minimum

or maximum assessed score (-0.12 correlation with the intercept and slope). Because the positive behaviour scores were considered less informative than the negative behaviour scores, the two scores were combined by adding the inverse, or absence, of positive behaviours to the negative behaviours.

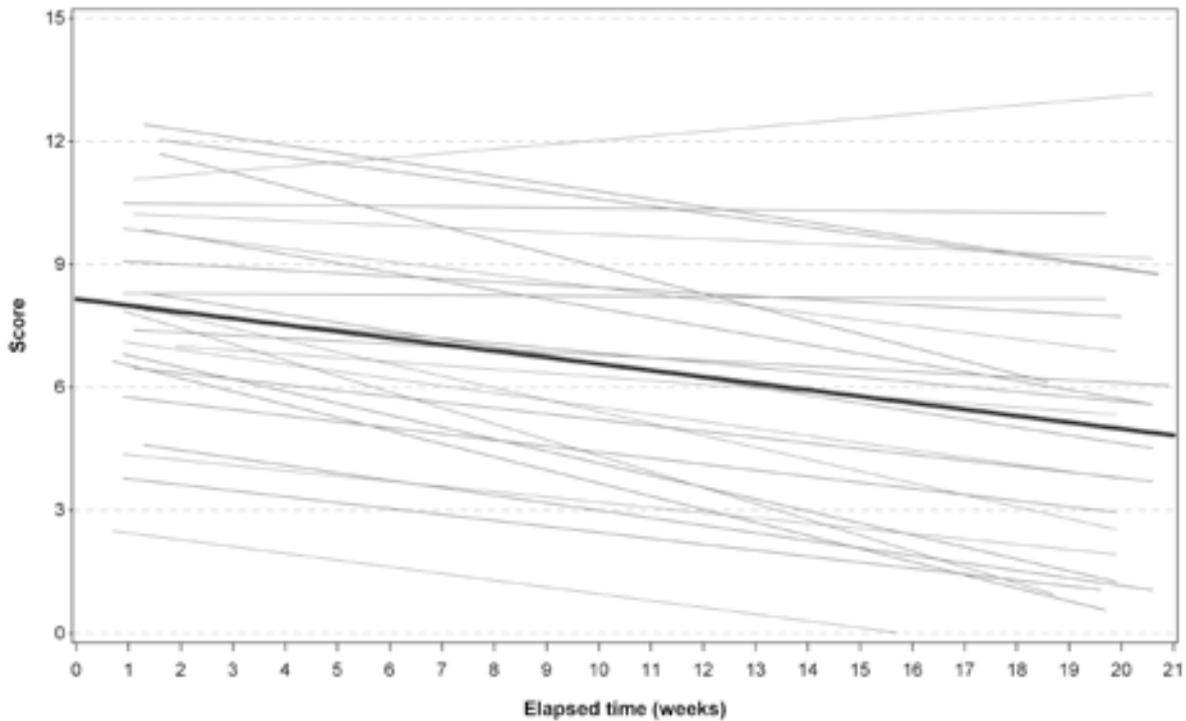
The results from the combined behaviour trends seen in Figure 8 demonstrate that trend lines can be used to describe the behaviour scores observed and that the combined behaviour score was an effective measure of the overall behaviour change.

Figure 6. Predicted positive behaviour scores for each child and for the IYP programme overall



Note: the thick line represents the linear regression of the IYP programme overall.

Figure 7. Predicted negative behaviour scores for each child and for the IYP programme overall



Note: the thick line represents the linear regression of the IYP programme overall.

Findings from the observations during the programme

The overall trend of behaviour change

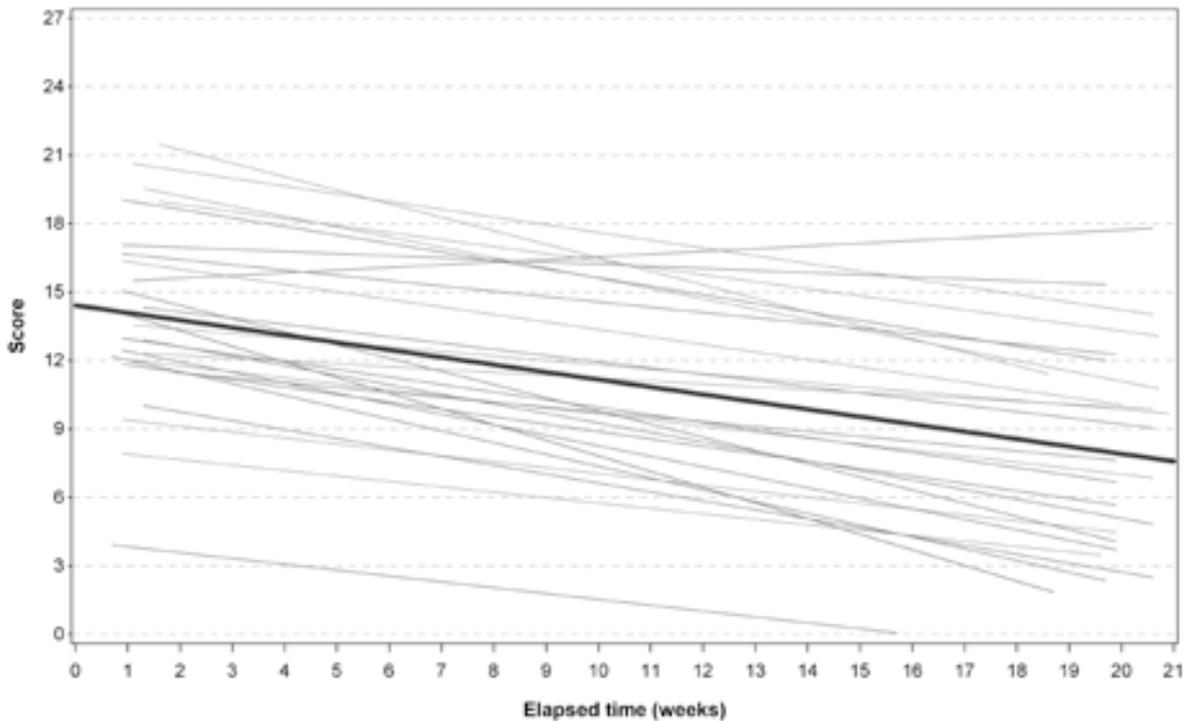
The mixed effects model provided trend lines for each child (random effects) and an overall IYP programme effect trend line (fixed effect). Using the combined behaviour scores, the overall trend throughout the course, represented by the thick line in Figure 8, was towards improved behaviour. From week 0 to week 21 of elapsed time, the estimated IYP programme effect size was $d=1.5^{13}$. This is in the upper range of effect sizes obtained from the Family Interview data.

Trends of behaviour change for each child

The individual trend lines presented in Figure 8 show improved behaviour reported for most of the children. The behaviour of at least one child got worse, while for some children behaviour remained the same. Figure 8 also highlights the wide range of initial behaviours exhibited by the children whose parents began the IYP programme and the varying rates of behaviour change their parents reported throughout the course.

13 An effect size calculation can be estimated based on the slope of the overall regression line, where the baseline behaviour score is taken from the intercept at week 0 of elapsed time ($\alpha=14.4$, $sd=4.58$), and using the regression formula ($\alpha=14.4$ and $\beta=-.32$ points per week) a post-course behaviour score at week 21 of $14.4 - 7.7$ (6.7) is taken. Therefore, based on Cohen's d the effect size of $d=1.5$ is calculated from 14.4 less 6.7 divided by 4.58 .

Figure 8. Predicted combined behaviour scores for each child and for the IYP programme overall



Note: the thick line represents the trend of behaviour change for the IYP programme overall.

Findings from the follow-up data

The Parent Reports at follow-up measured each child's behaviour 6 months after course completion. Both positive and negative behaviour scores were tracked at the follow-up by four weekly interviews over 1 month. An average score was calculated for each child individually and aggregated for all 25 children to provide trend lines from course completion to the follow-up.

As demonstrated by Figure 9 and Figure 10, positive and negative behaviours tended to level out to a horizontal line representing that both positive and negative behaviours had stayed much the same as at course completion. This indicated the effect of the IYP programme had been maintained. This pattern was reflected in most of the 25 individual lines. However, some children demonstrated further gains in behaviour after course completion while others appeared to lose ground.

Figure 11 demonstrates the combined positive and negative behaviours to the follow-up. The overall IYP programme trend (thick line) shows the general pattern of change throughout the programme followed by a levelling out after course completion. Again, while the overall trend confirms the effect of the IYP programme and the sustained improvement to the 6-month follow-up, Figure 11 demonstrates the different paths each child experienced as well as a few examples of children who did not show behaviour change and/or who did not maintain behaviour change.

Figure 9. Predicted positive behaviour scores for each child through the programme to the follow-up

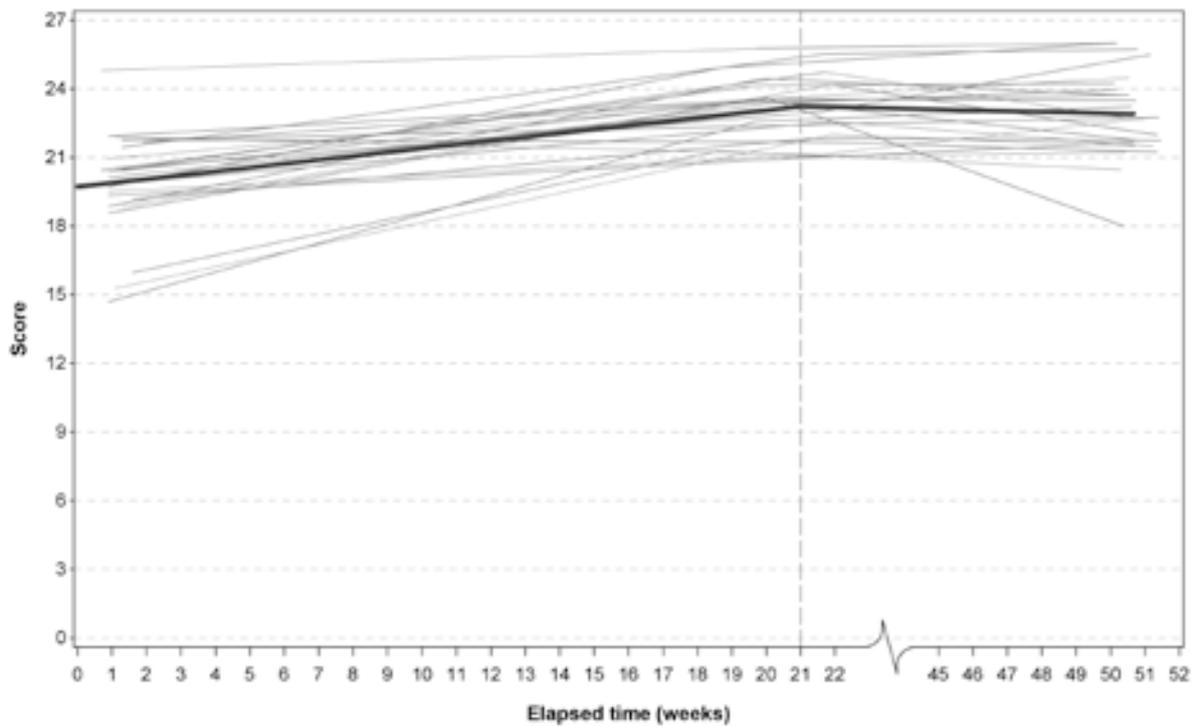
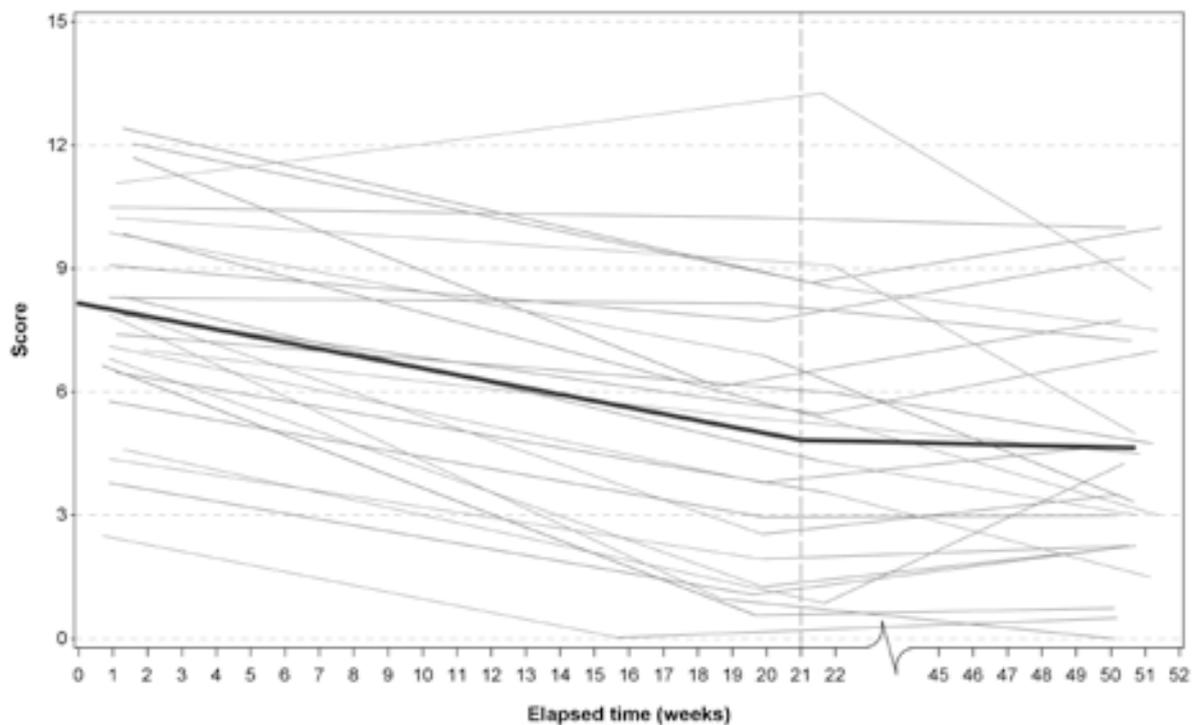
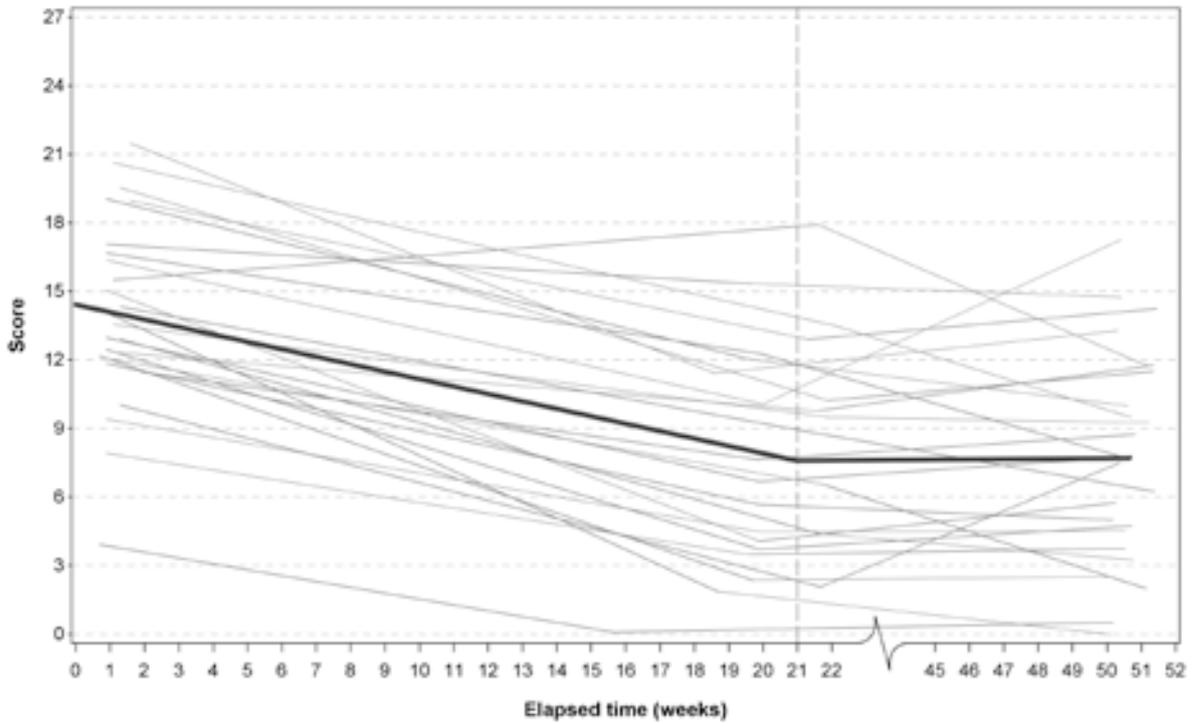


Figure 10. Predicted negative behaviour scores for each child through the programme to the follow-up



Note: a) the vertical line represents the week of IYP course completion; b) the horizontal axis break denotes a shortened time scale where no data were collected (follow-up data collection started in week 49); c) the thick line represents the trend of behaviour change for the IYP programme overall.

Figure 11. Predicted combined behaviour scores for each child through the programme to the follow-up



Conclusions from the Parent Reports

The results from the weekly Parent Reports obtained throughout the programme and at the 6-month follow-up confirmed the efficacy of the IYP programme evident from the Family Interviews.

Furthermore, the individual trend lines for each of the 25 children in the Single Case Studies

highlighted the variability in the children's behaviour and demonstrated that treatment effects differed across children. Parents reported improved behaviour for most children that ranged from large to moderate, while a small number showed no change.

Chapter 9:

Programme fidelity, Group Leader feedback and parent satisfaction



9.1 Programme fidelity: Unified protocols

To obtain similar results to those published by the developers of evidence-based programmes, it is important to ensure the programme is delivered with fidelity. 'Fidelity' means the programme is delivered in its entirety, using all the components and processes recommended by the developer.

Special Education developed a Unified Protocol for the IYP courses in the research that required Group Leaders to complete and return checklists that documented their adherence to the stipulated Incredible Years processes and

programme delivery. The Group Leaders provided checklists for pre-programme preparation, programme delivery and post-programme follow-up and reflection. Only one checklist was missing across the three sites and 12 programmes in the Incredible Years Pilot Study. The completed checklists provided documentation of fidelity to the Incredible Years processes and delivery.

Although Group Leaders cannot change the actual parenting skills taught in the IYP programme without reducing treatment integrity and hence treatment effectiveness, cultural enhancements are encouraged. Indeed, Webster-Stratton (2007) states that the IYP programme is effective across groups of culturally diverse parents due to its principles that promote and guide a culturally responsive structure for delivering the programme to diverse populations. She stresses the important role Group Leaders have in affirming the diversity of the families that take part in IYP training. Cultural diversity is viewed as integral to a successful learning process for parents.

Of particular interest, then, is the documentation of any culturally responsive actions that occurred in the evaluated IYP programmes. The following practices were adopted to a greater or lesser extent in all but one of the 12 courses: manaakitanga, whakawhanaungatanga, greetings, powhiri, waiata, te reo, whakatauki, Māori values (eg wairua), whare tapa whā concepts, wharenui metaphor, asking the group for cultural preferences around the opening and closing of sessions and session activities, consulting Kaitakawaenga/Kaumataua and having a Māori Group Leader. The exception was a course in Canterbury which reported no cultural enhancements at the specific request of the one Māori parent attending the programme.

9.2 Group Leader feedback

Group Leaders and their supervisors held debriefing sessions following the completion of the IYP courses. Their feedback included:

- Conducting two 18-week programmes in 1 year was challenging for the Bay of Plenty as the same Group Leaders facilitated Phase 1 and Phase 2 courses. In Canterbury, the tight timeframe exacerbated problems caused by the earthquake.
- Existing casework impinged on the .5 release time to deliver IYP training, resulting in Group Leaders effectively doing the same amount of work in less time.
- Programme resources were considered adequate and continue to be developed to suit the New Zealand context, particularly for Māori parents.
- Petrol vouchers and taxi chits were provided where required, as was assistance with childcare. Food, warm drinks and treats contributed to the welcoming environment.
- Venues were sometimes an issue when they had to be physically rearranged for each session. In Canterbury, earthquake damage meant inadequate facilities were the only ones available.
- Drop-outs were primarily due to life circumstances (work, sick children, moving) rather than to the IYP programme.
- As is standard practice for IYP programmes, Group Leaders did a lot of work between sessions to support or to find support for parents with personal and family issues.
- Where appropriate, sessions began and ended with a karakia or whakatauki.
- All three sites mentioned the need for Group Leaders to have experience in working with challenging children and families and to have sufficient IYP training to run successful programmes.

9.3 Parent satisfaction with the Incredible Years Parenting programme

Parent satisfaction from the Family Interview

In the baseline interview parents were asked to rate their satisfaction with the way the IYP programme was organised using a 3-point scale: not at all, somewhat and very. These same questions were repeated following course completion. In addition, parents were asked to rate their agreement with 11 statements about the IYP course overall after completion of the 18-week course. These items were based on the Incredible Years Parent Satisfaction Questionnaire, but the scoring was altered from a 7-point to a 3-point scale: not at all, somewhat and a great deal.

Satisfaction data are presented for the 146 parents (N=51 Māori and N=95 non-Māori) who attended some or all of the IYP sessions they were originally enrolled in. Of the 166 parents in the study, two parents did not respond to the question about attendance and a further 18 parents stated they did not attend their designated course.

Satisfaction with the organisation of the IYP programme was generally high for both Māori and non-Māori parents before the course started and at course completion, with no significant difference between the two groups of parents¹⁴ (Table 31). Parents were very satisfied with aspects such as the Group Leader's initial home visit, the information provided, and the professionalism and friendliness of the staff (88% to 100%). Satisfaction with the meeting times (77% to 88%) and suitability of the venue (75% to 78%) were slightly lower. Assistance with attendance was the exception to this high level of satisfaction. Approximately half (45% for Māori and 55% for non-Māori) of the parents were very satisfied with this aspect of the IYP programme's organisation before the start of the course, although the proportion increased (61% and 70% respectively) by course completion.

14 Differences between responses from the parents of Māori and non-Māori children were tested using chi-square tests of significance.

Table 31. Satisfaction with course organisation: opinions before and after the course by ethnicity

Very satisfied with:	Before course started		After course completion	
	Māori (N=51)	Non-Māori (N=95)	Māori (N=51)	Non-Māori (N=95)
Group Leader's first visit	94%	88%	NA	NA
Information provided	94%	90%	86%	91%
Professionalism of staff	100%	95%	88%	93%
Friendliness of staff	100%	98%	98%	96%
Suitability of venue	78%	77%	75%	76%
Course meeting times	85%	88%	77%	84%
Assistance with attendance	45%	55%	61%	70%

At course completion parents were also asked about their satisfaction with the overall IYP programme. The data presented in Table 32 show the generally high level of agreement expressed by both Māori and non-Māori parents with most statements about the Incredible Years approach and the progress they made. There was no significant difference between Māori and non-Māori parents in their level of satisfaction. Of particular note, nearly all parents said they would recommend the IYP course to others (92% for Māori and 94% for non-Māori). The exception to the high level of satisfaction was

the extent to which the IYP training helped with other personal and family problems. Only half of the respondents (49%; 48%) felt the learnings had generalised beyond the target child. Notably, the majority of both Māori parents (84%) and non-Māori parents (94%) agreed a great deal that the IYP programme was appropriate to their cultural identity. None of the Māori participants felt the IYP course was 'not at all' appropriate to their cultural identity. Similar generally high levels of satisfaction with the IYP programme have been reported in other studies (eg Gardner et al 2006; Larsson et al 2009).

Table 32. Satisfaction with the course overall at course completion by ethnicity

Agree 'a great deal' that:	Māori (N=51)	Non-Māori (N=95)
they would recommend the programme	92%	94%
IY was appropriate to their cultural identity	84%	94%
IY was the right approach	71%	79%
they can their achieve goals in the IY programme	71%	77%
they're optimistic that IY will give good results	69%	76%
they're confident they can manage problems by themselves	67%	70%
they can manage problems because of lessons learnt	65%	70%
the problems with their child are improving	55%	67%
they're satisfied with their child's progress	59%	61%
they are using lessons learnt to improve their child's behaviour	53%	60%
IY helped with other personal and family problems	49%	48%

Comments from the primary caregivers reflected the majority view of satisfaction with the IYP programme. For example:

"Hugely increased my confidence as a parent and this is reflected in a better relationship with my child."

"... it has been life changing for our family."

"IY changed the way I speak to my child. I respond rather than yell and scream."

"Really enjoyable – got a lot out of it."

"IY has given me more of a backbone in my parenting."

Parent satisfaction from the in-course measure

Parent satisfaction was also assessed at the end of the IYP course by the Group Leaders. Parents completed a Parent Satisfaction Survey that covered overall programme satisfaction, course content and specific parenting techniques. On a 7-point scale, the vast majority of responses for all categories were six and seven (positive). These ratings reflect parents' assessments that

their children's behaviour had improved and they were confident in their ability to manage behaviour problems. Course content was deemed extremely/useful, as were parenting techniques. These data provided further evidence of the high level of satisfaction with the IYP programme expressed by the majority of parents.

Chapter 10:

Policy implications



The aim of this chapter is to present the views of the Incredible Years EAG on the policy implications and applications of the findings of the Incredible Years Pilot Study described in the preceding chapters. Issues addressed are:

- The efficacy of the Incredible Years Parenting programme in New Zealand
- Extensions of the present research
- The selection of client families
- The effects of context
- Māori/non-Māori differences
- Participants who do not respond to Incredible Years Parenting training
- Site differences
- The need for an effectiveness study of the current Incredible Years Parenting programme in New Zealand
- The organisational implications of the Incredible Years Pilot Study.

10.1 The efficacy of the Incredible Years Parenting programme in New Zealand

The primary purpose of this evaluation was to obtain evidence of the efficacy of the IYP programme in New Zealand contexts using a before-and-after design applied to three sites running established IYP programmes. This research produces good evidence of the efficacy of the IYP programme at the chosen sites with estimates of effect sizes comparable with those found in previous randomised trials and before-and-after designs (see Section 2.2). On the basis of these findings there is adequate evidence to support the conclusion that the IYP programme, if well delivered, is likely to be effective in a New Zealand context.

This research was conducted as a pilot study preliminary to a possible further randomised controlled trial of IYP training. The authors of this report are of the view that a RCT of the IYP programme in New Zealand is probably unnecessary given the strong and consistent evidence from this Incredible Years Pilot Study. Two lines of evidence support this conclusion. First, there is extensive international evidence demonstrating the efficacy of IYP programmes in many settings including the United States, Europe and Asia (for reviews see Kazdin 2007; Ministry of Social Development 2009b). Given this international evidence there are strong *a priori* grounds for believing the IYP programme will be effective in New Zealand. Second, this conjecture was supported by consistent evidence of the efficacy of IYP courses in New Zealand using pre/post and single subject research designs.

Both designs showed that the IYP programme was an effective intervention, with evidence of efficacy being observed up to 6 months post-treatment. It can be argued, therefore, that the time and expensive of a RCT is not warranted.

While there is good evidence the IYP programme has good efficacy in New Zealand in the short to medium term, the Incredible Years Pilot Study findings also raise a number of issues that require consideration in the roll out of IYP courses as part of the Ministry of Education's PB4L programme. These issues are discussed below.

10.2 Extensions of the Incredible Years Pilot Study

The Incredible Years Pilot Study shows evidence of good efficacy in the short to medium term. However, a longer-term assessment of the programme is needed to examine the extent to which the benefits are maintained and the extent to which these benefits may dissipate with the passage of time. To address the issue of the maintenance of intervention gains the Advisory Group recommends the present study is extended to longer-term follow-ups of the children and parents in the present trial. This research design would require the re-assessment of the study group at 1 year and 2 years post-treatment using the Family Interview administered at the 6-month follow-up.

10.3 The selection of Incredible Years Parenting programme client families

In the present study all families enrolled in the IYP courses were included in the evaluation irrespective of their baseline scores on the initial ECBI assessment. This procedure was used to examine the extent to which the benefits of IYP training varied with the level of behavioural problems evident at the first assessment. The findings of the study (see Section 6.2) showed clear evidence of two general trends: i) the IYP programme was effective for client families with children in both the clinical and sub-clinical ranges of the ECBI scales; and ii) IYP training was more effective for client families with children in the clinical range.

These findings have complex implications for the identification of client families eligible for IYP courses. On the one hand it can be argued that, since IYP training is effective for families in both the sub-clinical and clinical ranges of the ECBI, the programme should be offered to all families irrespective of their baseline scores. The potential cost of this decision may be delivering IYP training to a number of families who do not need the programme and will not benefit from it. On the other hand it can be argued that, given the limited resources available for delivering the IYP programme, the greatest benefit can be obtained by delivering it to only those families with children in the clinical range. The potential cost of this decision will be failing to deliver the programme to families in the sub-clinical range who would benefit from it.

There is no clear solution to this dilemma. A case can be made for both treating all client families or for restricting the programme to only those families with children in the clinical range. The EAG suggests that consideration should be given to providing families whose children have ECBI scores in the sub-clinical range with a less intensive parent management training programme than the IYP programme. One such programme is Triple P Level 4 which provides an 8–10 week programme (Ministry of Social Development 2009b).

10.4 The effects of context

One of the important findings of the present study was that the benefits of IYP training were greater for parent reports of behaviour in the home context than for teacher reports of behaviour in the school context. These findings are consistent with a large body of evidence suggesting the rates of problem behaviours in children vary with the context within which the behaviour is assessed (Webster-Stratton and Reid 2010). The important implication of this observation is that programmes targeted at parents are likely to have their greatest impact in changing behaviours at home whereas programmes targeted at teachers are likely to have their greatest impact on behaviours at school. This conclusion suggests the importance of having programmes in different settings and very clearly supports the approach used in the PB4L strategy to invest in both parent behaviour management training and teacher behaviour management training (Ministry of Education 2011).

10.5 Māori/non-Māori differences

An important policy debate has concerned the extent to which the IYP programme is culturally appropriate, and effective, for Māori families (eg Altena and Herewini 2009). The findings of the Incredible Years Pilot Study suggest that, as a general rule : i) IYP training is effective for Māori and ii) the IYP programme is seen as culturally appropriate by over 80 percent of Māori parents.

At the same time, small but statistically detectable differences were found between Māori and non-Māori client families, with the effect sizes for Māori being slightly smaller than those for non-Māori. These findings suggest that, while IYP is an effective programme for Māori, there is a case for further investment in programme development to improve the programme's delivery and to maximise the gains for Māori. The Nga Tau Miharo programme resource developed by the Werry Centre is an important innovation in this area. It provides a manual and DVD for the delivery of IYP training to Māori families. Nga Tau Miharo was not in place during the current evaluation and it is possible that, had it been available, the small ethnic differences found in this study would have been eliminated.

More generally, views about the effectiveness of the IYP courses for Māori have been divided. Previous research (Fergusson et al 2009) found the programme was equally effective for Māori and non-Māori whereas Māori commentators have suggested it is not acceptable for Māori unless it is substantially changed (eg Altena and Herewini 2009). The findings of this evaluation suggest a middle position is the most tenable. Specifically, while there is good evidence to suggest the IYP programme is effective for Māori families there is also evidence indicating the benefits of IYP training are slightly but detectably smaller for Māori than the benefits for non-Māori. These findings clearly justify the delivery of the IYP programme to both Māori and non-Māori but also suggest the need for continuing investments in improving the delivery and efficacy of IYP training to Māori families. It is likely the implementation of Nga Tau Miharo will provide an important step in this process.

10.6 Participants who do not respond to the Incredible Years Parenting programme

While the evaluation has shown IYP training has good efficacy as a parent behaviour management programme, it must be recognised the programme is not effective for all families. An important finding from the Single Case Studies analyses reported in Section 8.1 was that in the region of one in four families enrolled in IYP courses failed to report significant change in the client child's behaviour. This suggests that, while the programme benefits the majority of families receiving IYP, it does not benefit all families. This is a pattern common for interventions of this type. Two extensions to the current IYP programme are indicated. The first is the need to identify those families and children who have not benefited from the IYP programme. The second is to develop further programmes and resources to meet the needs of families who have attended IYP courses but who have not received the expected benefits.

10.7 Site differences

An unexpected finding from this evaluation was the presence of significant between-site differences in the efficacy of the IYP programme, with one site producing greater benefits than the other two sites. The reasons for these differences are unknown. Nevertheless, the findings have important policy implications since they highlight the need for well-developed processes to audit the delivery of the IYP programme and to compare the performances of different providers with benchmarks based on the present evaluation. An important next step in the development of IYP training in New Zealand is the need to conduct a national effectiveness study which examines a range of issues relating to variability in the outcomes of the IYP programme.

10.8 The need for an effectiveness study

The various issues raised above regarding ethnic differences (Section 10.5), clients who do not improve (Section 10.6) and between-site differences (Section 10.7) highlight the need for a further evaluation of the IYP programme using data from the Incredible Years National Register held by the Ministry of Education. This register contains data on client ethnicity, pre and post scores on the ECBI and measures of client satisfaction. The availability of this database makes it possible to examine the extent to which the findings from the present evaluation can be applied to the IYP programme as it is being delivered by multiple providers in multiple centres. Some of the important questions regarding the IYP programme include:

- Overall, how effective is the IYP programme in reducing childhood problem behaviours?
- To what extent is the programme effective for Māori and non-Māori?
- What proportion of children enrolled in the programme fail to show significant improvements in behaviour?
- How much variation is there in the efficacy of the IYP programme among different sites?
- How well is the IYP programme being delivered at different sites?

It is the view of the Advisory Group that using the Incredible Years National Register data to address these and related questions is an essential next step in the development and evaluation of IYP courses in New Zealand. While the Incredible Years Pilot Study has provided in-depth information about the performance of the IYP programme at three sites, this evaluation fails to provide an overview of how well the programme is being delivered at a national level. Learning about this is important both for the quality control of programme delivery and for the further development of the IYP programme in a New Zealand context.

10.9 Organisational implications

An important feature of this evaluation was to develop and explore a new model for evaluating government-funded programmes for the prevention, treatment and management of conduct problems in childhood and adolescence. This model was based on a consortium approach that brought together the skills of a number of groups: the service providers delivering the programme; the government researchers who evaluated the programme; academic advisors with expertise in research design and analysis; and Māori researchers.

This new model was developed in response to concerns about previously used evaluation models in which new policies and programmes were evaluated using a tendering process. This resulted in external groups and agencies conducting evaluations on behalf of the Government. These processes were seen as being less efficient because of the limited participation of government service providers and research/policy makers in the evaluation process. The model proposed for the assessment of the IYP programme involved a collaborative process that included: representatives of the service developer (Ministry of Education); the funder of the fieldwork (Ministry of Health); the research evaluation team (Ministry of Social Development); and academic advisors with expertise in the area of conduct problems research and the collection of developmental data.

The present study provides a case history of the effectiveness of this model in producing a well-designed and effective evaluation of the IYP programme. The EAG believes the success of the evaluation speaks for itself and the model developed in the process of evaluating the IYP programme can be generalised to many other contexts. The critical features of this model are:

- It requires developing a consortium in which different members bring different skills to the project.
- The control of the evaluation remains within government and contributes to building government research capacity.
- The costs of research are reduced since the conduct of the research does not require the payment of the overhead and profit costs associated with outside tendering.

It is the view of the EAG that this model of evaluation should continue to be used in the evaluation of other programmes targeted at the prevention, treatment and management of childhood conduct problems. Some important areas for further development include:

- Longer-term follow-up of the families studied in this evaluation (see 10.3)
- Evaluation of the effectiveness of the IYP programme using the Incredible Years National Register data (see 10.8)
- Evaluation of the Incredible Years Teacher programme (see 10.4).

The EAG firmly believes the proposed programme of research will be better achieved by maintaining the current evaluation team and EAG who have overseen the Incredible Years Pilot Study, thus transferring the skills and learning acquired over the course of this evaluation to further projects.

10.10 Summary of Evaluation Advisory Group recommendations

- The Incredible Years Pilot Study should be viewed as providing adequate evidence of the efficacy of the IYP programme in New Zealand.
- Investments should be made into the longer-term follow-up of the families participating in this evaluation, with further assessments being made at 1 year and 2 years post-treatment.
- Consideration should be given to providing parent management training programmes, such as Triple P Level 4, that are less intensive than the IYP programme for client families with children whose behaviour scores on the ECBI scales place them in the sub-clinical range.
- Consideration should be given to developing parallel research to evaluate the Incredible Years Teacher programme to determine the benefits of this programme for behaviours in the classroom and for behaviours at home.
- Further investment is needed to increase the efficacy of the IYP programme for Māori families.
- There is a need to develop further programmes and interventions to meet the needs of families whose children do not show sufficient behaviour change at the end of the IYP programme.
- Investment should be made into a national effectiveness study using the Incredible Years National Register data held by the Ministry of Education.
- The present consortium structure for evaluating programmes targeted at childhood conduct problems should be maintained. A future research programme is outlined in Section 11.3.

Chapter 11:

Summary and conclusions



11.1 Summary of findings related to research questions

The findings from the Incredible Years Pilot Study presented in this report provide the following evidence relating to the research questions stated in Chapter 3.

1. *Programme efficacy: To what extent was there evidence to suggest that the IYP programme was effective in a New Zealand context?*

- Before-and-after comparisons of families enrolled in the IYP programme produced effect size estimates for change in child behaviour, parenting practice and conflict and relationships consistent with previous Incredible Years evaluations using randomised trials. These RCTs reported effect sizes for child behaviour change ranging from $d=.48$ to $d=.89$ and effect sizes for parenting practice ranging from $d=.38$ to $d=1.44$.
- Parents receiving IYP training reported significant improvements in the following areas: child behaviour (effect sizes $d=.51$ to $d=.96$), parenting practices (effect sizes $d=.26$ to $d=.83$) and conflict and relationships (effect sizes $d=.21$ to $d=.60$).
- Effect sizes for improvements in parenting practices and family relationships were typically lower ($d=.26$ to $d=.71$) than effect sizes for parent-reported improvements in child behaviour. This indicated small changes in parenting practice may produce substantial improvements in child behaviour.

- The Parent Report Single Case Studies reflected the findings of improved child behaviour at course completion reported for the main study. Furthermore, these data highlighted the variability in child behaviour and showed that treatment effects were not equal for all children in the study.
 - Small to medium reductions in caregiver stress and anxiety were reported at post-course and follow-up, as were fewer child visits to health services and fewer stressful life events. Parents also reported a decrease in depressive symptoms at follow-up.
 - Before-and-after comparisons of scores on the Social Development Scale completed by teachers produced small effect sizes ($d=.17$ at post-course and $d=.29$ at follow-up). This suggested some of the improvements in child behaviour carried over to the school setting.
2. *Sustainability of gains: To what extent were any benefits of IYP sustained over time?*
- The reported improvements evident at the completion of the IYP courses were generally sustained at the 6-month follow-up. Significant improvements were maintained in the following areas: child behaviour (effect sizes $d=.56$ to $d=1.0$), parenting practices (effect sizes $d=.25$ to $d=.79$) and conflict and relationships (effect sizes $d=.15$ to $d=.59$).
 - The Parent Report Single Case Studies reflected the findings of improved child behaviour sustained to follow-up reported for the main study.
3. *Benefits for Māori: To what extent were the programme outcomes for Māori similar to or different from the outcomes for non-Māori?*
- The evidence suggests the IYP programme was both effective for and accepted by Māori families. However, at both post-course and follow-up parents of Maori children reported slightly smaller improvements in child behaviour and slightly smaller reductions in negative parenting practices on the individual outcome measures than did parents of non-Maori children. These differences were not significant.
- On the combined outcome measures at programme completion the differences between Māori and non-Māori children were not significant. However, at follow-up a small but statistically significant difference between responses by parents of Māori and non-Māori children was detected for the overall child behaviour outcome measure ($p=.025$). This suggests there may be a particular challenge in maintaining the benefits of the IYP programme for Māori families.
4. *Programme entry measures: To what extent do the outcomes of the IYP programme vary with the child's behavioural adjustment at baseline?*
- Children with pre-course scores in the clinical range of ECBI Intensity improved to a greater extent at post-course on all child behaviour measures except Social Competence than children with scores below the clinical cut-off. Significant differences in reported Conduct Disorder and ODD remained at follow-up, whereas differences in ADHD, Self-control and Anxiety/Withdrawal were no longer evident. This perhaps reflected the continued improvement for children in the sub-clinical group.
 - The greater improvement reported for children in the clinical range on the pre-course SCS on measures of Conduct Disorder and ODD at post-course were sustained at follow-up. Self-control and Social Competence had also improved to a greater extent for children in the SCS clinical group than for children in the sub-clinical group 6 months following course completion.
5. *Between site differences: To what extent do the outcomes of IYP vary with the site at which the programme was delivered?*
- The IYP courses were effective at each of the three sites. However, the significant interaction between the global child behaviour measure and site ($p<.002$) indicated the courses delivered in Mid-Central were particularly effective. The reason for this difference is not known.

6. *Client satisfaction: To what extent do the parents who received the IYP training see the programme as: a) helpful; b) appropriately organised; c) effective in dealing with their child's behavioural problems; d) culturally appropriate?*
- Parent ratings in both the Family Interview and the in-course questionnaire revealed high to moderate satisfaction with the IYP programme for both Māori and non-Māori parents on most measures.
7. *Other findings*
- The study reached the target group of children and parents.
 - The programme was delivered with fidelity to the delivery, processes and techniques stipulated by Incredible Years.

11.2 Conclusions

These results lead to the following conclusions:

- The IYP programme can be successfully implemented in New Zealand and retain its general level of effectiveness in reducing conduct problems and enhancing parenting skills.
- Although these findings are generally consistent with the RCT studies summarised in Chapter 2, the before-and-after design of the Incredible Years Pilot Study may lead to an over-estimation of effect sizes. The comparison presented in Table 33 of the effect sizes for child behaviour and parenting practices obtained post-treatment in this research with the effect sizes reported in a Norwegian RCT before-and-after study (Larsson et al 2009) is reassurance the present study provides an adequate estimation of the efficacy of the IYP programme in New Zealand. On the basis of these data it seems unlikely a RCT study of IYP courses in New Zealand would produce results materially different from the estimates provided by the before-and-after comparisons reported here.
- The findings provide a series of performance benchmarks against which the future roll out of the IYP programme in New Zealand should be assessed.

Table 33. Effect sizes for child behaviour and parenting practices reported in Larsson et al (2009) and the Incredible Years Pilot Study

Measure	Larsson et al (2009)		Incredible Years Pilot Study	
	Effect size	Instruments used	Effect size	Instruments used
Child Behaviour	d=.65	ECBI Intensity	d=1.00	ECBI Intensity
Parenting Practice				
Positive parenting	d=1.44	Alabama Parenting Questionnaire	d=.46	Parenting Practices Interview
Harsh discipline	d=.61	Arnold-O'Leary Parenting Scale	d=.59	
Inconsistent discipline	d=1.05		d=.54	

11.3 Future research

The findings from the Incredible Years Pilot Study leave important unanswered questions that need to be addressed. The following future research programme is recommended:

- A longer-term follow-up of the children and parents in the present trial to examine the extent to which programme benefits are sustained and the extent to which these benefits may dissipate over time. The research design would extend the present study to re-assess the participants at 1 year and 2 years post-treatment using the Family Interview administered at the 6-month follow-up.
- A national effectiveness study based on the Incredible Years National Register data held by the Ministry of Education to investigate whether or not the effect sizes for the changes in the ECBI and SCS measures obtained in the Incredible Years Pilot Study can be replicated on national aggregated data and for individual courses at all sites. The study should evaluate:
 - i) the overall effectiveness of the IYP courses;
 - ii) the IYP programme's effectiveness for Māori and non-Māori;
 - iii) possible variation among providers in the effectiveness of IYP training;
 - iv) the proportion of client families who fail to show improvement;
 - v) the adequacy of the delivery of the IYP programme.
- The performance benchmarks reported here provide measures against which the future roll out of the IYP programme should be judged. Consistent and rigorous quality control monitoring of provider outcomes will ensure the programme is effective and the funding is well spent.
- Further investigation is required to understand the circumstances that produced site variation in outcomes to inform best practice principles and to ensure quality control. The site differences found in the Incredible Years Pilot Study demonstrate that differences may occur in programme outcomes for a variety of reasons even at well-established and well-functioning sites. This finding points to the need for consistent quality control of provider delivery and monitoring outcomes. It indicates the need for fidelity checks and an audit of each IYP programme in every site using the benchmarks provided by the Incredible Years Pilot Study to assess between-site heterogeneity and to determine the extent of variation.
- The suggestions that the benefits of IYP training may be slightly larger for non-Māori than for Māori families and that there appears to be a particular challenge for Māori in maintaining behaviour change point to the need for further research to examine Māori/non-Māori differences in greater detail. Findings from that research will inform the work on Māori enhancements for IYP courses. Regular monitoring of changes in ECBI and SCS scores for Māori and non-Māori children at every site is necessary to ensure the effectiveness of the IYP programme for Māori is maximised.
- An evaluation of Ngā Tau Miharo ō Aotearoa, the Incredible Years resources developed by the Werry Centre for Māori Group Leaders, is required to ascertain whether or not differences in outcomes for Māori and non-Māori families are evident in courses using these resources. A comparison of before-and-after ECBI and SCS scores at sites with Māori families that use Ngā Tau Miharo ō Aotearoa with sites that do not use these resources will reveal whether or not the differences reported here remain for IYP courses that have implemented culturally responsive refinements for Māori.
- The lower effect sizes found for the teacher measure suggests parent management training programmes like Incredible Years, although beneficial in reducing behaviour problems in schools, are not sufficient. It is likely parent management training should be used in conjunction with school-based programmes for maximum return in the school environment. The evaluation of the Incredible Years Teacher programme is required.
- Although the evidence presented in this report strongly supports the effectiveness of the IYP programme in the New Zealand context, it is possible existing parenting programmes produce similar effects. Further research is required to compare the cost benefit of providing the IYP programmes against the cost benefit of other parenting programmes.

Appendix 1.

The Incredible Years Pilot Study Evaluation Advisory Group

Dr John Church	Adjunct Senior Fellow, Psychology Department, University of Canterbury
Professor David Fergusson	Executive Director, Christchurch Health and Development Study, University of Otago
Professor Richie Poulton	Director, Dunedin Multidisciplinary Health and Development Research Unit, University of Otago
Dr Sarah Hayward	Formerly Senior Project Manager, Ministry of Health; currently Principal Advisor, Ministry of Justice
Brian Coffey	Group Manager, Strategy (Special Education), Ministry of Education
John McGovern	Service Manager, Ministry of Education (Special Education)
Ross Mackay	Principal Advisor, Ministry of Social Development
Dorian Gray	Senior Analyst, Ministry of Social Development
Fiona Sturrock	Research Lead, Ministry of Social Development (Chairperson)

Appendix 2.

Source instruments for the Incredible Years Pilot Study

See The Incredible Years Pilot Study Technical Appendix (Gray 2013) for more details on the source instruments used in the study.

The Family Interview Questionnaire

The Family Interview Questionnaire is based on a composite of instruments determined by the outcome variables of primary interest, and the mediating and contributing variables associated with them. There are four main blocks of outcome variables underlying the interview:

- Child behaviour – decreased problem behaviour
- Parent behaviour – increased positive parenting
- Family relationships – improved parent/child and parent/partner relationships
- Other relevant outcomes – parental depression, anxiety and stress; stressful life events.

Child behaviour

Child behaviour was measured with four instruments adapted for use in this context. Approval was granted to adapt the Eyberg Child Behaviour Inventory with the payment of copyright fees. Scale items were all rated on a consistent 3-point scale: not at all, somewhat, a great deal.

The Eyberg Child Behaviour Inventory (ECBI) (Eyberg and Ross 1978; Eyberg 1980) is a 36-item inventory measuring child problem behaviours perceived by the caregiver, and is normed for children aged 2–16 years. The ECBI measures

the number of problem behaviours and the frequency with which these behaviours occur. The scale demonstrates good stability, with reliability coefficients from 0.86 (test-retest) to 0.98 (internal consistency) (Robinson et al 1980). Good convergent validity is demonstrated by significant correlations with the Child Behaviour Checklist and the Parenting Stress Index (reported in Jones et al 2007). The ECBI is a well-respected and well-used measure for assessing the frequency of conduct problem behaviour that is reliable and valid, and identifies change due to intervention over time (Rhodes 2009). It has been used extensively within the field of parent training intervention including in several studies of the IYP programme (Webster-Stratton 1998; Scott et al 2001; Hutchings et al 2007). The ECBI has two scales, the Intensity scale and the Problem scale, but the latter was not used in the Family Interview Questionnaire.

The Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997) is a 25-item inventory designed as a behavioural screening measure to assess the occurrence of particular behaviours associated with conduct problems, hyperactivity, emotional symptoms and peer problems in children. The scale has demonstrated good stability as judged by internal consistency (mean Cronbach's $\alpha=0.73$), cross-informant correlation (mean=0.34) and test-retest stability after 4–6 months (mean=0.62) (Goodman 2001). There are versions for parents and teachers. Both versions contain five subscales: emotional problems, conduct problems, hyperactivity, peer problems and pro-social behaviour. The additional Impact Supplement scale, which measures how

the caregiver or teacher perceives the child's behaviour and the impact of the problem on the child's daily life, was not used in the Incredible Years Pilot study. The SDQ is a screening measure and not as sensitive to clinical change as the ECBI.

The Social Competence Scale – Parent Version (P-COMP), developed by the Conduct Problem Prevention Research Group (Fast Track), consists of 12 items that assess the child's positive social behaviours as perceived by the parent. It includes a measure of frustration tolerance and communication skills. This instrument is also used by the Incredible Years Group Leaders to assess the participating parents' children (CPPRG 1995; Corrigan 2002).

Some of the child behaviour items were based on those used in the Early Start Evaluation (Fergusson et al 2005b). These items have been tested in a New Zealand context and were found to provide robust measures of child behaviour.

Parenting practices

The Alabama Parenting Questionnaire (APQ) was designed to tap the parenting dimensions that are risk factors associated with child conduct disorder (Shelton et al 1996). It consists of 42 items, scored with a 5-point rating scale, that load onto five subscales: parental supervision, positive parenting, corporal punishment, parental involvement and inconsistent discipline. Dadds et al (2003) evaluated the APQ with a community sample of Australian children aged 4–9 years. The results showed good internal consistency, validity and test-retest reliability for the measure. Scale items were all rated on a consistent 3-point scale: never, sometimes, often.

The Arnold-O'Leary Parenting Scale (Arnold et al 1993) is a 30-item inventory of parenting competencies that measures dysfunctional and/or ineffective parenting practices of parents with younger children. The scale yields an overall score and three revised subscale scores of dysfunctional strategies used by parents tackling problem behaviour. 'Laxness', refers to insufficient monitoring of the child and her/his behaviour,

allowing rules to go unenforced or providing positive reinforcement for misbehaviour. 'Over-reactivity', refers to displays of anger, meanness or irritability. 'Hostility' refers to the use of verbal or physical force. The scale has adequate internal consistency and has been found to have good test-retest reliability. It looks specifically at parenting practices rather than at child misbehaviour. This measure provides an accurate assessment of parental discipline strategies, yielding correlations with independent blind observation of parenting behaviours. Scale items were all rated on a consistent 3-point scale: never, sometimes, often.

Relationships

The relationship between the caregiver and their child and partner were measured using three instruments. Approval was granted to adapt the Conflict Tactics Scales CTSPC and CTS2 with the payment of copyright fees.

The Conflict Tactics Scales (Straus et al 1993; Straus et al 1998) have been used for decades to evaluate violence within families and intimate relationships. Two updated versions, the Conflict Tactics Scale Parent/Child (CTSPC) and the Revised Conflict Tactics Scale 2 (CTS2), were used in the Incredible Years Pilot Study. The CTSPC items related to the parents' use of verbal aggression and physical assault in their relationships with their children. Scale items were all rated on a consistent 3-point scale: never, less than once a week, once a week or more. The CTS2 focused on violence between the parents. Scale items were all rated on a consistent 3-point scale: never, sometimes, often.

The Parent Problem Checklist (PPC) was developed as a measure of inter-parental conflict, especially as it relates to the parents' ability to cooperate and to act as a team in performing the executive parenting functions within the family. It contains 16 items measuring the presence or absence of parental disagreement over rules and discipline for child misbehaviour, the occurrence of open conflict over child-rearing issues and whether or not parents undermine each other's relationships with the children. The PPC is a unidimensional measure with moderately high internal consistency and high test-retest reliability (Dadds and Powell 1991). Scale items were all rated on a consistent 3-point scale: not at all, somewhat, a great deal.

The Partner Attachment Scale measures the quality of the relationship between parents. The items used in the study are based on a selected series of items from Braiker and Kelley (1979) as used in the Christchurch Health and Development Study (CHDS) 21-Year Interview. Scale items were all rated on a consistent 3-point scale: doesn't apply, somewhat applies, definitely applies.

Child health

A question from the Early Start Field Trial 5-Year Follow-up questionnaire (Fergusson et al 2005b) assessed the use of child health services.

Parental depression

The Composite International Diagnostic Interview (CIDI) version of the Diagnostic and Statistical Manual of Mental Disorders (DSM) assessed depression symptoms (World Health Organisation 1993) and the anxiety and stress scales of the Depression Anxiety Stress Scale (DASS) measured anxiety and stress (Lovibond and Lovibond 1995).

Life events

The occurrence of stressful life events was measured using items from the Early Start Field Trial 5-Year Follow-up questionnaire (Fergusson et al 2005b), which was an expansion of the items used in the Christchurch Health and Development

Study based on Holmes and Rahe (1967). These items have been tested in a New Zealand context and have been found to provide a robust measure of a family's exposure to stressful and adverse life events.

Cultural participation

Cultural identity and participation for parents of Māori descent was measured using a Mason Durie questionnaire adapted for the Christchurch Health and Development Study (Broughton et al 2000).

Parent satisfaction

Parent satisfaction was assessed with items from the Incredible Years Parent Program Satisfaction Questionnaire Basic Parent Programme (1–4), as well as specific questions about expectations of and satisfaction with the IYP programme.

Demographics

The background information obtained in the Family Interview Questionnaire covered such things as family size, housing, education, and employment. The New Zealand Deprivation Index 2006 was determined through 'meshblocking' the addresses of the Family Interview participants and extracting the relevant New Zealand Deprivation Index to each meshblock.

The Teacher Questionnaire

Child behaviour in the classroom was assessed using the Canterbury Social Development Scale (Church et al 2006). This is a 30-item rating scale which consists of brief descriptions of 15 antisocial behaviours and 15 positive social behaviours which are likely to occur in the classroom (or early childhood centre). Each item is rated on a 5-point scale from 'never' through to 'very frequently'. The CSDS exists in four versions: one for the preschool level, one for Years 1 to 4, one for Years 5 to 8, and one for Years 9 and 10. The versions differ only with respect to age appropriate wording for individual items. Teacher responses to the antisocial items are reverse

scored so the total score provides a standardised measure of the level of social development. For the present study the rating scale was changed from a 5-point to a 3-point scale which meant the clinical cut-offs for the scale could not be used.

The Single Case Studies

- The weekly Parent Report was based on the Parent Daily Report (PDR) developed at the Oregon Learning Centre (Chamberlain and Reid 1987). The PDR is administered to parents by the Interviewer via the telephone. The caller asks parents to report whether or not any of a list of child behaviour problems had occurred at home in the previous 24-hour period. If a behaviour has occurred, parents are asked how stressed the behaviour made them feel. This measure was chosen because it is well suited for repeated daily assessments and because it is especially effective in picking up low-frequency behaviour.

The weekly Parent Report in the Incredible Years Pilot Study was a 10-minute interview using a list of 23 questions on positive behaviours and 26 questions on negative behaviours to which parents answered 'yes' or 'no' regarding their child's behaviour the previous day. Parents were not asked about stress.

- For the Direct Observation trial, digital recordings were made of parent-child interaction during three consecutive 5-minute play activities: child initiated play, parent directed play and packing up. Recordings were made in a clinic using a standard collection of eight toys. Each 15-minute recording was coded using a 3-column coding form. Each line on the recording form recorded a single parent-child or child-parent interaction. Interactions were coded using commonly employed coding categories similar to those used by Eyberg (2010) and Forehand and McMahon (1981). The first column was used to record parent initiations (and responses to child initiations), the second column was used to record child responses and child

initiations, and the third column was used to code positive and negative reactions made by the parent in response to child statements and child behaviour (Church 2012b).

- Parent initiations (Column 1) were coded into three classes: compliance requests, questions and other initiations, and other talk.
- Child behaviour following a compliance request (Column 2) was classified as compliance/non-compliance and child responses to other parent talk were classified as appropriate/inappropriate reply. Other child talk was classified as other talk.
- Parent reactions (Column 3) to child responses to compliance requests, to child replies, and to child behaviour were classified either as positive reactions or as negative reactions.

All recording and all coding was done by a trained early childhood lecturer working under the supervision of Dr John Church at the University of Canterbury.

Appendix 3.

Statistical methods

Programme efficacy (outcome) and sustainability

The programme efficacy or outcome is determined by testing for a linear trend of change in the mean scores from the baseline through the mid-course to the post-course interview (observation points or time). The sustainability of programme outcomes is determined by testing for a linear trend between the baseline and the follow-up interviews. A general linear model is used to test the hypothesis that there was a linear trend in the scores by applying an Analysis of Variance (ANOVA) for repeated measures. 'Repeated measures' refers to the comparing of the three interviews over time for the same children. If the p-value for time is small ($p < .05$) then the null hypothesis that there was no linear trend of change through the course can be rejected, which suggests the programme was effective in changing behaviours.

The algebraic form of the general linear model can be expressed as:

$$Y_{ij} = \beta_0 + \beta_1 Time_j + \mu_{ij}$$

Y = outcome (Conduct Disorder, etc)

i = observations 1 to N

j = observation points (Time) 1, 2, and 3
(or 1 and 4 for the follow-up)

Programme effect sizes

In addition to the repeated measures test, the analysis includes 'effect size estimates' to examine the size of the change in behaviours. Effect size calculations are Cohen's d (expressed as ' d '). Cohen's d is the standardised difference between means of proportions (Cohen 1977). Cohen suggests an effect size of $d = .20$ is small, an effect size of $d = .50$ is medium, and an effect size of $d = .80$ is large. These interpretations are arbitrary but provide an indication of how large the behaviour change is. A positive effect size ($d > 0$) indicates improved behaviours while a negative effect size ($d < 0$) indicates worsened behaviours.

The effect size for each measure is calculated as the difference between the baseline and comparison mean scores divided by the standard deviation of the scores for that measure at baseline, where the comparison is either post-course (interview 3) or follow-up (interview 4). In both cases, the baseline standard deviation is used, which may produce conservative estimates of effect size.

Effect size between baseline and post-course results:

$$d = \frac{\bar{x}_1 - \bar{x}_3}{s_1}$$

Effect size between baseline and follow-up results:

$$d = \frac{\bar{x}_1 - \bar{x}_4}{s_1}$$

Where:

\bar{x}_1	Mean score at interview 1 (baseline)
\bar{x}_3	Mean score at interview 3 (post-course)
\bar{x}_4	Mean score at interview 4 (follow-up)
s_1	Standard deviation of scores at interview 1 (baseline)

Confidence intervals on the effect sizes

Because the effect sizes are estimates based on samples, 95% confidence intervals (CI) are calculated to provide a range for the estimated effect sizes. The confidence intervals are calculated using a SAS® Macro developed by Hess and Kromrey (2003), which calculates the confidence intervals based on three inputs: i) the effect size; ii) the sample size of the effect size calculation at the comparison interview (treatment); and iii) the sample size of the effect size calculation at baseline (control). The comparison sample size is the sample size at the post-course or follow-up interview, which is used in the macro instead of the baseline or control sample size as it is usually the smallest sample size. Hence the confidence limits calculated are conservative.

The calculated confidence intervals will be fairly wide for small sample sizes. In some cases of small sample sizes and small effect sizes, the lower confidence limit may be less than zero. The p-values for the tests of a linear trend should be used as evidence of change and not the effect size confidence intervals.

Programme outcomes between sub-groups

Programme outcomes are tested between sub-groups such as Māori/non-Māori children and clinical/sub-clinical Conduct Problem, for instance. This is because any improvements from the programme could differ across the sub-groups being compared. The test is based on the interaction between the sub-group and the change in the outcome scores as determined by a Multivariate Analysis of Variance. Evidence of a significant interaction between sub-group and the change in outcome scores is demonstrated where $p < .05$. The absence of a significant interaction was taken to imply the relative change was broadly similar across the sub-groups being compared.

In the example of the Māori/non-Māori ethnicity sub-group, the algebraic form of the general linear model can be expressed below, where the hypothesis that the programme outcomes were broadly similar between Māori and non-Māori is taken from the p-value for the ethnicity by time interaction.

$$Y_{ij} = \beta_0 + \beta_1 Time_j + \beta_2 Ethnicity_i + \beta_3 (Ethnicity \cdot Time)_{ij} + \mu_{ij}$$

Where:

Y = outcome (Conduct Problem, etc)

j = observation points 1 and 3 (or 1 and 4)

i = observations 1 to N

Corrections to multiple hypothesis testing

The tables reported contain multiple hypotheses testing, which requires some consideration of the possibility of low p-values occurring by chance. Two methods are employed to reduce the chance of false-positive conclusions from hypothesis testing. First, a Bonferroni adjustment to the $\alpha=.05$ is applied to the overall programme and sustained

outcome tables (Table 34). It was thought a Bonferroni adjustment was sufficient for the overall outcome tables because of their strong findings. Table 34 shows the Bonferroni adjustments made to the $\alpha=.05$ level for each of the programme outcome and sustained outcome tables.

Table 34. Bonferroni adjustments applied

Tables	Bonferroni adjustment to $\alpha=.05$	Number of tests
Child Behaviours (programme outcomes)*	p-values <.006 are significant	8
Child Behaviours (sustained outcomes)	p-values <.008 are significant	6
Parenting Practices	p-values <.005 are significant	9
Conflict Tactics/Relationships	p-values <.006 are significant	8
Other outcomes	p-values <.010 are significant	5

*Includes the ECBI and SCS scores as collected by the IYP course Group Leaders.

Second, the sub-group interaction tables are summarised with a doubly Multivariate Analysis of Variance, which tests the hypothesis that there is no interaction between the broad outcome being tested (combining all the measures in a table), observation point (time), and the sub-group of interest. The purpose of this is to ensure, where only a few of the measures show interactions between sub-groups, the sub-groups were or were not similar.

The multivariate analysis provides a p-value that represents a broad outcome for each table such as Child Behaviour or Parenting Practices as a group. In the example of the Māori/non-Māori ethnicity sub-group, the algebraic form of the general linear model can be expressed below, where the hypothesis that $\beta_3^k = 0$ for all $k=1$ to K where each k is an individual measure such as Conduct Disorder, Oppositional Defiant Disorder, etc.

$$Y_{ij}^k = \beta_0^k + \beta_1^k Time_j + \beta_2^k Ethnicity_i + \beta_3^k (Ethnicity \cdot Time)_{ij} + \mu_{ij}^k$$

Where:

Y = outcome (Child Behaviour, etc)

j = observation points 1 and 3 (or 1 and 4)

i = observations 1 to N

k = outcome k ($k= 1$ to K , Conduct Disorder, Oppositional Defiant Disorder, etc)

Summary of statistics and tests

The following table summarises the statistics calculated, and the tests for linear trends that provide the evidence to reject the null hypothesis of no linear trend in scores throughout the Incredible Years Pilot Study, meaning no improvement.

Table 35. Summary of analysis methods used

Outcome results for each measure	Statistical tests
<p>Scores from each measure</p> <ul style="list-style-type: none"> Mean (average) scores and standard deviations (std) at baseline, mid-course, and post-course. Mean scores and std at baseline and follow-up. <p>Effect sizes (Cohen’s d)</p> <p>The effect sizes ‘d’ with 95% CI:</p> <ul style="list-style-type: none"> Effect size of the programme outcomes, calculated for baseline and post-course. Effect size of the outcomes 1 year after baseline, calculated for baseline and the follow-up. 	<p>Test for a linear trend of improvement</p> <p>A repeated measures Analysis of Variance is used to test for a linear trend through the observation points. A p-value, with a Bonferroni adjustment, provides evidence of a linear trend in the mean scores.</p>
<p>Effect sizes of sub-groups for each measure</p> <ul style="list-style-type: none"> Effect size of the programme outcomes by sub-groups with 95% CI, calculated for baseline to post-course. Effect size of the outcomes 1 year after baseline by sub-groups with 95% CI, calculated for baseline to follow-up. 	<p>Test for interactions between sub-groups and time (changes between observations over time)</p> <p>A repeated measures Analysis of Variance is used to test for interactions between sub-group and time or the change in outcomes scores:</p> <ul style="list-style-type: none"> through baseline to post-course observations baseline with follow-up observations. <p>A p-value of $p < .05$ suggests the sub-groups experienced different effect sizes.</p>
<p>Overall outcomes and sub-groups</p> <ul style="list-style-type: none"> Range and median of effect sizes for each overall outcome group (table) by sub-group calculated for baseline to post-course and baseline to follow-up. 	<p>Test for interactions between sub-group, grouped outcome measures, and time</p> <p>A doubly multivariate (with repeated measures) Analysis of Variance was used to test for interactions between sub-groups, grouped outcomes (in each table), and the changes through observation points in time. This test was used to confirm there were no overall trends in the outcomes for the sub-groups.</p>

References

- Altena, I and Herewini, T (2009) *Incredible Years Marae Based Group*. Auckland: Werry Centre
- Aos, S, Miller, M and Drake, E (2006) *Evidence-based public policy options to reduce future prison construction, criminal justice costs and crime rates*. Olympia: Washington State Institute for Public Policy
- Arnold, P S, O'Leary, S G, Wolff, L S and Acker, M M (1993) The Parenting Scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment* 5 (2), 137–144
- Belsky, J, Vandell, D L, Burchinal, M, Clarke-Stewart, K S, McCartney, K and Owe, M T (2007) Are there Long-Term Effects of Early Child Care? *Child Development*, March/April, Vol 78, No. 2, 681–701
- Berryman, M, Woller, P and Glynn, T (2009) *The Incredible Years: Learning from the experiences of Māori whānau and Māori staff in Special Education*. Wellington: Ministry of Education
- Berryman, M, Woller, P and Glynn, T (2012) *Māori Whānau and Māori Facilitators: A Look at their Experiences in Incredible Years in 2011*. Wellington: Ministry of Education
- Braiker, H and Kelley, H H (1979) Conflict in the development of close relationships. In R L Burgess and T L Huston (Eds) *Social exchange in developing relationships*. New York: Academic
- Brestan, E V and Eyberg, S M (1998) Effective psychosocial treatments of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of Clinical Child Psychology*, 1998 27(2):180–189
- Broughton, J, Fergusson, D M, Rimene, C, Horwood, L J and Sporle, A (2000) *Nga Tini Aho O Te Ao Hou. The many strands of contemporary Māori society: Māori ethnicity and identity in the Christchurch Health and Development Study*. The Ngai Tahu Māori Health Research Unit (NTMHRU)
- Cargo, T (2008) *Māori experiences of delivering the Incredible Years Parenting Programme (Reflections)*. Auckland: Werry Centre
- Chamberlain, P and Reid, J B (1987) Parent observation and report of child symptoms. *Behavioral Assessment*, 9, 97–109
- Church, R J (2012a) Incredible Years Pilot Study: The Repeated Measures Direct Observation Study (unpublished)
- Church, R J (2012b) *Parent-Child Interaction Coding Scheme (P-CICS): Coding manual*. University of Canterbury: School of Educational Studies and Human Development
- Church, R J, Tyler-Merrick, G and Hayward, H (2006) *Identifying antisocial development in New Zealand children. Phase 1: Development of the screening procedure*. Report to the Ministry of Education. University of Canterbury: School of Education
- Cohen, J (1977) *Statistical Power Analysis for the Social Sciences*. New York: Academic Press
- Conduct Problems Prevention Research Group (CPPRG) (1995) *Psychometric Properties of the Social Competence Scale – Teacher and Parent Ratings*. (Fast Track Project Technical Report). University Park, PA: Pennsylvania State University
- Corrigan, A (2002) *Social Competence Scale – Parent Version, Grade 1/Year 2*. (Fast Track Project Technical Report). Available from the Fast Track Project Web site, <http://www.fasttrackproject.org>
- Dadds, M, Maujean, A and Fraser, J A (2003) Parenting and Conduct Problems in Children: Australian Data and Psychometric Properties of the Alabama Parenting Questionnaire. *Australian Psychologist*, Vol 38, No. 3, 238–241

- Dadds, M and Powell, M (1991) The relationship of interparental conflict and global marital adjustment to aggression, anxiety and immaturity in aggressive and non-clinic children. *Journal of Abnormal Child Psychology*, 19, 553–561
- Durie, M (2004) *Race and ethnicity in public policy. Does it work?* The Social Policy, Research and Evaluation Conference, 25 November. Wellington
- Durie, M (2006) *Māori Education 2026*. Post Primary Teachers Association Conference. Wellington
- Eyberg, S M (1980) Eyberg Child Behaviour Inventory. *Journal of Clinical Psychology* 9, 27
- Eyberg, S M (2010) *Abridged Manual for the Dyadic Parent-Child Interaction Coding System*, 3rd edition. Child Study Laboratory, University of Florida
- Eyberg, S M and Ross, A W (1978) Assessment of child behavior problems: The validation of a new inventory. *Journal of Clinical Child & Adolescent Psychology*, 7(2), 113-116
- Fergusson, D M, Horwood, L J and Ridder, E (2005a) Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. *J Child Psychol Psychiatry*, 2005, 46(8):837–849
- Fergusson, D M, Horwood, L J, Ridder, E M and Grant, H (2005b) *Early Start: Evaluation report*. Retrieved 7 May 2008 from <http://www.earlystart.co.nz/pdf/evalreport.pdf>
- Fergusson, D M, Stanley, L and Horwood, L J (2009) Preliminary Data on the Efficacy of the Incredible Years Basic Parent Programme in New Zealand. *Australian and New Zealand Journal of Psychiatry* 43, 76–79
- Fergusson, D M, Boden, J and Hayne, H in P Gluckman (2011) *Improving the Transition*. Wellington
- Feinstein, L and Sabates, R (2006) Predicting adult life outcomes from earlier signals: Identifying those at risk. Centre for Research on the Wider Benefits of Learning, Institute of Education, University of London
- Forehand, R and McMahon, R J (1981) *Helping the noncompliant child: A clinician's guide to parent training*. New York: Guilford Press
- Gardner, F, Burton, J and Klimes, I (2006) Randomised Controlled Trials of a Parenting Intervention in the Voluntary Sector for Reducing Problems in children: Outcomes and Mechanisms of Change. *Journal of Child Psychology and Psychiatry*, 47:11, 1123–1132
- Goodman, R (1997) The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology, Psychiatry, and Allied Disciplines* 38 (5), 581–586
- Goodman, R (2001) Psychometric properties of the Strengths Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337–1345
- Gray, A (2011) *Investing in the early years* from <http://www.nzfamilies.org.nz/web/investing-early-years/index.html>
- Gray, D (2013) *Incredible Years Pilot Study: Technical Appendix*. Ministry of Social Development, Wellington (unpublished)
- Gluckman, P (2011) *Improving the Transition*. Wellington
- Heckman, J J (2006) Skill formation and the Economics of Investing in Disadvantaged Children. *Science* 30 June 2006, Vol 312, No. 5782, 1900–1902
- Hendricks, A K and Balakrishnan, R (2005) *Review of parenting programmes*. Wellington: The Families Commission
- Hess, M R, Kromrey, J D (2003) *EFFECT_CI: A SAS® Macro for Constructing Confidence Intervals Around Standardized Mean Differences*. University of South Florida. Retrieved from <http://analytics.ncsu.edu/sesug/2003/SD04-Hess.pdf>

- Holmes, T H and Rahe, R H (1967) The social readjustment rating scale. *Journal of Psychosomatic Research*, Vol 11(2), 1967, 213–218
- Horwood, L J, Gray, D S and Fergusson, D M (2011) *The Psychometric Properties of the Child Behaviour Rating Scales used in the Incredible Years Pilot Study* (unpublished)
- Hutchings, J, Bywater, T, Daley, D, Gardner, F and Whitaker, C, Jones, K, et al (2007) Parenting intervention in Sure Start services for children at risk of developing conduct disorder: Pragmatic randomised controlled trial. *British Medical Journal*, 334(7595), 1–7
- Jones, K, Daley, J, Hutchings, T, Bywater, T and Eames, C (2007) Efficacy of the Incredible Years Basic parent training programme as an early intervention for children with conduct problems and ADHD. *Child: Care, Health and Development*, 33, (6) 749–756
- Kazdin, A E (2007) *Psychosocial treatments for conduct disorder in children and adolescents*. New York, NY: Oxford University Press. (2007). A guide to treatments that work (3rd edition). 71–104, xxv, 758
- Karoly, L A, Kilburn, M R and Cannon, J S (2005) *Early childhood interventions: Proven results, future promise*. Santa Monica: RAND Corporation. Retrieved 7 May 2008 from http://www.rand.org/pubs/monographs/2005/RAND_MG341.pdf
- Kelleher, C and McGilloway, S (2006) The Incredible Years Basic Parenting Programme: The Clondalkin Partnership Study in Ireland. Retrieved December 2012 from <http://www.incredibleyears.com/library>
- Larsson, B, Fossum, S, Clifford, G, Drugli, M B, Handegård, B H and Mørch, W-T (2009) Treatment of oppositional defiant and conduct problems in young Norwegian children: results of a randomized controlled trial. *European Child and Adolescent Psychiatry*, 18(1):42–52
- Lindsay, G, Strand, S, Cullen, M A, Cullen, S, Band, S, Davis, H, Conlon, G, Barlow, J and Evans, R (2011) *Parenting Early Intervention Programme evaluation: DFE-RR121A*. London: DfE
- Lovibond, S H and Lovibond, P F (1995) *Manual for the Depression Anxiety Stress Scales* (2nd edition). Psychology Foundation, Sydney
- MacFarlane, A (2011) Te Ao Māori perspective on understanding conduct problems, a chapter in Advisory Group on Conduct Problems (2011) *Conduct Problems: Effective Programmes for 8–12 Year Olds*. Wellington: Ministry of Social Development
- Ministry of Education (2011) *Positive Behaviour for Learning Action Plan Update*. Wellington, New Zealand
- Ministry of Social Development (2007) *Interagency plan for Conduct Disorder/Severe Antisocial behaviour 2007–2012*. Wellington: Ministry of Social Development
- Ministry of Social Development (2009a) *Conduct problems: Best practice report*. Wellington: Ministry of Social Development
- Ministry of Social Development (2009b) *Conduct problems: Effective programmes for 3–7 year olds*. Wellington: Ministry of Social Development
- Odgers C L, Caspi, A, Broadbent, J M, Dickson, N, Hancox, R J, Harrington, H (2007) Prediction of differential adult health burden by conduct problem subtypes in males. *Archives of General Psychiatry*, 2007, 64:476–84
- O'Neill, D (2009) A Cost-Benefit Analysis of Early Childhood Intervention: Evidence from a Randomised Evaluation of a Parenting Programme. IZA Discussion Paper No. 4518, October
- Posthumus, J A, Raaijmakers, M a J, Maassen, G H, van Engeland, H and Matthys, W (2011) Sustained effects of Incredible Years as a Preventive Intervention in Preschool Children with Conduct Problems. *J Abnorm Child Psychol*

RAND (2007) *Promising Practices Network: Programmes that Work: Incredible Years*. Retrieved 25 June 2012 from <http://www.promisingpractices.net/program.asp?programid=134>

Rhodes, H (2009). Briefing sheet: Examples of Effective Measuring Tools, Family and Planning Institute

Robinson, E A, Eyberg, S M and Ross, A W (1980) The standardisation of an inventory of child conduct problem behaviours. *Journal of Clinical Child Psychology*, 57, 628–635

Romeo, R, Knapp, M and Scott, S (2006) Economic cost of severe antisocial behaviour in children – who pays it? *British Journal of Psychiatry* (2006) 188, 547–553

Scott, S, Knapp, M, Henderson, J and Maughan, B (2001) Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *British Medical Journal*, 323, 194–203

Scott, S (2007) Conduct disorders in children. *British Medical Journal*, March, 334:646

Shelton, K K, Frick, P J and Wootton, J (1996) Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*, 25, 317–329

Smith, L T (1999) *Decolonizing methodologies: Research and indigenous peoples*. Dunedin: University of Otago Press

Straus, M, Hamby, S, Boney-McCoy, S and Sugarman, D (1996) The revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues*, 1998, 17(3), 283–316

Straus, M A, Hamby, S L, Finkelhor, D, Moore, D W and Runyan, D (1998) Identification of child maltreatment with Parent-Child Conflict Tactics Scales: Development and psychometric data for a national sample of American parents. *Child Abuse and Neglect*, 22(4), 249–270

Webster-Stratton, C (1994) Advancing videotape parent training: A comparison study. *Journal of Consulting and Clinical Psychology*, 62(3), 583–593

Webster-Stratton, C (2007) Affirming diversity: multi-cultural collaboration to deliver the Incredible Years Parent Programs. Retrieved 8 June 2012 from <http://www.incredibleyears.com/library>

Webster-Stratton, C and Hammond, M (1997) Treating children with early-onset conduct problems: a comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology*, 65(1), 93–109. Supplementary analysis retrieved December 2012 from <http://www.incredibleyears.com/library>

Webster-Stratton, C and Reid, M J (2010) The Incredible Years Parents, Teachers and Children Training Series: A Multifaceted Treatment Approach for Young Children with Conduct Disorders, book chapter in J Weisz and A Kazdin (Eds) *Evidence-based Psychotherapies for Children and Adolescents*, 2nd edition. New York: Guilford Publications

Webster-Stratton, C, Reid, M J and Hammond, M (2004) Treating Children with Early-Onset conduct Problems: Intervention Outcomes for Parent, Child, and Teacher Training. *Journal of Clinical Child and Adolescent Psychology*, Vol 33, No. 1, 105–124

Webster-Stratton, C (2012) Presentation given at The Incredible Years Parent Management Training Conference. Wellington, New Zealand

Werry Centre (2008) *Māori Experiences of Delivering the Incredible Years Parenting Programmes (Reflections)*. Auckland: Werry Centre

Werry Centre (2010) *Incredible Years Ngā Tau Miharo Hui*. Auckland: Werry Centre

Werry Centre (2012) *Ngā Tau Miharo o Aotearoa*. Auckland: Werry Centre

World Health Organisation (1993) Composite International Diagnostic Interview (CIDI). World Health Organisation, Geneva, Switzerland

