



# Preparing for Life at Age 9

## Assessing the Continuity or Fade-out of Effects\*

### Technical Report

PFL Evaluation Team  
UCD Geary Institute for Public Policy  
University College Dublin

November 2019

#### Abstract

This report examines the impact of *Preparing for Life (PFL)*, an Irish prenatally commencing early childhood intervention programme which targets disadvantaged communities and focuses on parents as the key mechanism of change, when the children are *nine years old*. Prior evidence on the sustained effect of home visiting and parent-training programmes in middle childhood is inconclusive, with many of the pioneering studies experiencing a dissolution of treatment effects once the programmes end. This report finds little evidence of cognitive fade-out at age 9, with effect sizes of 0.67 of a standard deviation (SD) on general conceptual ability, and significant treatment effects on executive functioning (inhibitory control (0.61SD), attention flexibility (0.66SD), working memory (0.56SD) and standardised school achievement tests of reading (0.74SD) and maths (0.47SD). The significance and magnitude of these effects are similar to those observed at age five. The programme, however, has no impact on absenteeism or the use of school resources and the significant treatment effects found for children's socio-emotional skills and behaviour at age 4 are no longer present. While about 50% of the original sample recruited during pregnancy is retained at the Age 9 Follow-up, the treatment groups are still balanced on all key baseline characteristics. In sum, these findings indicate that *PFL* continues to have a significant impact on important dimensions of children's skills five years after the families finished the programme, yet continued investment may be needed to break long-standing inequalities in health and socio-emotional wellbeing. That said, the sizable cognitive advantages are likely to have positive impacts as the children progress from primary to secondary school, as well as subsequent outcomes in adolescence.

---

\* The evaluation of the *Preparing for Life* programme was funded by the Northside Partnership through the Department of Children and Youth Affairs and The Atlantic Philanthropies. Additional support was provided by Tusla, the Community Foundation of Ireland and the HSE for the Age 9 Follow-up. We would like to thank all those who supported this research, especially the participating families and community organizations, the *PFL* intervention staff, and the programme's Expert Advisory Committee. Thanks also to the Scientific Advisory Committee, including Sylvana Côté, Colm Harmon, James Heckman, Cecily Kelleher, Sharon Ramey, Craig Ramey, and Richard Tremblay, for their guidance and advice, and to the Early Childhood Research Team at the UCD Geary Institute for Public Policy who significantly contributed to this study. The trial was registered with controlled-trials.com (ISRCTN04631728) and the AEA RCT Registry (AEARCTR-0000066). All study procedures were approved by the UCD Human Research Ethics Committee, the Rotunda Hospital's ethics committee, and the National Maternity Hospital's ethics committee. E-mail: [orla.doyle@ucd.ie](mailto:orla.doyle@ucd.ie)

## Section 1 Background

### 1.1 Introduction

This report examines the impact of *Preparing for Life (PFL)*, an Irish prenatally commencing early childhood intervention programme which targets disadvantaged communities and focuses on parents as the key mechanism of change, when the children are *nine years old*. There is a growing evidence base that circumstances early in life are critical for developing the skills required to lead a successful life. Children exposed to adverse prenatal and postnatal environments typically experience poorer health, education, and labour market outcomes in the long run (Cunha *et al.* 2006; Heckman 2006; Almond, Currie, Duque 2017). Intervening early in life to eradicate or compensate for these adverse environments through early childhood intervention programmes is becoming an increasingly accepted strategy (see Council of Economic Advisors 2014; OECD 2016). Such investments are considered efficient from both a biological and an economic perspective (Doyle *et al.* 2009). Physiologically, there is evidence of greater brain plasticity and neurogenesis in the early years, particularly between pregnancy and age three (Thompson and Nelson 2001; Knudsen *et al.* 2006). Such investments may be also economically efficient, as by investing early the returns from the improved skill set can be reaped over a longer period (Heckman and Kautz 2014).

Previous reports of the *PFL* programme have identified some important effects at earlier ages, primarily using parent report measures of children's outcomes.<sup>1</sup> The final report based on data collected at the end of the *PFL* programme, found that the programme had a

---

<sup>1</sup> Doyle (2013) describes the design of the *PFL* evaluation. Doyle *et al.* (2014) examine the impact of the *PFL* programme on birth outcomes utilising hospital data and identify a significant treatment effect regarding a reduction in caesarean sections, yet no impact on neonatal outcomes. Doyle *et al.* (2017a) examine the impact of the programme on parent reported cognitive and non-cognitive skills at 6, 12, and 18 months, and find no evidence of effects, yet there are significant improvements in the quality of the home environment. Doyle *et al.* (2015) examine the impact of the programme on child health measured at 6, 12, 18, 24, and 36 months and identify a number of significant treatment effects at 24 months in terms of reducing the incidence of asthma, chest infections, and health problems. O'Sullivan, Fitzpatrick and Doyle (2017) examine the impact of the programme on dietary intake at 12, 18, 24, and 36 months, and its mediating effect on cognitive development at 24 and 36 months; and find evidence of improved nutrition at 24 months in terms of increased protein intake. Doyle *et al.* (2017b) examine the impact of the programme on maternal wellbeing using daily data collected over a 24-hour period using the Day Reconstruction Method, finding little evidence of effects on maternal wellbeing. Cote *et al.* (2018) investigate whether the impact of the programme varied according to children's developmental trajectories and find a positive impact on trajectories of cognitive development and number of health clinic visits for all children, whereas positive impacts on externalizing behaviour problems are restricted to children with the most severe problems. Finally, Doyle (2019) examines the impact of the programme on children's cognitive and socio-emotional skills at the end of the programme and finds significant effects on all dimensions of children's skills, with large effects for cognitive ability.

large and substantive impact on children's cognitive, social, and behavioural development, as well as their health (Doyle and the PFL Evaluation Team, 2016). The programme raised general conceptual ability, which is a proxy for IQ, by 0.77 of a standard deviation, indicating the malleability of IQ in the early years. Gains were found across all dimensions of cognitive skill including spatial ability, pictorial reasoning, and language ability. The programme significantly reduced the proportion of children scoring below average and increased the proportion of children scoring above average, thus impacting the entire distribution of cognitive skills. These results, based on direct assessment, were supported by significant treatment effects found for parent-reported scores eliciting children's ability from age two onwards. Although weaker, the programme also impacted several dimensions of children's non-cognitive skills including externalising problems such as aggressive behaviour, and prosocial behaviour such as helping other children. In particular, the programme reduced the proportion of children scoring in the clinical range for behavioural problems by 15 percentage points. The programme also had an impact on child health. It reduced the amount of hospital services the children used and improved how families used these services. It also had an impact on weight status – 24% of children in the high treatment group were classified as being overweight or obese at age four, compared to 41% in the low treatment group. The size of these treatment effects exceeded current meta-analytic estimations based on the home visiting literature (e.g. Sweet and Appelbaum 2004; Gomby 2005; Filene *et al.* 2013).

The *PFL* programme effects may be attributed to both its length and intensity. By delivering a five-year intervention, i.e. the first 2,000 days,<sup>2</sup> the programme provided early *and* sustained investment in families during a critical stage of development. This is important as the technology of skill formation, proposed by Cunha and Heckman (2007), shows that children's early skills facilitate the development of more advanced skills through a process of self-productivity, and this in turn makes investment throughout the lifecycle more productive through a process of dynamic complementarity (Cunha, Heckman, and Schennach 2010; Heckman and Mosso 2014). Although there is a genetic basis for the development of skills (Nisbett *et al.* 2012), they can be modified and enhanced by environmental conditions (Weaver *et al.* 2004).

The traditional human capital model proposes that skills are determined by inputs of parental time and income (Becker 1965; Michael and Becker 1973), and that inequalities in

---

<sup>2</sup> Participants joined the program during their 21<sup>st</sup> week of pregnancy on average, and finished when their children started primary school when they were 4 years, 9 months old on average, thus ~1,855 days is the precise figure.

skills arise from inequalities in the availability of these resources. This contributes to the large and well-documented socioeconomic gap in children's cognitive and socio-emotional skills that can be observed as early as 18 months of age (Cunha and Heckman 2007; Fernald, Marchman, and Weisleder 2013). These inequalities have been partly explained by poverty, credit constraints (e.g. Carneiro and Heckman 2003), and parental time investments (e.g. Del Boca, Flinn, and Wiswall 2014; Del Bono *et al.* 2016), however such factors may also influence and/or serve as proxies for the child's environment. Indeed, empirical research has identified the quality of the home environment (Todd and Wolpin 2007), parenting skills (Dooley and Stewart 2007; Fiorini and Keane 2014), and parental stimulation (Miller *et al.* 2014) as important predictors of children's skills.

In particular, families from disadvantaged backgrounds face financial constraints which may limit their ability to invest in their children; however they may also be constrained in their capacity to parent. Evidence suggests that parents from low socioeconomic status backgrounds often engage in poorer parenting styles and behaviours (Cunha, Elo, and Culhane 2013) such as permissive or harsh parenting (Bradley and Corwyn 2002), and provide less stimulating materials and experiences to their children (Bradley *et al.* 1989). This partly may be attributed to a knowledge gap regarding appropriate parenting practices and techniques for optimising child development. Specifically, Cunha *et al.* (2013) identify a lack of parenting knowledge and differing beliefs about the importance of parenting among low socioeconomic status parents. There is also evidence of less pre-academic stimulation, such as reading to children and helping them to recognise letters, in disadvantaged homes (Miller *et al.* 2014). Thus, by increasing parenting knowledge and encouraging parental stimulation in developmentally appropriate activities the *PFL* programme aimed to counteract the adverse effects of disadvantage on children's skills.

## **1.2 Description of report**

This report examines the impact of the *PFL* programme at nine years of age on child, school, and parent outcomes. The child outcomes include the children's cognitive development, socio-emotional and behavioural development, and health based on parent report, child report, and direct assessment. The school outcomes include administrative records of test scores, absenteeism, and educational resources use (e.g. special needs assistants). The parent outcomes include parenting behaviour and the family environment using parent reports.

The remainder of the report is structured as follows. Section 2 reviews the literature assessing the impact of home visiting programmes and group-based parenting programmes on short and medium term outcomes. Section 3 describes the design of the original study and the Age 9 Follow-up. Section 4 outlines the statistical methods that are used to estimate the results. Section 5 presents the main results and robustness tests. Finally, Section 6 concludes.

## **Section 2 Literature**

### **2.1 Evidence base for home visiting and group-based parenting programmes**

There is a growing literature on the effectiveness of parent-focused interventions. A recent meta-analysis of 28 individual meta-analyses on the effectiveness of parenting programmes summarising over 500 studies, finds that such interventions are, on average, effective at reducing externalising problems, with an effect size (Cohen's  $d$ ) of 0.46 (Mingebach *et al.* 2018). Another meta-analysis of 37 studies focusing on internalising problems finds a smaller, although significant, effect ( $d=0.12$ ) (Yap *et al.* 2016). Regarding cognitive skills, an older meta-analysis of 260 family support programmes also finds significant effects ( $d=0.29$ ), yet concludes that the majority of early childhood intervention programmes have very small effects (Layzer *et al.* 2001). Another meta-analysis focusing on interventions with parents of infants finds no effects on externalising ( $d=0.16$ ) or internalising ( $d=0.16$ ) problems or cognitive skills ( $d=0.13$ ), based on 16 studies (Rayce *et al.* 2017). The majority of these meta-analyses are based on studies conducted in developed countries, yet one meta-analysis of 21 parenting interventions operating in lower and middle income countries, focusing on parental stimulation specifically, finds significant effects on early cognitive ( $d=0.42$ ) and language ( $d=0.47$ ) skills (Aboud and Yousafzai 2015). A similar meta-analysis of 13 studies in low and middle income countries finds significant effects on non-cognitive skills ( $d=0.35$ ) (Britto *et al.* 2015). Meta-analyses on the impact of parenting programmes on health outcomes are less numerous and the results are inconsistent (Avellar and Supplee 2013; Filene *et al.* 2013; Peacock *et al.* 2013).

The majority of these meta-analyses include studies of home visiting programmes, group parent training programmes, and centre-based preschool programmes, or some combination. The main *PFL* support, home visiting programmes, have become increasingly popular, yet evidence on the effectiveness of these programmes is mixed. A meta-analysis of 60 home visiting programmes by Sweet and Appelbaum (2004) finds an average effect size

(*d*) of 0.18 for cognitive skills and 0.10 for socio-emotional skills. While Miller, Maguire, and Macdonald (2011), based on seven studies, find average effect sizes of 0.30 for cognitive skills. Finally, Filene *et al.* (2013), based on 51 studies, find an average effect size of 0.25 for cognitive skills and a non-significant effect size of 0.11 for physical health outcomes. Thus, the effects are typically modest in size and not consistent across programmes (Gomby 2005; Peacock *et al.* 2013; Avellar *et al.* 2016).

The secondary *PFL* support is group-based parent training. Such programmes typically involve weekly group sessions facilitated by trained practitioners in standardised interventions such as Incredible Years and Triple P. The aim of these programmes is to improve parenting skills using behavioural and/or cognitive therapy approaches, with a focus on reducing child behavioural problems. Meta-analyses of such group-based programmes have identified some evidence of effectiveness. For example, a review of 24 studies focusing on parents of children under the age of three finds significant reductions in emotional and behavioural problems ( $d=0.81$ ), externalising problems ( $d=0.23$ ), and improvements in social skills ( $d=3.59$ ) (Barlow *et al.* 2016). Another meta-analysis of 13 studies focusing on parents of three to 12 year olds finds significant reductions in child conduct problems ( $d=0.44-0.53$ ) (Furlong *et al.* 2012). There are no specific meta-analyses examining the impact of group parent training on cognitive skills, as it is rarely the focus of these interventions, yet three studies included in the Furlong *et al.* (2012) review measure child educational and cognitive outcomes, and find little evidence of effect ( $d=0.07$ ). Similarly, none of the meta-analyses of group-based parenting programmes report on health outcomes, however, an individual study of Triple P delivered to parents of children with asthma and/or eczema concluded that it led to a better family quality of life and reduced symptom severity (Morawska *et al.* 2016).

## **2.2 Evidence base for sustained impact of home visiting and group-based parenting programmes**

Empirical research regarding the sustained effects of home visiting programmes and group parent training has yielded mixed results. Systematic reviews report that home visiting programmes have effects in multiple domains such as child health, child development, behavioural problems, family economic self-sufficiency, and positive parenting practices (Minkowitz, O'Neill, and Duggan 2016; Peacock *et al.* 2013). However, overall these effects tend to be short term and programmes can be limited in improving the lives of children from disadvantaged families (Burger 2010; Peacock *et al.* 2013). In terms of early childhood intervention programmes more generally, some US based studies have demonstrated a fade-

out of cognitive benefits of the programmes starting later in childhood (Heckman *et al.* 2010; Heckman *et al.* 2013). For example, the Perry Preschool Program found that the large and significant impact on cognitive development of 0.75 of a standard deviation (SD) identified at the end of the programme had fallen to 0.08SDs by age 8. However, other studies do not report such cognitive fade-outs (e.g. Gertler *et al.* 2014). For example, the Jamaica study, which is based on weekly home visits for two years, found maintenance of effects at the 11, 17, and 22 year follow-ups (Grantham-McGregor and Smith 2016). In addition, even among the programmes which do experience cognitive fade-out, they still observe improvements in other outcomes later in life such as involvement in crime and being in receipt of social welfare (Campbell *et al.* 2014; Heckman *et al.* 2017).

A study by Baily *et al.* (2017) examine the persistence or fadeout of 67 high-quality early intervention programmes and finds a general pattern of declining effects sizes, with end of programme effect sizes averaging 0.23SDs, which then fall to 0.10SDs by the end year one, and 0.05SDs up to one to two years after treatment has ended. A meta-analysis focusing on programmes targetting of early phonological awareness specifically, conducted by Bus and van IJzendoorn (1999), found large initial impacts of children's reading skills (0.44SD) which tend to fade to 0.16SD at the 18 month assessment on average. While a number of mathematic interventions also find that large initial gains tend to fade over time (Clements *et al.* 2013; Smith *et al.* 2013). Thus the evidence base for both more general ECI programmes as well as home visiting programmes specifically, tends to exhibit a pattern of fade out over time.

Of the sparse research that has evaluated the sustained impact of home visiting programmes throughout childhood, the results are inconsistent. Tables 1a-d summarise the results of home visiting programmes with follow-ups between the ages of five and 12. Table 1a shows that of the four studies assessing the impact of home visiting programmes on children's cognitive development, only one has a significant impact. In particular, Bierman *et al.* (2017) found that children who participated in Early Head Start (home based option) had improved cognitive ability at ages seven to nine. Table 1b shows that of the five studies assessing the impact on children's achievement tests, three identified a significant impact. Bierman *et al.* (2017) also found an impact on children's reading and language skills at ages seven to nine, and two studies of the Healthy Families America programme also identified significant effects on the percentage of children in a gifted programme, receiving special education, and excelling academically in all three behaviours that promote learning in school at ages six to seven (DuMont *et al.*, 2010; Kirkland and Mitchell-Herzfield 2012). Table 1c

shows that of the four studies assessing the impact on children's health, only one found a significant impact. In particular, Kitzman *et al.* (2010) found that children who participated in the Nurse Family Partnership programme had a reduced incidence of substance use at age 12. Note that none of the studies observed any sustained impacts in relation to children's physical and mental health (Dumont *et al.* 2010; Kitzman *et al.* 2010; Minkowitz *et al.* 2007; Olds *et al.* 2004; Olds *et al.* 2007). Table 1d shows that of the ten studies assessing the impact of home visiting programmes on children's socio-emotional skills and behaviour, only three identify significant impacts. In particular, two studies of the Early Head Start programme found effects on children's behaviours, perceived competence, and approaches to learning at ages five and seven to nine (Bierman *et al.* 2017; Chazan-Cohen, Raikes, and Vogel 2013). In addition, the Nurse Family Partnership programme found a reduction in internalising disorders at age 12 (Kitzman *et al.* 2010). In general, however, the majority of evaluations found no significant effects for sustained impact of home visiting programmes on children's social, cognitive, behavioural, physical, or academic development (e.g. Kitzman *et al.* 2010; Minkowitz *et al.* 2012; Olds *et al.* 2014; Sidora-Arcoleo *et al.* 2010; Sitnick *et al.* 2015; Smith *et al.* 2014).

There is less research available on the long-term impact of group-based parenting programmes, and meta-analytic evidence of sustained impact is not consistent. For example, De Graaf *et al.*'s (2008) meta-analysis assessing the effectiveness of Triple P (based on 14 studies) indicates that post-intervention reductions in disruptive behaviours were maintained over time. However, Wilson *et al.*'s (2012) review of 23 Triple P studies concluded that there was less convincing evidence of long-term benefits.

Overall, the evidence for the sustained effectiveness of home visiting and parent – training programmes in middle childhood is inconclusive, with much of the literature finding little evidence of a sustained effect. In addition, much of the evidence base stems from US studies. One may expect lower SES inequalities in children's skills in Europe where many countries are characterised by universal health insurance, generous welfare payments, and a social safety net which protects the most vulnerable in society. Yet inequalities in children's skills continue to persist in Europe despite these arguably more redistributive policies.<sup>3</sup> The existence of such inequalities suggests that family economic circumstances alone may not be the primary driver alone. Thus, a human capital model, which moves beyond income and

---

<sup>3</sup> Martins and Veiga (2010) find that socio-economic status represents between 14.9% and 34.6% of overall inequality in mathematics scores in the EU using PISA (Programme for International Student Assessment) data, with Germany scoring the highest and Sweden the lowest. In Ireland, the figure is 25%.



time investments as the main determinants of children's skills, may provide a more informative model for testing the impact of early childhood interventions in countries with extensive welfare provision such as Ireland.

With this in mind, this report explores the role of intensive and continued investment in parenting from pregnancy until entry into formal schooling within a disadvantaged community in Dublin, Ireland. Theoretically, if the *in utero* and infancy periods are critical for optimising brain development, and parenting and the quality of the home environment is strongly implicated in the development of children's skills, then intervening early and focusing on parents may generate larger effects than centre-based pre-school programmes on which much of the early childhood intervention literature is based.

**Table 1a** *Evaluations of Cognitive Outcomes for Home Visiting Programmes from Ages 5-12*

Outcome	Author	Sample Size	Programme	Measures	Significant Finding	Effect	Age (years)	% of original sample retained
<b>Cognitive tests</b>	Bierman <i>et al.</i> (2017)	556 children	Early Head Start Home Based Option + REDI-P	Woodcock Johnson tests of Achievement (Letter-Word Identification Scale), School Readiness Questionnaire, Learning Behaviours Scale	Woodcock Johnson tests of Achievement	Favourable	7-9	87%
	Chazan-Cohen, Raikes and Vogel (2013)	927 families	Early Head Start Home Based Option	English receptive vocabulary, Woodcock-Johnson revised test, Leiter R sustained attention test, speech problems survey	None	None	5	Original N not reported. Retention between ages 2-5: 96% parent interviews, 92% child assessments
	Kitzman <i>et al.</i> (2010)	635 children	Nurse Family Partnership	Leiter-R sustained attention test	None	None	12	80% parent interviews, 76% child interviews, 85% school records
	Olds <i>et al.</i> (2014)	411 children	Nurse Family Partnership	Conners continuous performance test (attention dysfunction)	None	None	6-9	81% at age 6 78% at age 9

**Table 1b** *Evaluations of Achievement Test Outcomes for Home Visiting Programmes from Ages 5-12*

Outcome	Author	Sample Size	Programme	Measures	Significant Finding	Effect	Age (years)	% of original sample retained
<b>Achievement tests</b>	Bierman <i>et al.</i> (2017)	556 children	Early Head Start Home Based Option + REDI-P	Test of Word Rereading Efficiency, Academic Competence Evaluation Scales (ACES; reading/language skills, mathematics), Academic Performance Rating Scale	ACES (reading/language skills)	Favourable	7-9	87%
	DuMont <i>et al.</i> (2010)	897 mothers	Healthy Families America	Partaking in a gifted programme, receiving remedial services, receiving special education, repeating a grade	Participating in a gifted programme, receiving special education	Favourable	7	80% of baseline sample
	Kirkland and Mitchell-Herzfield (2012)	577 mother and child pairs	Healthy Families America	Doing poorly academically (below grade level in reading, math or positive behaviours that promote learning), excelling academically (reading and math, behaviours that promote learning), retained in 1 <sup>st</sup> grade	Excelling academically with all 3 behaviours that promote learning, retained in 1 <sup>st</sup> grade	Favourable	6-7	49% academic reports, 68% child interviews, 80% parent surveys
	Kitzman <i>et al.</i> (2010)	635 children	Nurse Family Partnership	Placement in special education, ever retained in a grade, GPA, Group achievement test scores, Peabody Individual Achievement Tests	None	None	12	80% parent interviews, 76% child interviews, 85% school records
	Sidora-Arcoleo <i>et al.</i> (2010)	721 mother and child dyads	Nurse Family Partnership	Peabody Picture Vocabulary Test-Revised	None	None	6-12 years	Not reported

**Table 1c** *Evaluations of Child Health Outcomes for Home Visiting Programmes from Ages 5-12*

Outcome	Author	Sample Size	Programme	Measures	Significant Finding	Effect	Age (years)	% of original sample retained
<b>Health</b>	Kitzman <i>et al.</i> (2010)	578 children	Nurse Family Partnership	Days of substance use in the last 30 days (interview)	Incidence of substance use, used cigarettes, alcohol or marijuana in the last 30 days.	Favourable	12	80% parent interviews, 76% child interviews, 85% school records
	Minkowitz <i>et al.</i> (2007)	1308 children	Healthy Steps	Hospitalised in previous year, used emergency department in past year for injury	None	None	5.5	57%
	Olds <i>et al.</i> (2004)	641 mothers	Nurse Family Partnership	Subsequent admissions of child to ICU/Special Care	None	None	6	54% child assessment, 56% parent interviews
	Olds <i>et al.</i> (2007)	627 mothers	Nurse Family Partnership	Infant/childhood deaths	None	None	9	55% parent interviews, 49% teacher reports, 53% school records, 50% achievement tests

**Table 1d** *Evaluations of Socio-emotional Outcomes for Home Visiting Programmes from Ages 5-12*

Outcome	Author	Sample Size	Programme	Measures	Significant Finding	Effect	Age (years)	% of original sample retained
<b>Socio-emotional development</b>	Bierman <i>et al.</i> (2017)	556 children	Early Head Start Home Based Option + REDI-P	Social Competence Scale, Student Teacher Relationship Scale, Child Behaviour Scale (excluded by peers), Perceived Competence Scale for Children, Loneliness Scale, Friendship Questionnaire	Perceived Competence Scale for Children, Child Behaviour Scale	Favourable	7-9	87%
	Chazan-Cohen, Raikes and Vogel (2013)	927 families	Early Head Start Home Based Option	Child Behaviour Checklist (CBCL), Family and Child Experiences Survey (FACES), negativity towards parents during play	FACES (positive approaches to learning), FACES (social behaviour problems)	Favourable	5	Original N not reported. Retention between ages 2-5: 96% parent interviews 92% child assessments
	Dumont <i>et al.</i> (2010)	897 mothers	Healthy Families America	CBCL (attention problems, rule breaking and aggressive behaviours, social problems, and the anxious-depressed and withdrawn-depressed syndrome)	None	None	7	80% of baseline sample
	Kitzman <i>et al.</i> (2010)	635 children	Nurse Family Partnership	Child conduct	None	None	12	80% parent interviews, 76% child interviews, 85% school records
	Kitzman <i>et al.</i> , (2010)	594 mothers, 578 children	Nurse Family Partnership	Ever arrested, externalising disorders and internalising disorders, total problems	Internalising disorders	Favourable	12	80% parent interviews, 76% child interviews, 85% school records
	Minkowitz <i>et al.</i> (2007)	1308 children	Healthy Steps	CBCL (internalising, externalising, total behaviour problems)	None	None	5.5	57%
	Olds <i>et al.</i> (2014)	411 children	Nurse Family Partnership	CBCL (internalising, externalising, total behaviour problems)	None	None	6-9	81% at age 6, 78% at age 9

	Sidora-Arcoleo <i>et al.</i> (2010)	721 mother and child dyads	Nurse Family Partnership	CBCL (physical aggression items)	None	None	6-12	Not reported
	Sitnick <i>et al.</i> (2015)	614 families	Family Check Up for Children	CBCL (oppositional-aggressive items)	None	None	5	85%
	Smith <i>et al.</i> (2014)	612 children	Family Check Up for Children	CBCL (oppositional-aggressive items), oppositional behaviour in the classroom	None	None	7-8	62%

## Section 3 Study Description

### 3.1 Setting and programme design

In an effort to break the intergenerational cycle of disadvantage, *PFL* was developed as part of the Irish Government's and The Atlantic Philanthropies' Prevention and Early Intervention Programme (Office of the Minister for Children and Youth Affairs 2008). The programme was developed by 28 local agencies and community groups who collaborated to design an evidence-based intervention tailored to meet the needs of the local community. The study took place between 2008 and 2015 in a highly disadvantaged Dublin community. Census data collected prior to programme delivery demonstrates high rates of unemployment (12% vs national average of 3.5%) and public housing uptake (42% vs national average of 7.2%), and low levels of education (7% completed college degree vs national average of 19.4%) (Census 2006). Survey data also shows that children in the community consistently score below the norm in cognitive and language development, communication and general knowledge, physical health and wellbeing, social competence, and emotional maturity (Doyle *et al.* 2012). The *PFL* programme was developed to reduce these socioeconomic inequalities in children's skills by working directly with parents to improve their knowledge of child development and parenting.

### 3.2 Initial recruitment and randomisation

Recruitment into the *PFL* programme took place between the 29<sup>th</sup> of January 2008 and the 4<sup>th</sup> of August 2010 through two maternity hospitals and/or self-referral using a community-based marketing campaign. The inclusion criteria included all pregnant women residing in the designated *PFL* catchment area during this period, regardless of their social or family circumstances. Based on estimates of a two to five point difference on cognitive development scores (i.e., average standardised effect size of 0.18) from a meta-analysis of home visiting programmes (Sweet and Appelbaum 2004), a sample size of approximately 117 in each group was required to power the study.

In total, 233 participants were recruited by the *PFL* recruitment officers. This represents a recruitment rate of 52% based on the number of live births during the recruitment period. For the remaining 48%, initial contact was made with 26%, but they could not be subsequently contacted or they refused to join the programme, and a further 22% never had any contact with the recruiters. Of those who joined the programme, an

unconditional probability randomisation procedure, with no stratification, assigned 115 to a high treatment group and 118 to a low treatment group. During recruitment, the participants initiated their own randomisation by touching the screen of a tablet laptop which generated their assignment condition.<sup>4</sup> This process produced an email which was automatically sent to the programme manager and the principal investigator listing the participant's treatment condition and identification code. Any attempts to compromise randomisation by reassigning participants would trigger an additional email highlighting any intentional subversion.

Baseline data from 205 participants (representing 90% of the high treatment group and 86% of the low treatment group) was collected after randomisation yet prior to treatment delivery.<sup>5</sup> The baseline variables include 117 measures of socio-demographics, physical and mental health, IQ, parenting attitudes, and self-control, among others. To assess the effectiveness of the randomisation procedure, the baseline characteristics of the high and low treatment groups were compared using separate permutation tests. At the 10% significance level, the two groups differed on 7.7% (9/117) of measures, which is consistent with pure chance and indicates the success of the randomisation process (see Doyle and *PFL* Evaluation Team 2010). Regarding the few observed statistically significant differences, there were no systematic patterns in the data.<sup>6</sup>

### 3.3 Treatment

Figure 1 below describes the supports provided to the high and low treatment groups. The high treatment consisted of three primary components - a five year home visiting programme, a baby massage course, and the Triple P Positive Parenting Programme. The treatments are built upon the theories of human attachment (Bowlby 1969), socio-ecological development (Bronfenbrenner 1979), and social-learning (Bandura 1977). The home visiting programme aimed to promote children's health and development by building a strong mentor-parent relationship and focusing on the identification of developmental milestones,

---

<sup>4</sup> Actively involving participants in the randomisation procedure helped to ensure that they trusted that the procedure was truly random and that a judgement on their parenting ability was not being made. Data capturing participants' response to treatment assignment shows that 98% were 'happy' with their group assignment.

<sup>5</sup> Of the 233 randomly assigned participants, two (high=1; low=1) miscarried, 19 (high=6; low=13) withdrew from the programme before the baseline assessment, and seven (high=4; low=3) did not participate in the baseline but participated in subsequent waves. An analysis of a subset (n = 12) of this group on whom recruitment data but no baseline data are available, implies they do not differ on age, education, employment, and financial status from those who did complete a baseline assessment, however the limited sample size should be noted.

<sup>6</sup> High treatment mothers were more likely to be at risk of insecure attachment, reported lower levels of parenting self-efficacy, were more likely to have a physical health condition, were less considerate of future consequences, and were less likely to intend to use childcare, however they also demonstrated greater knowledge of infant development and reported using more community services than the low treatment group.



appropriate parenting practices, and encouraging enhanced stimulation. The visits started in the prenatal period and continued until school entry. Twice monthly home visits of approximately one hour were prescribed with mentors from different professional backgrounds including education, social care, and youth studies. The mentors were hired to deliver the programme on a full-time basis and they received extensive training prior to treatment delivery. Mentor supervision took place on a monthly basis to ensure fidelity to the programme model<sup>7</sup>, and families were allocated the same mentor over the course of the intervention where possible.

Each home visit was structured around *PFL*-developed 'Tip Sheets' which included information on pregnancy, parenting, health, and development. The 210 Tip Sheets were developed by the *PFL* implementation team based on pre-existing and publicly available information from local organisations such as the Health Service Executive, the Department of Health and Children, and Barnardos Children's Charity. The mentors could choose when to deliver the Tip Sheets based on the age of the child and the needs of the family, yet the full set of Tip Sheets must have been delivered by the end of the programme. The mentors delivered the intervention using techniques such as role modelling, coaching, discussion, encouragement, and feedback, as well as directly interacting with the *PFL* child. Each home visit began with an update on the family's situation and a discussion of whether the goals agreed at the previous visit were achieved. The mentor would then guide the parent through the Tip Sheet(s) selected for that visit and following this, new goals would be agreed. The Tip Sheets typically targeted multiple aspects of development. An analysis of Tip Sheet content found that 12% (n=22) encouraged the development of cognitive skills; 14% (n=25) focused on language development; 16% (n=30) encouraged children's development of positive approaches to learning; 33% (n=60) dealt with social and emotional development; and finally, the majority of Tip Sheets addressed physical wellbeing and motor development (59%, n=105).

Participants in the high treatment group were also encouraged to take part in a baby massage course in the first year, which consisted of five two-hour individual or group sessions delivered by the mentors. The purpose of these classes was to equip parents with

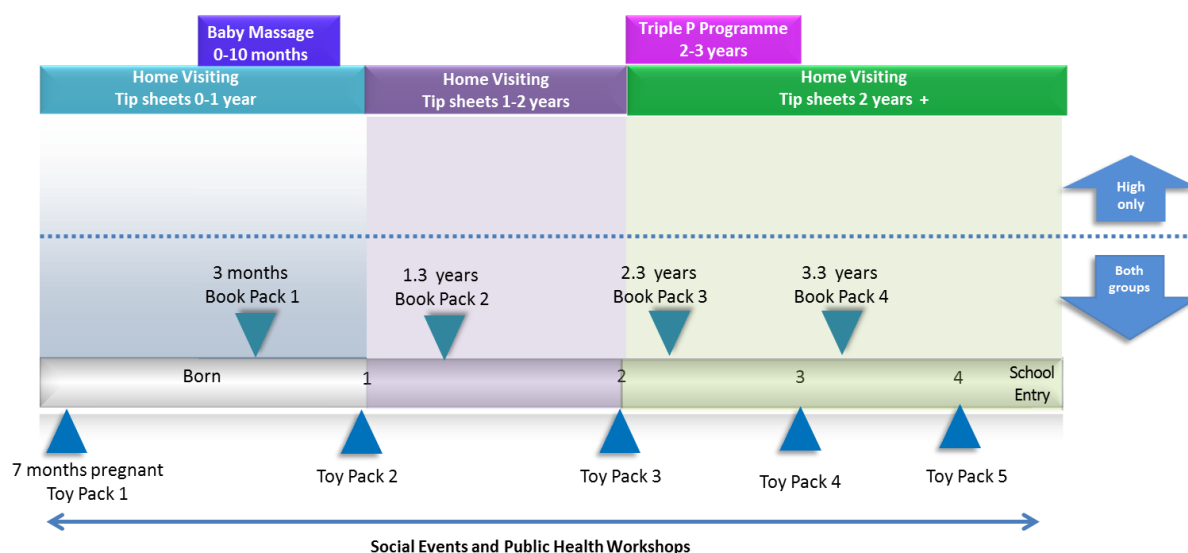
---

<sup>7</sup> The training included an intensive two-day workshop on the *PFL* programme, with a focus on the programme manual, and included topics such as the evidence base for mentoring programmes, relationships and activities, outcomes and evaluation, policy and practice alignment, and the *PFL* logic model. They also received 21 other relevant courses conducted over a six-month period including child protection, attachment theory, and team building. Mentor supervision was based on the model commonly used by social workers in Ireland and was provided for two hours per month. Key areas addressed included participant work, teamwork, support, administration, and training/development.

baby massage skills and to emphasise the importance of early reciprocal interactions and communication between parents and infants. A systematic review of 34 RCTs of infant massage found limited effects on child outcomes, although the authors note the low quality of many of the included studies (Bennett, Underdown, and Barlow 2013). Baby massage was included as part of the *PFL* treatment as an enjoyable activity which encouraged early engagement with the programme.

When the *PFL* children were between two and three years old, the high treatment group was invited to participate in the *Triple P Positive Parenting Programme* (Sanders, Markie-Dadds, and Turner 2003) which was delivered by the mentors. The goal of *Triple P* is to encourage positive, effective parenting practices in order to prevent problems in children's development. The programme is based on five principles including providing a safe, engaging environment, the home as a positive place to learn, setting of rules and boundaries, realistic expectations of children, and parental self-care (Sanders 2012). Meta-analysis of the impact of *Triple P* has identified improved parenting practices and child social, emotional, and behavioural outcomes (Sanders *et al.* 2014). *Triple P* consists of five treatment levels of increasing intensity including a media campaign and communication strategy, a positive parenting seminar series, single session discussion groups, intensive small group and individual programs, and intensive family intervention. The high treatment participants were specifically encouraged to take part in the small group programme which consisted of five two-hour group discussion sessions and three phone calls. The mentors also used the *Triple P* principles and techniques when delivering the home visits to ensure consistent messaging across the programme components.

In addition to the standard services available to pregnant women and young children, both the high and low treatment groups received a supply of developmental toys annually (to the value of ~€100 per year) including a baby gym, safety items, and developmental toys. They also received four book packs containing six to eight developmentally appropriate books. The groups were also encouraged to attend community-based public health workshops on stress management and healthy eating, as well as social events such as coffee mornings and Christmas parties organized by the *PFL* staff. Programme newsletters and birthday cards were sent annually to each family, in addition to two framed professional photographs of the child. The low treatment group also had access to a *PFL* support worker who could help them avail of community services if needed, and this function was provided by the mentors for the high treatment group. Note that the low treatment group did not receive the home visiting programme, Tip Sheets, baby massage classes, or the *Triple P* programme.



**Figure 1** Timing of PFL treatments

### 3.4 Dosage

The average number of home visits delivered to the high treatment group between programme entry and programme end was 49.7 (SD = 38.1, range 0 - 145), which equates to just less than one visit per month. This represents 38% of prescribed visits which is somewhat less than the 50% that is typically found in shorter home visiting programmes (Gomby *et al.* 1999). The average duration of each visit was just under one hour, and on average participants received 50.6 hours of the home visiting treatment. There was, however, large variability in treatment intensity with 17% of families not participating in any home visits and 16% receiving over 90 visits. Regarding the other treatments, 43% of all randomised high treatment families participated in some form of the *Triple P* programme. Of those, the majority took part in the small group programme (86%), with smaller proportions participating in the discussion groups (42%) and the intensive individual programme (12%). The baby massage course was attended by 62% of all randomised high treatment participants.

In terms of the common supports available to both groups, 81% of the high treatment group and 77% of the low treatment group received at least one developmental pack, and 68 and 52% respectively attended a *PFL* social event. Finally, 77% of the low treatment group contacted the *PFL* support worker at least once during the course of the programme.

### 3.5 Design of the Age 9 Follow-up

#### 3.5.1 Recruitment

The Age 9 Follow-up study sought to include as many of the original *PFL* participants as possible. As such, all families that were recruited and randomised in the original *PFL* study between 2008 and 2010 were eligible to take part. However, participants who had officially dropped out or left the study due to death or miscarriage were not contacted. Recruitment was conducted through two phases: community event recruitment followed by targeted recruitment.

##### *i) Community event recruitment*

For the first phase of recruitment, trained members of the UCD research team attended a community event organised by the *PFL* implementation team in February 2019 for all families that took part in Phase I of the programme. Attendees were asked if they were interested in hearing about the Age 9 Follow-up study and if they were happy to speak to a researcher from UCD about it. A total of 59 families were recruited at the community event.

##### *i) Targeted recruitment*

As a second phase of recruitment, eligible participants who did not attend the community event and who had previously agreed to future contact from the UCD research team were contacted directly and were invited to participate in the Age 9 Follow-up study. In total, 66 participants were recruited through targeted recruitment.

In both phases of recruitment, eligible participants were provided with information about the follow-up study. Researchers explained the procedures, articulating all relevant study information, consent details, and participant rights. Parents were provided with an information sheet detailing the study, as well as an age appropriate information booklet for their child.

#### 3.5.2 Data collection procedure

Unlike previous data collection waves which took place over a two and a half year period in order to capture children who were the same age at the time of the interview, data for the Age 9 Follow up were collected between February and June 2019. Therefore, while the majority of children were nine years old at the time of the interview (average age 9.4 years), the children ranged in age from eight to 11 years old. In particular, 30% of the children were eight years old, 42% nine years old, 27% 10 years old, and 1% 11 years old.

Importantly, there were no statistically significant differences in the age of the children in the high and low treatment groups at the time of the interview (high treatment=9.5 years, low treatment=9.4 years;  $p$ -value=0.355). As there are no differences in the ages of the children across groups, we do not control for age in the results.

Data were collected through direct assessments with children, information provided by schools, and interviews with parents. To minimise detection bias, all assessments were conducted by trained researchers who were blind to the treatment condition (Eble, Boone, and Elbourne 2016). Detailed information on all instruments may be found in Appendix A.

#### *i) Direct assessment*

A total of 117 child assessments were conducted. Children were invited to participate in assessments of their cognitive skills using the British Ability Scales III: School Age Battery (BAS III; Elliott, Smith and McCulloch 2011) (an updated version of the assessment used in the phase I evaluation). The BAS III yields an overall score reflecting general cognitive ability (General Conceptual Ability, GCA), as well as three standardised scores for Verbal Ability, Pictorial Reasoning Ability, and Spatial Ability. The children also conducted several tasks assessing self-regulation/executive functions using the National Institutes of Health Toolbox for Assessment of Neurological and Behavioral Function Cognition Battery (NIH Toolbox; Zelazo and Bauer 2013) (e.g. the Flanker task to assess inhibitory control, the Dimensional Change Card Sort task to assess attention flexibility, and the List Sorting task to assess working memory). Finally, the Social Skills Improvement System Rating Scales (SSIS-RS; Elliot and Gresham, 2008) was used to assess self-reported socio-emotional development. Most assessments (99%) were conducted in the child's school. Where it was not possible to conduct the assessment in school, a venue familiar to the child (e.g. the home) was used. The direct assessments lasted approximately 45-60 minutes. Children's height and weight were also measured during either recruitment or the direct assessment.

#### *ii) School information*

Information on standardised test scores, absenteeism and resource supports were gathered from 32 schools. Data on test scores were available for 118 children, while data on absenteeism and school resources were available for 123 children. All schools in Ireland must complete standardised testing for reading and maths in second class, and they can choose to administer the tests more frequently if they wish. The test scores capture children's literacy and numeracy skills based on either the Drumcondra Tests for Reading and Mathematics or

the Micra-T for English Reading and Sigma-T for Mathematics tests. Schools choose which of the two tests to administer. Within our sample, 64% of children completed the Micra and Sigma tests and 34% the Drumcondra tests. The norm-referenced standard scores and categorisation of above and below average performance based on STen scores were used as indicators of children's reading and mathematics ability.

The absenteeism records include information on school attendance in 2017/18 and 2018/19 based on the number of days of attendance as a proportion of days in the school year. The resource supports records provide details on resource/learning support hours and special needs assistant support depending on the child's assessed level of need. Schools were asked what resource supports, if any, each child receives.

Table 2 compares the school characteristics of the high and low treatment groups and shows that there are no statistically significant differences across the groups in terms of the proportion attending designated disadvantaged status (DEIS) schools, single-gender schools, schools outside the *PFL* catchment area, and the distance of schools from the catchment area. In addition, there are no differences across the high and low treatment groups regarding the proportion who took the Drumcondra tests or the Micra/Sigma tests. This is important as it allows us to combine the test scores from both tests to estimate treatment effects for the whole sample. However, robustness tests where we analyse both tests separately are also conducted.

Although the *PFL* children attended a wide range of schools (32 in total), 55% attended one of the four schools in the *PFL* catchment area. In 26 of the other schools, only one or two *PFL* children were attending. A large proportion of the children were attending a DEIS school (71%). 76% were attending coeducational schools and 24% were attending single gender schools. The vast majority were attending Catholic schools (98%).

**Table 2** Comparison of high and low treatment groups: School data

	$M_{\text{HIGH}}$ (SD)	$M_{\text{LOW}}$ (SD)	$p^1$
School has DEIS status	0.73 (0.45)	0.68 (0.47)	0.558
Single gender school	0.27 (0.45)	0.21 (0.41)	0.408
School outside <i>PFL</i> catchment area	0.46 (0.50)	0.45 (0.50)	0.976
Driving distance from school to <i>PFL</i> village centre (in kms)	5.51 (14.98)	3.97 (7.25)	0.543
School conducts Micra/Sigma tests	0.67 (0.48)	0.60 (0.50)	0.432
<i>N</i>	70	53	

Notes: <sup>1</sup> two-tailed *p*-values calculated from permutation tests with 100,000 replications.

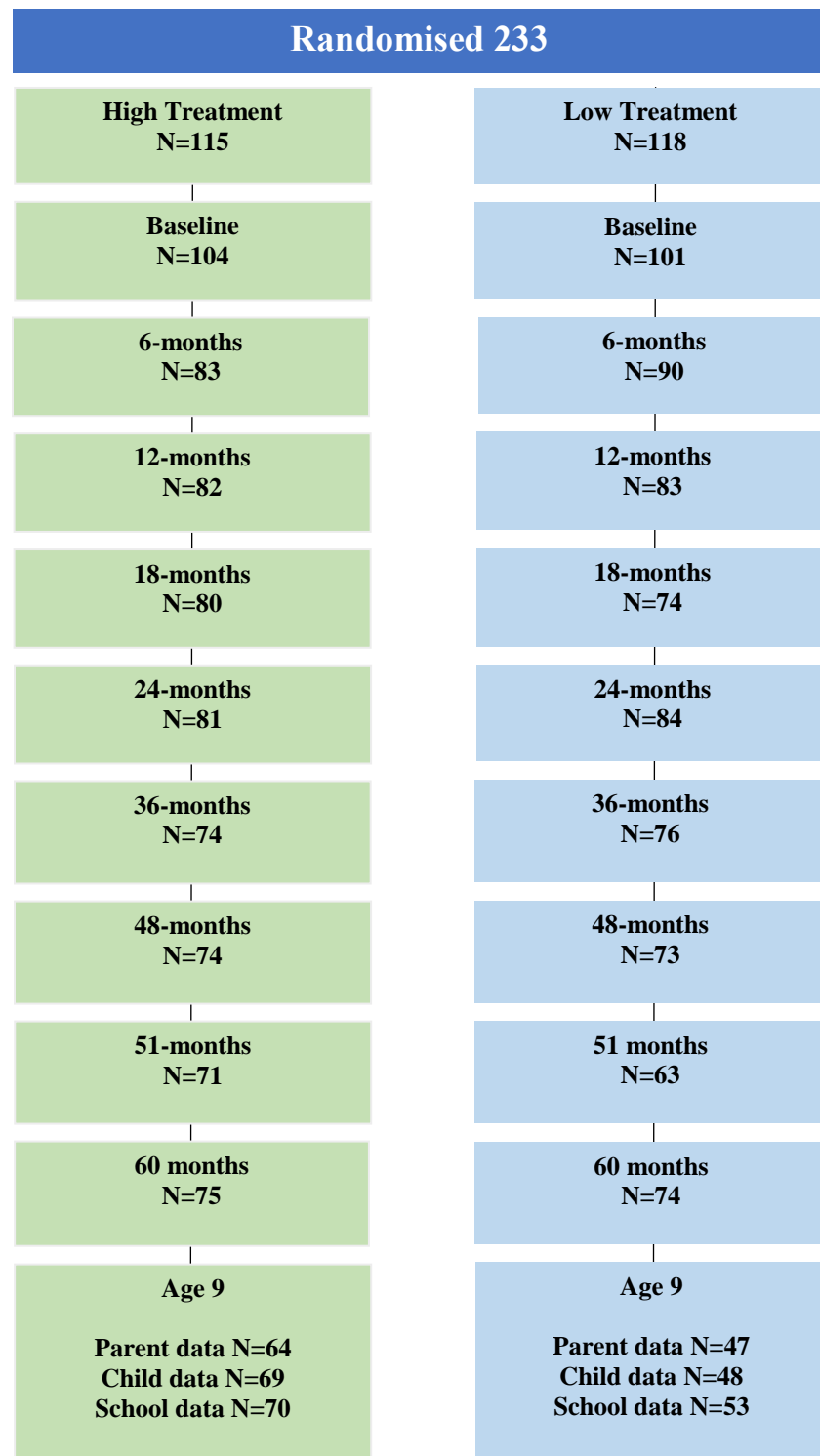
### *iii) Parent interview*

In total 111 parent interviews were conducted either online (n=39), over the phone (n=49), or in person (n=23) depending on the participant's preference. The questionnaire included measures to assess children's socio-emotional development (*Brief Problems Monitor* (BPM; Achenbach McConaughy, Ivanovaa and Rescorla 2011); *Strengths and Difficulties Questionnaire* (SDQ; Goodman 1997)), parenting attention (*Attentional Control Scale* (ACS; Derryberry and Reed 2002), and the home learning environment ((*Family Involvement Questionnaire* home-based subscale (FIQ: Fantuzzo, Tighe and Childs 2000)). The interview did not exceed 20 minutes duration

### *3.5.3 Age 9 Follow-up sample and attrition*

Figure 2 depicts the families' participation in the study between programme entry and the Age 9 Follow-up. At the follow-up, child assessment data were collected for 117 of the original 233 randomly assigned participants, representing an overall retention rate of 50% (high=59%; low=41%). For the school data, information on 123 participants were available, representing a retention rate of 53% (high=61%; low=45%). For the parent data, information was collected from 111 of the participants, representing an overall retention rate of 48% (high=56%; low=40%). While the level of attrition during the trial was largely equivalent across both groups, at the Age 9 Follow-up, attrition was significantly higher among the low treatment group. This may, in part, be attributed to the recruitment strategy which largely depended on the *PFL* mentors, with whom the high treatment families have greater contact. As shown in Table 1 above, these rates are somewhat lower than the follow-up studies of the mainly US-based home visiting programmes.

**Figure 2** Participant flow





A re-examination of the comparability of the high and low treatment groups at baseline using the Age 9 estimation samples shows that the two groups differ on 6.8% (8/117) of measures for the child assessment data, 10.3% (12/117) of measures for the school data, and 12.8% (15/117) of measures for the parent data. These are largely consistent with pure chance and indicate that the groups remain balanced at the Age 9 Follow-up, as confirmed by a joint test of all baseline variables. Table 3 compares the Age 9 participants in the high and low treatment group across the three different samples for a selection of baseline variables. It shows that there are no statistically significant differences across the two groups for a range of key socio-demographic and health factors including parental age, education, employment status, and health.

Although the estimation samples are largely balanced in terms of baseline characteristics, it is important to test for differential attrition in the high and low treatment group. To investigate this, the factors predicting participation in the child, school, and parent assessments were tested using bivariate tests with 50 baseline measures. Analyses were conducted separately for the high and low treatment groups to allow for differential attrition processes. In general, there is some evidence of differential attrition, with 17 to 36% of baseline measures predicting attrition from the high treatment group at age 9, and between 11 and 19% of measures predicting attrition from the low treatment group at age 9 (in two-tailed tests, using the 10% significance level).<sup>8</sup> Overall, there is less evidence of differential attrition in the school sample than the parent or child samples. The factors predicting attrition from both groups are largely similar, however the number of variables predicting attrition is lower in the low treatment group than the high treatment group. In line with much of the home visiting literature (see Roggman *et al.* 2008), families who did not take part in the Age 9 assessment had more risk factors, for example, they are less likely to be employed, have lower levels of education and IQ, are younger, and have poorer parenting skills. Table 4 compares a selection of baseline characteristics of those who participated in the Age 9 parent assessment - ‘stayers’ - to those who did not - ‘non-stayers’. It shows that high treatment parents who completed the age 9 assessment were older, had higher IQ, and were more likely to be employed at baseline, while low treatment parents who completed the age 9 assessment were less likely to be first time mothers. In order to account for differential attrition across

---

<sup>8</sup> For the child sample, 32% and 11% of baseline measures significantly predict attrition from the high and low treatment groups respectively. For the school sample, the corresponding figures are 17 and 19%. For the parent sample, the corresponding figures are 36 and 17%.

the high and low treatment groups, treatment effects are estimated using the Inverse Probability Weighting procedure detailed in the Methods section.

**Table 3** Baseline comparison of high and low treatment groups: Child, School, and Parent samples

	Child Sample			School Sample			Parent Sample		
	$M_{\text{HIGH}}$ (SD)	$M_{\text{LOW}}$ (SD)	$p^1$	$M_{\text{HIGH}}$ (SD)	$M_{\text{LOW}}$ (SD)	$p^1$	$M_{\text{HIGH}}$ (SD)	$M_{\text{LOW}}$ (SD)	$p^1$
Age	26.50 (5.59)	25.07 (5.82)	0.197	26.23 (5.67)	25.90 (6.10)	0.763	26.66 (5.69)	26.20 (6.12)	0.694
Married	0.147 (0.36)	0.133 (0.34)	0.890	0.145 (0.36)	0.180 (0.39)	0.515	0.156 (0.37)	0.178 (0.39)	0.782
No. of children	1.97 (1.27)	1.96 (1.07)	0.940	1.91 (1.25)	2.04 (1.07)	0.531	1.98 (1.30)	2.09 (1.10)	0.638
First time mother	0.500 (0.50)	0.422 (0.50)	0.415	0.522 (0.50)	0.380 (0.49)	0.127	0.516 (0.50)	0.378 (0.49)	0.154
Low education (left $\leq$ age 16)	0.294 (0.46)	0.311 (0.47)	0.920	0.304 (0.46)	0.320 (0.47)	0.875	0.328 (0.47)	0.311 (0.47)	0.851
Weschler Abbreviated Scale of Intelligence (WASI)	84.66 (11.35)	81.91 (14.03)	0.275	84.46 (11.38)	82.12 (13.68)	0.328	84.54 (11.55)	83.22 (13.54)	0.600
Employed	0.485 (0.50)	0.444 (0.50)	0.658	0.493 (0.50)	0.440 (0.50)	0.525	0.516 (0.50)	0.467 (0.51)	0.576
Resides in social housing	0.529 (0.50)	0.533 (0.51)	0.877	0.522 (0.50)	0.560 (0.50)	0.666	0.531 (0.50)	0.556 (0.50)	0.777
Medical card	0.574 (0.50)	0.667 (0.48)	0.289	0.565 (0.50)	0.660 (0.48)	0.310	0.547 (0.50)	0.622 (0.49)	0.445
Prior physical health condition	0.794 (0.41)	0.644 (0.48)	0.116	0.783 (0.41)	0.640 (0.49)	<b>0.078</b>	0.781 (0.42)	0.644 (0.48)	0.129
Prior mental health condition	0.309 (0.47)	0.267 (0.45)	0.642	0.290 (0.46)	0.280 (0.45)	0.949	0.281 (0.45)	0.311 (0.47)	0.747
Smoking during pregnancy	0.459 (0.50)	0.422 (0.50)	0.768	0.464 (0.50)	0.440 (0.50)	0.782	0.453 (0.50)	0.467 (0.51)	0.912
Drinking alcohol during pregnancy	0.309 (0.47)	0.311 (0.47)	0.946	0.319 (0.47)	0.300 (0.46)	0.844	0.297 (0.46)	0.333 (0.48)	0.692
<i>N</i>	~113			~119			~109		

**Notes:** All baseline measures were assessed during pregnancy prior to treatment delivery except for WASI which was assessed at 3 months postpartum. Baseline data are missing for four participants in the age 9 assessment. <sup>1</sup> two-tailed  $p$ -values calculated from permutation tests with 100,000 replications.

**Table 4** Baseline characteristics predicting attrition from the Age 9 Parent sample

	High Treatment Group			Low Treatment Group		
	$M_{\text{STAYER}}$ (SD)	$M_{\text{NON-STAYER}}$ (SD)	$p^1$	$M_{\text{STAYER}}$ (SD)	$M_{\text{NON-STAYER}}$ (SD)	$p^1$
Age	26.66 (5.69)	23.55 (5.66)	<b>0.006</b>	26.20 (6.12)	24.57 (5.74)	0.179
Married	0.156 (0.37)	0.125 (0.34)	0.679	0.178 (0.39)	0.179 (0.39)	0.897
First time mother	0.516 (0.50)	0.575 (0.50)	0.627	0.378 (0.49)	0.589 (0.50)	<b>0.037</b>
No. of children	1.98 (1.30)	1.88 (1.30)	0.692	2.09 (1.10)	1.77 (1.18)	0.164
Low education (left $\leq$ age 16)	0.328 (0.47)	0.350 (0.48)	0.712	0.311 (0.47)	0.464 (0.50)	0.124
Weschler Abbreviated Scale of Intelligence (WASI)	84.47 (11.32)	78.20 (13.00)	<b>0.013</b>	83.22 (13.54)	79.05 (12.13)	0.114
Employed	0.516 (0.50)	0.125 (0.34)	<b>0.000</b>	0.467 (0.51)	0.339 (0.48)	0.183
Resides in social housing	0.531 (0.50)	0.590 (0.50)	0.569	0.556 (0.50)	0.554 (0.50)	0.845
Medical card	0.547 (0.50)	0.675 (0.47)	0.182	0.622 (0.49)	0.696 (0.46)	0.444
Prior physical health condition	0.781 (0.42)	0.700 (0.46)	0.416	0.644 (0.48)	0.607 (0.49)	0.784
Prior mental health condition	0.281 (0.45)	0.275 (0.452)	0.978	0.311 (0.47)	0.179 (0.39)	0.133
Smoking during pregnancy	0.453 (0.50)	0.600 (0.50)	0.122	0.467 (0.51)	0.482 (0.50)	0.929
Drinking alcohol during pregnancy	0.297 (0.46)	0.175 (0.39)	0.167	0.333 (0.48)	0.214 (0.41)	0.226
<i>N</i>	64	40		45	56	

**Notes:** All baseline measures were assessed during pregnancy prior to treatment delivery except for WASI which was assessed at 3 months postpartum. Baseline data are missing for four participants in the age 9 assessment. <sup>1</sup> two-tailed  $p$ -values calculated from permutation tests with 100,000 replications.

### 3.6 Statistical methods

Using an intention-to-treat approach, the standard treatment effect framework defines the observed outcome  $Y_i$  of participant  $i \in I$  by:

$$(1) \quad Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) \quad i \in I = \{1 \dots N\}$$

where  $I = \{1 \dots N\}$  represents the sample space,  $D_i$  represents treatment assignment for participant  $i$  ( $D_i = 1$  for the high treatment group,  $D_i = 0$  for the low treatment group) and  $(Y_i(0), Y_i(1))$  are the potential outcomes for participant  $i$ . The null hypothesis of no treatment effect on outcomes is tested via:

$$(2) \quad Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$

Given the relatively small sample size, traditional hypothesis testing techniques which are based on large sample assumptions are not appropriate, thus the treatment effects are estimated using exact permutation-based hypothesis testing (see Good 2005).<sup>9</sup> The permutation tests are estimated by calculating the observed  $t$ -statistic. The data are then repeatedly shuffled so that the treatment assignment of some participants is switched (100,000 replications are used). The observed  $t$ -statistic is then compared to the distribution of  $t$ -statistics that result from the permutations. The mid- $p$  value is reported and is calculated as follows:

$$(3) \quad MP(t) = P(t^* > t) + 0.5P(t^* = t)$$

where  $P(.)$  is the probability distribution,  $t^*$  is the randomly permuted  $t$ -statistic, and  $t$  is the observed  $t$ -statistic. Similar to other early childhood intervention studies (e.g. Heckman *et al.* 2010; Campbell *et al.* 2014; Gertler *et al.* 2014; Conti, Heckman, and Pinto 2016), one-sided tests with the accepted Type I error rate set at 10% are used given the hypothesis that the high treatment will have a positive effect on children's outcomes.

As there was an imbalance in the proportion of girls and boys in the treatment groups at baseline, and given differential developmental trajectories by gender, all analyses control

---

<sup>9</sup> As permutation testing does not depend on the asymptotic behaviour of the test statistic, it is a more appropriate method to use when dealing with non-normal data (Ludbrook and Dudley 1998). A permutation test is based on the assumption of exchangeability under the null hypothesis, therefore if the null hypothesis is true, taking random permutations of the treatment variable does not change the underlying distribution of outcomes for the high or low treatment groups. Permutation testing has been shown to exhibit power advantages over parametric  $t$  tests in simulation studies, particularly when the degree of skewness in the outcome data is correlated with the size of the treatment effect (e.g. Mewhort 2005). Although this method is useful for dealing with non-normal data, it cannot be used to compensate for an under-powered study.

for gender.<sup>10</sup> As the assumption of exchangeability under the null hypothesis may be violated when controls are included, conditional permutation testing is applied. Using this method, the sample is proportioned into subsets, called orbits, each including participants with common background characteristics, in this case, there is one orbit for boys and one for girls. Under the null of no effect, the outcomes of the high and low treatment groups have the same distributions within an orbit. The exchangeability assumption is thus limited to strata defined by gender.

Although the few observed baseline differences are likely to be random, controlling for baseline covariates may improve the precision of treatment effects (Duflo, Glennerster, and Kremer 2008). Thus, as a robustness test, conditional permutation tests are estimated by controlling for key baseline differences on which the high and low treatment groups differ and may also affect child outcomes i.e., maternal knowledge of child development, parenting self-efficacy, maternal attachment, and maternal consideration of future consequences. Partitioning the sample into multiple orbits based on these variables can prove difficult, as the strata may become too small leading to a lack of variation within each orbit. To address this, a linear relationship is assumed between the control variables and the outcomes. Each outcome is regressed on the four variables assumed to share a linear relationship with child outcomes and the predicted residuals are permuted from these regressions within the orbits. This method, known as the Freedman–Lane procedure (Freedman and Lane 1983), has been demonstrated to be statistically sound in Monte Carlo studies (e.g., Anderson and Legendre 1999).

In order to account for any potential bias due to differential attrition, an Inverse Probability Weighting (IPW) technique (Robins, Rotnitzky, and Zhao 1994) is applied. First, logistic models are estimated to generate the predicted probability of participation in the Age 9 assessment. Given the number of significant predictors from the individual bivariate tests (up to 17) and the relatively small sample size, the Bayesian Information Criterion (BIC; Schwarz 1978) is used to reduce the number of variables included in the logistic models

---

<sup>10</sup> The high treatment group has more boys than the low treatment group (54% vs 36%). As recruitment occurred during pregnancy, this difference cannot be attributed to the treatment.

while estimating the model with best fit.<sup>11</sup> The predicted probabilities from these logistic models are then used as weights in the permutation tests so that a larger weight is given to participants that are underrepresented in the sample due to attrition.<sup>12</sup>

The issue of testing multiple outcomes at multiple time points, and thus increasing the likelihood of a Type-I error, is mitigated using the stepdown procedure which controls the Family-Wise Error Rate (Romano and Wolf 2005). Using this method all outcome measures are placed into a series of stepdown families each representing an underlying construct. For the child cognitive outcomes, the variables were placed into four stepdown categories capturing BAS Composite Scores, BAS Above Average Scores, BAS Below Average Scores, and NIH Toolbox Executive Functioning Scores. For the school outcomes, variables were placed in four stepdown families capturing Academic Standardised Scores, Academic Above Average Cutoff Scores, Academic Below Average Cutoff Scores, and Absenteeism/School Resources. For the child socio-emotional and health outcomes, the variables were placed into three stepdown categories capturing SSIS Socio-Emotional Scores, SSIS Socio-Emotional Below Average Cutoffs Scores, and Child Health. For the parent reported socio-emotional and parenting outcomes, variables were placed in three stepdown families capturing Child Socio-emotional Standardised Scores, Child Socio-emotional Cutoff Scores, and Parenting Outcomes. For the parent reported child health outcomes, variables were placed in three stepdown families capturing Child Health, Child Health Service Use in the Last 12 Months, and Child Meets Dietary Recommendations.

The stepdown procedure is conducted by calculating a  $t$ -statistic for each null hypothesis in the stepdown family using permutation testing. The results are placed in descending order. The largest  $t$ -statistic is then compared with the distribution of maxima permuted  $t$ -statistics. If the probability of observing this statistic is  $p \geq 0.1$  we fail to reject the

---

<sup>11</sup>The BIC measures goodness of fit and penalizes for the number of variables included in the model. The procedure is implemented in an iterative process. First, 50 baseline variables are included in an OLS regression modelling attrition and the BIC is calculated and stored. The process continues by testing each combination of 49 baseline variables in order to determine whether dropping any baseline variable would result in an increase in the predictive power as measured using the BIC. Prior to beginning this iterative process, the 50 baseline variables are placed in ascending order according to their effect size (in terms of predicting attrition). When iterating through the combinations of variables, the order in which variables are excluded depends on the effect size. Variables with the lowest effect size will be excluded first. For each combination of 49 variables, the new BIC is calculated and compared with the stored BIC. If the new BIC is smaller than the stored BIC the new BIC is stored and the excluded variable is dropped. A model resulting in a BIC that is within two points of the stored BIC is considered to have similar predictive power. Thus, only when the BIC is more than two points smaller is it considered a meaningful improvement in predictive power. This process is repeated by testing all combinations of 48 variables, and so on, until the optimal set of baseline variables has been found. Separate models for the high and low treatment groups are conducted at each time point.

<sup>12</sup> Participants who did not complete the baseline assessment, yet completed later assessments are assigned the average weight.

joint null hypothesis. If the probability of observing this  $t$ -statistic is  $p < 0.1$  the joint null hypothesis is rejected, and the most significant outcome is excluded, and the remaining subset of outcomes are tested. This process continues until the resulting subset of hypotheses fails to be rejected or only one outcome remains. By stepping down through the outcomes, the hypothesis that leads to the rejection of the null is found.

The results are discussed using  $p$ -values to indicate statistically significant effects, where  $p < 0.1$  is considered statistically significant, and Cohen's  $d$  effect sizes, where a small effect is 0.2, a medium effect is 0.5, and a large effect is 0.8.

## Section 4 Results

### 4.1 Child cognitive outcomes

Table 5 reports the Inverse Probability Weighted (IPW) adjusted means, standard deviations, and  $p$ -values that result from weighted individual and stepdown permutation tests, controlling for gender, alongside the effect size (as measured by the ratio of the treatment effect and the pooled standard deviation), for children's cognitive outcomes.<sup>13</sup> Children's cognitive and executive functioning skills are measured by the British Ability Scales III (BASIII) and NIH Toolbox respectively. The results indicate that the *PFL* programme had a significant and large impact on children's skills. The treatment increased children's overall GCA score by 0.67 of a standard deviation, which demonstrates that the high treatment group is better at thinking logically, making decisions, and learning. The results also demonstrate that *PFL* had a significant impact on each dimension of cognitive ability including spatial ability (0.48 of a standard deviation) which involves problem solving, spatial visualisation, and short-term visual memory; non-verbal reasoning ability (0.76 of a standard deviation) which involves inductive reasoning; and also verbal ability (0.39 of a standard deviation) which involves children's verbal reasoning, verbal knowledge, and expressive language. In addition, all four composite scores survive adjustment for multiple hypotheses testing.

---

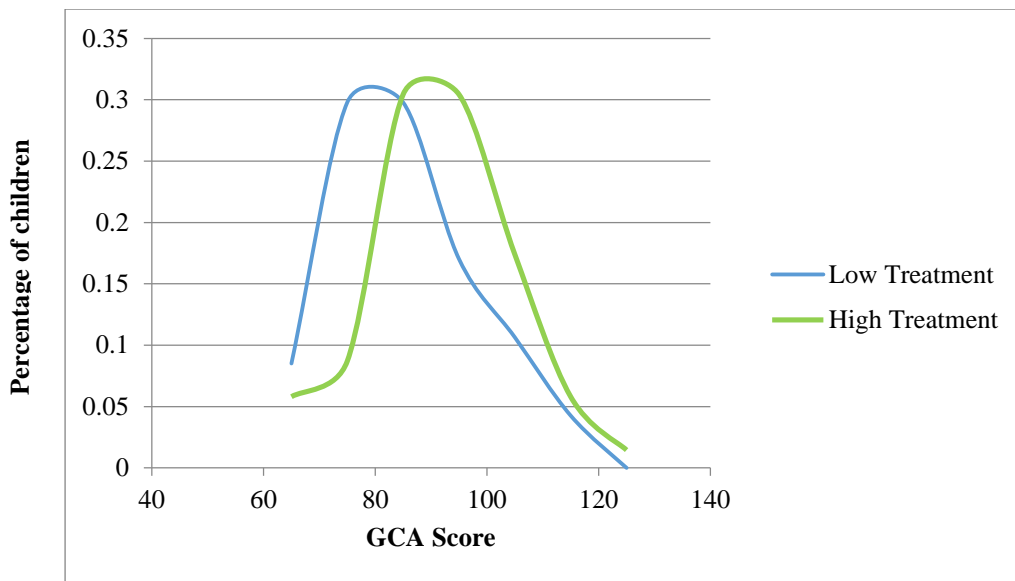
<sup>13</sup> Results excluding gender are largely similar. The non-IPW adjusted results are slightly more conservative with somewhat lower effect sizes. For example, 14 of the 16 cognitive outcomes reach conventional levels of significance in the individual IPW-adjusted permutation tests, compared to 9 in the non-adjusted results. However, all three stepdown families are still statistically significant in the non-adjusted results.



Children are classified as scoring above the norm if their score are above 110 points and below the norm if their scores are less than 90 points (range 51-122). High treatment children are also more likely to score *above* the norm in terms of their overall cognitive ability and their spatial ability; however neither result survives multiple hypotheses adjustment with effect sizes ranging from 0.01 to 0.36 of a standard deviation. In contrast, the high treatment children are less likely to score *below* the norm across all three cognitive domains, as well as overall ability, results which are robust to multiple hypothesis adjustment. The effect sizes range from 0.42 to 0.67 of a standard deviation. It is important to note that relatively few children, in either the high or low treatment group, score above the norm, while large proportions of children score below the norm. For example, only 2% of the high treatment group has above the norm GCA scores, while 58% have below the norm scores. The BAS III norms are based on a representative UK sample which includes children across all social groups. The scores identified here thus reflect the disadvantaged nature of the *PFL* cohort, where lower levels of cognitive ability are expected to be observed. Yet the counterfactual (low treatment group) reveals that without *PFL* intervention, a significantly greater proportion of the high treatment children would have scored below the norm, thus demonstrating the effectiveness of the programme.

The significant results regarding the proportion of children scoring below the norm suggests that the programme impacted the lower end of the distribution of children's skills. This is demonstrated in Figure 3 which shows that the distribution of GCA scores for the high treatment group is shifted to the right of the low treatment group's, with larger differences at the lower end of the distribution than the upper end.

***Figure 3 Distribution of BAS GCA cognitive scores at age 9***



The effects on the BAS scores at age 9 are similar in magnitude to the BAS results measured at the end of the programme (at approx. 51 months), demonstrating the sustained impact of the programme almost five years after the treatment ended.

The treatment also had an impact on children's executive functioning, the meta-cognitive processes that support concentration, reasoning, problem solving, and planning. Children in the high treatment group score significantly higher than the low treatment group on all three executive functioning tests assessing inhibitory control, the ability to override impulse responses; attention flexibility, the ability to deliberately focus and maintain attention; and working memory, the ability to retain and manipulate or use information over brief periods of time. All three tests survive multiple hypotheses testing, with effect sizes ranging from 0.56 to 0.66 of a standard deviation. The treatment also increased children's overall executive functioning by 0.73 of a standard deviation. These results are somewhat in contrast to the results for executive functioning reported at the end of the programme, where a significant treatment effect was found for children's ability to control their attention, but not their ability to delay gratification. These differences could be attributed to the tests administered at ages four and nine which measure different aspects of executive functioning and/or the validity of the tests at age four. The measures of executive functioning used at age nine are drawn from a battery of standardised cognition tasks, normed on a large reference population, and assessing a more comprehensive set of abilities.

**Table 5 Comparison of high and low treatment groups: Child Cognitive outcomes**

	N (HIGH/LOW)	M <sub>HIGH</sub> (SD)	M <sub>LOW</sub> (SD)	p <sup>1</sup>	p <sup>2</sup>	ES
<i>BAS Composite Scores</i>						
General Conceptual Ability	116 (69/47)	88.12 (11.85)	80.13 (12.11)	<b>0.002</b>	<b>0.006</b>	0.67
Spatial Ability	117 (69/48)	94.09 (14.26)	86.75 (16.27)	<b>0.032</b>	<b>0.045</b>	0.48
Non-Verbal Ability	117 (69/48)	84.63 (11.67)	76.53 (9.70)	<b>0.000</b>	<b>0.001</b>	0.76
Verbal Ability	116 (69/47)	92.22 (11.70)	87.27 (13.67)	<b>0.043</b>	<b>0.043</b>	0.39
<i>BAS Above the Norm %</i>						
General Conceptual Ability	116 (69/47)	0.02 (0.14)	0.00 (0.00)	<b>0.038</b>	0.239	0.29
Spatial Ability	117 (69/48)	0.11 (0.31)	0.10 (0.31)	0.406	0.679	0.01
Non-Verbal Ability	117 (69/48)	0.03 (0.17)	0.00 (0.00)	<b>0.023</b>	0.156	0.36
Verbal Ability	116 (69/47)	0.05 (0.22)	0.07 (0.26)	0.659	0.659	0.08
<i>BAS Below the Norm %</i>						
General Conceptual Ability	116 (69/47)	0.58 (0.50)	0.78 (0.42)	<b>0.029</b>	<b>0.029</b>	0.42
Spatial Ability	117 (69/48)	0.34 (0.48)	0.66 (0.48)	<b>0.007</b>	<b>0.018</b>	0.67
Non-Verbal Ability	117 (69/48)	0.70 (0.46)	0.89 (0.31)	<b>0.009</b>	<b>0.017</b>	0.51
Verbal Ability	116 (69/47)	0.33 (0.47)	0.60 (0.49)	<b>0.027</b>	<b>0.036</b>	0.58
<i>NIH Toolbox Executive Functioning</i>						
Flanker Task - Inhibitory Control	116 (69/47)	98.01 (16.64)	89.51 (11.39)	<b>0.046</b>	<b>0.049</b>	0.61
Dimensional Change Card Sort Task - Attention Flexibility	115 (69/44)	102.33 (21.68)	91.07 (12.45)	<b>0.035</b>	<b>0.054</b>	0.66
List Sorting Task - Working Memory	113 (69/44)	96.27 (13.43)	89.83 (9.48)	<b>0.008</b>	<b>0.008</b>	0.56
<i>Other</i>						
Composite Executive Function Score	113 (69/44)	0.39 (1.03)	0.22 (0.64)	<b>0.016</b>	~	0.73

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional *p*-value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional *p*-value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the pooled standard deviation.

## 4.2 School outcomes

Table 6 reports the IPW-adjusted means, standard deviations, and *p*-values that result from weighted individual and stepdown permutation tests, controlling for gender, alongside the effect size, for children's school outcomes.<sup>14</sup> Note that the sample size is lower for children's third class test scores as complete data were only available for children in second class as many of the children had not yet entered third class.

As shown in Table 6 the results indicate that the *PFL* programme had a significant and large impact on children's academic test scores at age nine, however there were no effects on absenteeism or use of school resources. Children in the high treatment group have significantly higher second and third class reading and maths standardised scores. All the standardised results survive adjustment for multiple hypotheses testing, and the effect sizes are 0.33 to 0.74 of a standard deviation. Children in the high treatment group are significantly less likely to score *below* the norm in terms of their reading and maths scores in both years, and are significantly more likely to score *above* the norm in terms of their third class reading and maths scores, as well as their third class maths score. Two of the four *above* the norm results survive multiple hypotheses adjustment – third class reading scores and second class maths scores. The effect sizes are 0.19 to 0.64 of a standard deviation. All four of the *below* the norm scores survive adjustment for multiple hypothesis testing, with effect sizes of 0.34 to 0.57 of a standard deviation. Finally, only one of the six absenteeism and school resources outcomes are statistically significant in the individual tests (high treatment children were less likely to have out-of-class educational supports), however it does not survive adjustment for multiple hypothesis testing.<sup>15</sup>

As a robustness test, the school results were re-estimated by conditioning on the school's DEIS status and an indicator of whether the child conducted the Drumcondra tests or the Micro/Sigma tests. The results, provided in Appendix Table B1, show that the results are robust to the inclusion of these controls.

---

<sup>14</sup> Results excluding gender are largely similar. The non-IPW adjusted results are also similar. There are no differences in the number (or level) of statistically significant results between the IPW and non-IPW adjusted models and the effect sizes are largely similar. However, more of the results survive multiple hypothesis adjustment.

<sup>15</sup> A separate analysis which included all 18 school outcomes in one stepdown family found that the standard reading score in third class, the standard maths score in second class, and the proportion scoring above average on maths scores in second class remained statistically significant in the stepdown analysis.

**Table 6** Comparison of high and low treatment groups: School outcomes

	N (HIGH/LOW)	M <sub>HIGH</sub> (SD)	M <sub>LOW</sub> (SD)	p <sup>1</sup>	p <sup>2</sup>	ES
<i>Academic Standardised Scores</i>						
2 <sup>nd</sup> class reading standardised score	118 (66/52)	99.55 (15.10)	94.93 (12.63)	<b>0.038</b>	<b>0.038</b>	0.33
2 <sup>nd</sup> class maths standardised score	117 (66/51)	97.42 (14.62)	89.73 (12.89)	<b>0.007</b>	<b>0.015</b>	0.56
3 <sup>rd</sup> class reading standardised score	70 (41/29)	97.51 (11.97)	89.67 (9.30)	<b>0.002</b>	<b>0.013</b>	0.74
3 <sup>rd</sup> class maths standardised score	70 (41/29)	94.42 (14.05)	88.05 (12.96)	<b>0.060</b>	<b>0.080</b>	0.47
<i>Academic Above the Norm Cutoff Scores %</i>						
2 <sup>nd</sup> class above average reading score	118 (66/52)	0.26 (0.44)	0.19 (0.39)	0.180	0.180	0.19
2 <sup>nd</sup> class above average maths score	117 (66/51)	0.22 (0.42)	0.05 (0.21)	<b>0.002</b>	<b>0.010</b>	0.55
3 <sup>rd</sup> class above average reading score	70 (41/29)	0.26 (0.44)	0.05 (0.22)	<b>0.006</b>	<b>0.020</b>	0.64
3 <sup>rd</sup> class above average maths score	70 (41/29)	0.14 (0.36)	0.05 (0.22)	<b>0.064</b>	0.164	0.34
<i>Academic Below the Norm Cutoff Scores %</i>						
2 <sup>nd</sup> class below average reading score	118 (66/52)	0.31 (0.47)	0.47 (0.50)	<b>0.043</b>	<b>0.043</b>	0.34
2 <sup>nd</sup> class below average maths score	117 (66/51)	0.34 (0.48)	0.55 (0.50)	<b>0.041</b>	<b>0.087</b>	0.43
3 <sup>rd</sup> class below average reading score	70 (41/29)	0.34 (0.48)	0.62 (0.49)	<b>0.024</b>	<b>0.098</b>	0.57
3 <sup>rd</sup> class below average maths score	70 (41/29)	0.45 (0.50)	0.72 (0.46)	<b>0.042</b>	<b>0.059</b>	0.55
<i>Absenteeism &amp; School Resources %</i>						
Proportion of days present in previous school year	104 (59/45)	0.92 (0.05)	0.93 (0.06)	0.762	0.762	0.20
Proportion of days present in current school year	122 (70/52)	0.94 (0.04)	0.94 (0.06)	0.561	0.763	0.00
In class supports	123 (70/53)	0.40 (0.49)	0.54 (0.50)	0.126	0.325	0.27
Out of class supports	123 (70/53)	0.24 (0.43)	0.38 (0.49)	<b>0.092</b>	0.307	0.31
SNA supports	123 (70/53)	0.06 (0.25)	0.05 (0.21)	0.499	0.798	0.08
Other supports	123 (70/53)	0.11 (0.31)	0.14 (0.35)	0.198	0.606	0.10

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional *p*-value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional *p*-value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the pooled standard deviation.

### 4.3 Child socio-emotional and health outcomes

Table 7 reports the IPW-adjusted means, standard deviations, and *p*-values that result from weighted individual and stepdown permutation tests, controlling for gender, alongside the effect size, for children's socio-emotional and health outcomes.<sup>16</sup>

The results indicate that the programme had no impact on children's socio-emotional skills at age nine as measured using the Social Skills Improvement System Behaviour Problems Subscale. Children in the high treatment group have fewer internalising problems, but more externalising problems, however these differences are not statistically significant, and the effect sizes are low. These results suggest that the moderate programme effects identified at age four, using the Child Behavioural Checklist, are no longer present at age nine.

The results also demonstrate that the significant effect on the proportion of overweight/obese children identified at age four is no longer statistically significant at age nine. In particular, there are no differences in the average BMI scores or the proportion of children in the high and low treatment groups classified as overweight/obese. Although a greater proportion of the low treatment group is obese (33% v's 30%), this difference is not statistically significant. This is higher than the 22% categorised as overweight/obese in the nationally representative Growing Up in Ireland cohort at age nine (GUI, 2018).

The programme also had no significant impact on the probability of having elevated resting pulse or blood pressure (systolic or diastolic) as measured using a wrist blood pressure monitor. Pulse and blood pressure are important vital signs and elevated recordings are indicative of cardiovascular risk (Sarganas *et al.*, 2017). Being classified as having an 'elevated' heart rate/blood pressure is based on scoring within the 95<sup>th</sup> percentile of the normed distribution, therefore children in the *PFL* sample have typical rates of elevated heart rates (~5%) and elevated rates of blood pressure (~10%).

---

<sup>16</sup> Results excluding gender are largely similar. The non-IPW adjusted results are similar to the IPW adjusted results, however two statistically significant differences emerge in the non-IPW adjusted individual permutation tests – children in the high treatment group are significantly less likely to fall within the above average range for bullying problems and overall behavioural problems – however, neither of these results survive multiple hypotheses adjustment.

**Table 7** Comparison of high and low treatment groups: Child socio-emotional and health outcomes

	N (HIGH/LOW)	$M_{\text{HIGH}}$ (SD)	$M_{\text{LOW}}$ (SD)	$p^1$	$p^2$	ES
<i>Child Socio-emotional Skills</i>						
SSIS Internalising Problems	117 (69/48)	7.50 (5.92)	8.24 (5.38)	0.264	0.523	0.13
SSIS Externalising Problems	117 (69/48)	6.42 (4.02)	5.24 (5.60)	0.760	0.760	0.25
SSIS Bullying	117 (69/48)	1.11 (1.23)	0.89 (1.90)	0.606	0.700	0.14
SSIS Hyperactivity/Inattention	117 (69/48)	6.43 (3.53)	6.40 (4.36)	0.428	0.643	0.01
<i>Child Socio-emotional Skills Cutoff %</i>						
SSIS Internalising Problems	117 (69/48)	0.10 (0.31)	0.10 (0.30)	0.505	0.505	0.02
SSIS Externalising Problems	117 (69/48)	0.07 (0.26)	0.14 (0.35)	0.285	0.377	0.22
SSIS Bullying	117 (69/48)	0.01 (0.10)	0.10 (0.31)	0.112	0.177	0.45
SSIS Hyperactivity/Inattention	117 (69/48)	0.12 (0.33)	0.11 (0.32)	0.495	0.625	0.01
<i>Child Health %</i>						
Overweight/Obese	116 (67/49)	0.30 (0.46)	0.33 (0.48)	0.383	0.735	0.07
Elevated heart rate	113 (67/46)	0.07 (0.25)	0.04 (0.19)	0.760	0.760	0.15
Elevated systolic blood pressure	117 (69/48)	0.12 (0.32)	0.10 (0.30)	0.649	0.835	0.06
Elevated diastolic blood pressure	113 (67/46)	0.11 (0.32)	0.10 (0.31)	0.573	0.820	0.03
<i>Other</i>						
SISS Problem Behaviours Standardised Total Score	117 (69/48)	96.65 (10.20)	97.42 (14.64)	0.381	~	0.06
SISS Problem Behaviours ~ 85%ile	117 (69/48)	0.04 (0.20)	0.10 (0.30)	0.869	~	0.23
BMI	116 (67/49)	18.19 (3.48)	18.79 (3.90)	0.279	~	0.16
Obese %	116 (67/49)	0.14 (0.35)	0.25 (0.44)	0.150	~	0.27

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional  $p$ -value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional  $p$ -value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the pooled standard deviation.

#### 4.4 Parent reported outcomes

Tables 8 and 9 report the IPW-adjusted means, standard deviations, and *p*-values that result from weighted individual and stepdown permutation tests, controlling for gender, alongside the effect size, for parent reported outcomes.<sup>17</sup> Table 8 reports the results of parent reported child socio-emotional development, as well as parenting behaviour. The results indicate that the *PFL* programme did not have a significant impact on children's socio-emotional development or parenting behaviour at age nine. Although children in the high treatment group have statistically significantly fewer peer problems, attention problems, and total behavioural problems according to the individual permutation tests, these results do not survive adjustment for multiple hypothesis testing. The programme also had no impact on the proportion of children with ADHD or learning difficulties. These parent-reported results for socio-emotional problems are consistent with the child-reported results detailed in Table 7 where no evidence of treatment effects are found.

In addition, none of the parenting outcomes concerning the home-based learning interactions, internet security, or parental attention are statistically significant. In most cases, children in the high treatment group have better outcomes than the low treatment group, however for the parenting outcomes, the reverse was true. In sum, these results suggest a fade-out of the programme's impact on socio-emotional and behavioural outcomes that were observed at age four.

Table 9 reports the results on parent reported child health. The results indicate that the programme has no impact on children's health at age nine as measured by the child's general health, health service use in the previous 12 months, and diet quality. In most cases, the high treatment group has better health than the low treatment group, but none of these differences are statistically significant in either the individual or stepdown tests.

---

<sup>17</sup> The non-IPW adjusted results and results excluding gender are also similar.



**Table 8** Comparison of high and low treatment groups: Parent reported child socio-emotional & parenting outcomes

	N (HIGH/LOW)	M <sub>HIGH</sub> (SD)	M <sub>LOW</sub> (SD)	p <sup>1</sup>	p <sup>2</sup>	ES
<i>Child Socio-emotional Standardised Scores</i>						
SDQ Peer problems	111 (64/47)	1.59 (1.61)	2.11 (1.73)	<b>0.060</b>	0.209	0.31
SDQ Prosocial behaviour	111 (64/47)	8.85 (1.71)	8.78 (1.34)	0.279	0.490	0.05
BPM Internalising problems	110 (63/47)	57.61 (6.35)	57.74 (6.13)	0.348	0.348	0.02
BPM Externalising problems	110 (63/47)	53.79 (5.74)	54.79 (6.25)	0.282	0.437	0.17
BPM Attention problems	110 (63/47)	53.89 (7.46)	56.71 (7.10)	<b>0.038</b>	0.136	0.39
<i>Child Socio-emotional Cutoff Scores %</i>						
SDQ Peer problems	111 (64/47)	0.13 (0.34)	0.18 (0.39)	0.224	0.508	0.15
SDQ Prosocial behaviour	111 (64/47)	0.08 (0.27)	0.01 (0.12)	0.899	0.899	0.34
BPM Internalising problems	110 (63/47)	0.21 (0.41)	0.24 (0.43)	0.244	0.566	0.07
BPM Externalising problems	110 (63/47)	0.07 (0.27)	0.18 (0.39)	0.187	0.325	0.31
BPM Attention problems	110 (63/47)	0.13 (0.33)	0.15 (0.36)	0.350	0.459	0.07
<i>Parenting Outcomes</i>						
Family Involvement Questionnaire (Home-based subscale)	111 (64/47)	38.53 (6.28)	39.24 (6.25)	0.606	0.929	0.11
Attentional Control Scale (Attentional focusing subscale)	110 (63/47)	26.94 (4.29)	27.88 (4.35)	0.717	0.717	0.22
Always supervises internet use %	105 (60/45)	0.54 (0.50)	0.56 (0.50)	0.513	0.943	0.06
Rules in place to restrict internet access %	103 (59/44)	0.90 (0.31)	0.90 (0.30)	0.495	0.933	0.01
Restrictions in place to restrict internet access %	103 (59/44)	0.79 (0.41)	0.84 (0.37)	0.617	0.873	0.12
<i>Other</i>						
BPM Total problems standardised score	110 (63/47)	54.60 (6.87)	56.70 (6.46)	<b>0.079</b>	0.144	0.32
Child has ASD-ADHD %	111 (64/47)	0.08 (0.27)	0.04 (0.20)	0.622	0.622	0.14
Child has learning difficulty %	111 (64/47)	0.06 (0.23)	0.12 (0.33)	0.177	0.213	0.22

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional *p*-value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional *p*-value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the pooled standard deviation.

**Table 9** Comparison of high and low treatment groups: Parent reported child health outcomes

	N (HIGH/LOW)	M <sub>HIGH</sub> (SD)	M <sub>LOW</sub> (SD)	p <sup>1</sup>	p <sup>2</sup>	ES
<i>Health</i>						
Child has good health %	111 (64/47)	0.97 (0.16)	0.94 (0.24)	0.370	0.637	0.18
Child is diagnosed with asthma %	111 (64/47)	0.13 (0.34)	0.18 (0.39)	0.202	0.392	0.12
Child has ongoing diagnosed chronic illness (other than asthma) %	111 (64/47)	0.18 (0.39)	0.21 (0.41)	0.286	0.487	0.06
No. of hours sleep on average night	111 (64/47)	10.58 (0.80)	10.70 (0.79)	0.676	0.676	0.14
Child has sleep problems %	111 (64/47)	0.45 (0.50)	0.52 (0.50)	0.260	0.515	0.14
<i>Health Service Use in last 12 months %</i>						
Attended GP	111 (64/47)	0.55 (0.50)	0.65 (0.48)	0.232	0.706	0.20
Attended Emergency Department	111 (64/47)	0.21 (0.41)	0.23 (0.42)	0.493	0.862	0.04
Attended routine dental check-up	111 (64/47)	0.53 (0.50)	0.63 (0.49)	0.727	0.727	0.20
Attended health facility due to accident	110 (63/47)	0.15 (0.36)	0.10 (0.30)	0.805	0.891	0.15
Attended health facility due to chest infection	110 (63/47)	0.14 (0.35)	0.17 (0.38)	0.277	0.839	0.09
Attended health facility due to wheezing or asthma	110 (63/47)	0.08 (0.28)	0.11 (0.32)	0.249	0.866	0.10
Received dental treatment	111 (64/47)	0.44 (0.50)	0.44 (0.50)	0.619	0.888	0.02
<i>Meets dietary recommendations %</i>						
Grains	111 (64/47)	0.59 (0.50)	0.62 (0.49)	0.687	0.905	0.06
Protein	111 (64/47)	0.28 (0.45)	0.29 (0.46)	0.638	0.935	0.04
Dairy	111 (64/47)	0.11 (0.31)	0.11 (0.31)	0.540	0.951	0.00
Fruit & Vegetables	111 (64/47)	0.06 (0.24)	0.09 (0.29)	0.783	0.783	0.13

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional *p*-value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional *p*-value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the pooled standard deviation.

## **4.5 Conditioning on baseline differences**

As a robustness test, the results are re-estimated conditioning on four variables on which there were significant differences between the high and low treatment groups at baseline and may impact child outcomes – namely maternal knowledge of child development, parenting self-efficacy, maternal attachment, and maternal consideration of future consequences.

Appendix Table B2 shows that the conditional results for child cognitive outcomes are largely similar to the main results, however two of the outcomes which reach conventional levels of significance in the individual permutation tests are not significant in the conditional results – general conceptual ability and non-verbal above average scores. The remaining results are similar in terms of effect size and significance. Appendix Table B3 shows that the conditional results for the school outcomes are similar to the main results in terms of the number of statistically significant differences and the effect sizes. Appendix Table B4 shows that the conditional results for child socio-emotional and health outcomes are largely similar to the main results, however one outcome is statistically significant in the conditional results which was not significant in the unconditional results – proportion of children with above average behavioural problems – however it does not survive multiple hypothesis adjustment. Appendix Table B5 shows that the three individual significant treatment effects found in the main results for children’s socio-emotional standardised scores are no longer significant once controls are added. The results for child health, reported in Appendix Table B6, are also similar to the main results in that no significant effects are found. This again confirms a lack of programme impact on children’s socio-emotional wellbeing and health at age nine.

## Section 5 Conclusions

The aim of the Age 9 Follow-up study was to examine whether the large and significant impacts of *PFL* found at the end of the programme were sustained at age 9. Prior evidence on the medium-term impact of home visiting and parent-training programmes in middle childhood is inconclusive, with many studies experiencing a dissolution of effects once the programme ends. In contrast, this study finds that *PFL* continues to have a sizeable impact on children's cognitive skills and achievement tests approximately five years after the families have finished the programme. There is no evidence of cognitive fade-out, with effect sizes of 0.67 of a standard deviation on overall cognitive ability, and significant effects on executive functioning and standardised school achievement tests. In particular, there is an eight-point gap between the high and low treatment groups' general conceptual ability scores, which is a close proxy for IQ. While this is slightly smaller than the 10-point gap found at end of the programme, it is still a sizeable difference considering that the low treatment group has received five years of formal schooling.

Overall, the IQ scores of the *PFL* children are above that of their parents (i.e., the Flynn effect), however the correlation between the high treatment children and their mothers is small and not statistically significant at either age five ( $r^{18} = 0.07, p = 0.562$ ) or age nine ( $r = 0.18, p = 0.148$ ), compared to the large and significant correlation between the low treatment children and their mothers at age five ( $r = 0.31, p = 0.018$ ) and age nine ( $r = 0.57, p = 0.001$ ).<sup>19</sup> Indeed, the correlation for the low treatment group is similar to the correlation of 0.38 between fathers and sons found in Black, Devereux, and Salvanes (2009). Thus the programme appears to be effective in reducing the intergenerational transmission of IQ scores. Note that the correlation between parents and children's IQ is growing over time, therefore it will be informative to track this correlation as the children enter secondary school.

The programme impacted all dimensions of cognitive skill including spatial ability, non-verbal ability, and verbal ability, in addition to reducing the proportion of children scoring below the standardised norm. The programme is particularly beneficial for children at

---

<sup>18</sup>  $r$  is Pearson's correlation coefficient.

<sup>19</sup> Maternal IQ was measured using the Weschler Abbreviated Scale of Intelligence which assesses cognitive ability across four subscales: vocabulary, similarities of constructs, block design, and matrix reasoning. From this, standardised measures of verbal ability, perceptual reasoning, and a full-scale measure of cognitive functioning, standardised to have a mean of 100 and standard deviation of 15, are generated. The full-scale measure was used in this analysis to correspond with the measure of General Conceptual Ability from the BAS.

the lower end of the skills distribution. This contrasts somewhat from earlier results where there was evidence that the programme shifted the entire distribution of cognitive skills. This difference may have occurred as naturally higher ability children, regardless of their treatment status, may be better placed to take advantage of the learning and supports provided in school. The magnitudes of the cognitive effects at age nine (0.39 - 0.76 SD) are similar, albeit a little smaller, to those found at the end of the programme (0.56 - 0.77 SD). An additional analysis shows that controlling for age five cognitive scores, reduces the size of the age nine treatment effects, however the impact of the programme on the age nine scores is still statistically significant.<sup>20</sup> This suggests that *PFL* is continuing to have an impact on children's development beyond the lifetime of the programme.

This provides evidence in support of the skill formation model (Cunha and Heckman, 2007) which posits that developing children's skills early in life helps them to develop more advanced skills later in life (a process called self-productivity), and this raises the effectiveness of later investments, such as investments in schooling (a process called dynamic complementarity). If this process continues, and the high treatment group continue to utilise their higher cognitive skills, this is likely to translate into improved outcomes throughout the life cycle. Indeed, the large and significant treatment effects found for the children's school achievement tests at age nine suggest that these higher cognitive skills are already having an impact on the children's performance in school. Children who received the high treatment supports had better second and third class standardised test scores in reading and maths, with effect sizes ranging from 0.33 - 0.74 of a standard deviation. The high treatment group was significantly less likely to score below the norm on their second and third class reading and maths tests and more likely to score above the norm on their third class reading and second class maths tests.<sup>21</sup>

Unlike the cognitive tests which were only conducted on the *PFL* cohort, these tests were conducted on all children in all schools in Ireland as part of national standardised testing, therefore they provide an independent assessment of the children's reading and maths ability. In addition, they allow us to compare the *PFL* cohort to the national norm. Although the high treatment group have significantly better reading and maths scores than the low treatment group, a large proportion of both the high and low treatment group are still scoring below the national norm. For example, 31% and 47% of the high and low treatment group

---

<sup>20</sup> Available upon request.

<sup>21</sup> Note that the third class results are based on a smaller sample size as some of the children had not yet started third class at the time of data collection, thus the third class results should be interpreted with caution.

respectively score below the norm on their reading scores, and 34% and 55% score below the norm on their maths scores. This suggests that continued investment in schools is required to break long-standing socio-economic inequalities in children's skills.

An interesting finding which emerged from the Age 9 Follow-up data collection was the number and diversity of primary schools attended by the *PFL* cohort. In total, the 123 children are attending 32 different schools. While the majority are attending DEIS schools (73% and 68% in the high and low treatment group respectively), there is still some heterogeneity. In addition, although 55% are attending schools within the *PFL* catchment area, many children are attending schools outside the catchment area. As equal proportions of the high and low treatment groups are attending schools outside the catchment area, this suggests that the programme did not change parents' preferences regarding the type or location of their child's primary school. From an evaluation perspective it is useful as it suggests that the significant treatment effects found for children's cognitive ability and achievement tests are unlikely to be driven by differences in school type.

A key innovation of the Age 9 Follow-up data collection was to conduct a battery of tests assessing the children's executive functioning skills. Although two dimensions of executive functioning were assessed at age four, the validity of these tests was limited given constraints in data collection. Increasingly emphasis is being placed on promoting children's executive functioning skills, especially self-regulation, as independent of IQ, self-regulation has been shown to predict later academic performance, health, and finances in adulthood (Blair and Raver 2012; Liew 2012). However, children from disadvantaged backgrounds typically have poorer self-regulation (Evans and Rosenbaum 2008). Thus improving children's early skills in these domains could yield cascading benefits into adulthood (Diamond and Lee 2011). The programme had a large and substantive impact on all dimensions of the children's executive functioning skills with effect sizes ranging from 0.56 – 0.65 of a standard deviation. Children in the high treatment group are better able to override their automatic impulses (inhibitory control); they are better at maintaining and focusing their attention (attention flexibility); and they are better able to retain, manipulate and use information over brief periods of time (working memory). These skills are particularly important in the school environment as they allow children to assimilate knowledge, as well as fulfil classroom expectations (Neuenschwander *et al.* 2012), which may have contributed to the higher achievement scores observed above.

In contrast to the results for children's cognitive development, the programme has no impact on school absenteeism or the use of school resources. The result for absenteeism is not

particularly surprising given that most of the schools have home-school liaison officers who directly address this issue, thus resulting in very low proportions of children missing school (94% of the high and low treatment group are present throughout the school year). The rates of absenteeism among the *PFL* cohort is also very similar to the national average rate of absenteeism which ranged from 5.4% to 5.9% between 2012/13 and 2016/17 (Miller 2018). However, the lack of significant treatment effects for the use of additional school resources, such as a special needs assistant and additional literacy and maths support, are more surprising, especially given the observed differences in the children's test scores. Indeed, only 6% and 5% of the children in the high and low treatment group respectively have a special needs assistant. While the low treatment group avail of a higher proportion of in-class and out-of-class support (54% and 38% respectively) compared to the high treatment group (40% and 24% respectively), these differences were not statistically significant. The lack of differences across the two groups may reflect the move to provide additional support to all children within the classroom environment rather than singling individual children out for treatment.

The results also indicate that the significant effects observed for children's socio-emotional development at age four are no longer present at age nine. At earlier time points we found that the programme was effective in reducing the proportion of children within the clinical range of behavioural problems, however, few effects were identified for continuous scores of children's socio-emotional development. These earlier measures were based on parent reports only. At age nine we assessed children's socio-emotional skills using both parent and child reports. The lack of treatment effects across both measures, for either the continuous or clinical range, suggests a dissolution of the *PFL*'s earlier impact on children's wellbeing. Given the size of the effects, these results are unlikely to be driven by small sample size issues. These results are in-line with the home visiting literature reported in Table 1 which found significant treatment effects in only three of the 10 follow-up studies conducted. The *PFL* treatment effects observed at age four were driven by a small number of children exhibiting significant behavioural problems, thus it is possible that exposure to the school environment, as well as natural child maturation, has helped to reduce these behavioural issues. Indeed, the overall incidence of clinically significant problems at age nine is low in the *PFL* cohort. For example, only 8% of the high treatment group and 4% of the

low treatment group are diagnosed with ASD or ADHD.<sup>22</sup> The higher rate in the high treatment group may reflect the high treatments groups' ability to advocate for their children in terms of receiving a diagnosis and/or that the children's conditions were identified by the *PFL* mentors during the programme, and thus an early diagnosis was sought.

Also in-line with studies of other home visiting programmes, there is little evidence that the programme continues to have an impact on children's health or parenting behaviour. While few studies examine the long-term impact of home visiting programmes on health, those that do typically find little evidence of effects on children's physical or mental health (Dumont *et al.* 2010; Kitzman *et al.* 2010; Minkowitz *et al.* 2007; Olds *et al.* 2004; Olds *et al.* 2007). At age nine, the *PFL* programme had no impact on children's general health, health service use, dietary intake, heart rate, blood pressure, and BMI. Approximately 30% of high treatment children and 33% of low treatment children are categorised as overweight or obese. By comparison, a lower proportion of high treatment children (26%) and a larger proportion of low treatment children (41%) were categorised as overweight or obese at age 4. As the *PFL* children are now spending a greater proportion of their day outside the home environment, their diet and exercise are becoming more independent and may be more open to influence from peers and/or the school environment. Thus, the family, for whom the intervention is targeted, may have less input on the child's weight status at age nine than at age four. That said, the results indicate that a significant proportion of the *PFL* cohort are not meeting their daily dietary recommendations. For example, while ~60% of children meet the daily recommended amount of grains, only 28% of children meet their daily protein in-take, 11% meet their daily dairy in-take, and only 7% meet the daily recommended in-take of fruit and vegetable. While these deficits cannot be attributed to either the home or school environment alone, it does indicate that the overall dietary health of this cohort is poor, and that specific supports targeting the school and home environment may be required to bolster children's development in this area.

The Age 9 Follow-up study also sought to assess changes in parenting behaviour as a result of earlier participation in the programme. The results suggest that the programme had no impact on parental involvement in their children's education, parents' ability to control their attention, parental supervision of child internet use at age nine. These results somewhat align with the limited differences observed on parenting throughout the programme, and with the general home visiting literature. The *PFL* programme is based on the premise that

---

<sup>22</sup> The higher proportion in the high treatment group may be attributed to the parents' increased ability to recognise problems and seek out a diagnosis.



providing support to parents will increase their knowledge of appropriate parenting practices and change their attitudes and parenting behaviours. These positive changes would then impact on children's development as a result of improved parental stimulation, interactions, and strategies. Although very few treatment effects were observed for parental wellbeing during the programme (see Doyle *et al.* 2017b), parents did make some important behavioural changes which may have contributed to their children's advanced skills. For example, Doyle *et al.* (2017a) identify significant treatment effects on parenting skills at six and 18 months in terms of improving the quality of the home environment, O'Sullivan *et al.* (2017) find positive treatment effects regarding improved nutrition at 24 months, and Doyle and PFL Evaluation Team (2015) report improved parenting behaviour regarding the use of appropriate disciplinary techniques and increased parental interactions. These practices, interactions, and activities are recognised as key means of stimulating children's development (Farah *et al.* 2008). Thus, while there appears to be a fade-out of the programmes impact on parents at age nine, the cumulative improvements in parenting over the course of the programme may account for the significant effects on children's cognitive development found in this report.

The size of the cognitive effects are substantially larger than those found in much of the existing literature. For example, the meta-analyses discussed earlier in the report find effect sizes of less than 0.30 for cognitive outcomes (e.g. Layzer *et al.* 2001; Sweet and Appelbaum, 2004; Miller *et al.* 2011; Filene *et al.* 2013; Rayce *et al.* 2017). The effects are also larger than the German home visiting programme, *Pro Kind*, which finds average effect sizes for cognition of 0.20 - 0.30 SD for girls only at age 2 (Sandner and Jungmann 2017). The effect sizes are also larger than a recent re-analysis of the Nurse Family Partnership Memphis trial which reported effects of 0.13 - 0.27 SD for cognitive skills at age six (Heckman *et al.* 2017). The PFL effects are more similar in magnitude to those found in studies of low and middle income countries. The Jamaica study, which is based on weekly home visits for two years starting between nine and 24 months, identified no significant effects on IQ at ages seven to eight, however the cognitive effects re-emerged at the 11, 17, and 22 year follow-ups with effect sizes ranging from 0.40 to 0.60 (Grantham-McGregor and Smith, 2016).

Although it is difficult to fully compare the results from different home visiting studies due to wide variations in programme goals, target groups, and implementation practices (Gomby *et al.* 1999), the larger effect sizes identified for the PFL programme, particularly for the cognitive outcomes, may be attributed to its prenatal start, its longer

programme length, its multiple connected treatments, and its inclusive eligibility criteria. In particular, *PFL* both starts earlier and is longer in duration than most other home visiting programmes. The *PFL* mentors worked with participants for a substantial and critical period of their children's lives; therefore the positive and sizable treatment effects may be a result of the strength and quality of the mentor-parent relationship which was given an appropriate length of time to build and develop. This is consistent with the home visiting literature which finds that the bond between parents and programme staff is key for understanding programme effects (Wesley, Buysse, and Tyndall 1997).

The larger effects may also be attributed to the extensive and diverse supports offered to the high treatment group. The *PFL* treatment included baby massage classes during infancy and the *Triple P* programme from age two, yet the majority of the other standalone home visiting programmes, such as Nurse Family Partnership and its European equivalents, do not provide such supports. Therefore, a multi-component approach, which offers supports in a variety of formats and settings may help to engage families who favour one form of treatment over another. However, as participants were not randomised to receive different components of the treatment, it is not possible to tease out the impact of the three different provisions. The larger effect sizes may also be attributed to the nature of the sample. Compared to many other home visiting programmes which include ethnically diverse samples, the *PFL* cohort is relatively homogenous, consisting mainly of ethnically-Irish born participants. This, coupled with the individual-level randomisation in a confined geographical space, reduces variability within the sample, and allows us to uncover treatment effects if indeed they exist.

Much of the policy focus on early childhood interventions has been attributed to the long-run findings from preschool programmes which target children directly (e.g. Head Start) and typically start at age three. Interventions which target parents and/or start in pregnancy have a smaller evidence base concerning their long-term effectiveness, yet are less intensive and are therefore less costly than their centre-based counterparts. On average, *PFL* participants received ~50 hours of treatment over five years, compared to, for example, the Perry Preschool Programme which involved >50 hours of treatment per month for eight months over one/two years (~400/800hrs). In addition, unlike most centre-based programmes, home visiting programmes and parent training programmes target parenting behaviour and childrearing practices, thus they can potentially impact all children in the family rather than the target child alone. Thus the lower level of treatment intensity, and

possible spillovers to other children in the family, may further improve the economic return to programmes such as *PFL*.

Cost-benefit analyses of some of the most well-known US-based home visiting programmes finds returns ranging from \$US1.61 for the Nurse Family Partnership programme, \$US3.29 for Parents as Teachers, and \$US1.21 for Healthy Families America per \$US invested, with total programme costs of \$US10,049, \$US2,688, and \$US4,797 respectively (Washington State Institute for Public Policy 2016). In addition, cost-benefit analyses of the Head Start programme by Ludwig and Philips (2007) and Deming (2009) find that effect sizes on cognitive skills of 0.10 - 0.20 SDs and 0.06 SDs respectively, are enough to satisfy cost-benefit tests, based on an average cost per child of ~\$US7,000. Therefore, if the significantly larger effects identified in this study translate into future financial gains both for the individual participants and wider society, the *PFL* programme is likely to generate similar positive returns.

To conclude, the sizable cognitive advantages generated by the *PFL* programme are likely to have positive impacts on the children's outcomes throughout life. Thus it is critical to continue to track the *PFL* cohort as they progress from primary to secondary school. One concern as we move forward with the *PFL* evaluation is sample size. While a response rate of 50% was achieved at the Age 9 Follow-up, this figure is likely to reduce further as the children become older and start attending secondary school. It would be useful to identify a means of tracking the participants through administrative data (for example through the Department of Education or using health records) that would allow us to identify participants without the reliance on personal contact details which are subject to change. Therefore maintaining the *PFL* cohort should be a priority if we are to assess the long term impact of the programme. This is particularly important given the magnitude of the cognitive effects, especially in comparison to other intervention programmes, as *PFL* can provide a model for other communities aiming to reduce long-term socioeconomic inequalities.

## References

- Aboud, Frances E., and Aisha K. Yousafzai.** 2015. "Global Health and Development in Early Childhood". *Annual Review of Psychology* 66: 433–457.
- Achenbach, T.M., McConaughy, S.H., Ivanova, M.Y., & Rescorla, L.A.** 2011. *Manual for the ASEBA Brief Problem Monitor (BPM)*. Burlington, VT: ASEBA, 1-33.
- Almond, Douglas, Janet Currie, and Valentina Duque.** 2017. "Childhood Circumstances and Adult Outcomes: Act II". NBER Working Paper No. 23017.
- Anderson, Marti J., and Pierre Legendre.** 1999. "An Empirical Comparison of Permutation Methods for Tests of Partial Regression Coefficients in a Linear Model." *Journal of Statistical Computation and Simulation* 62 (3): 271–303.
- Avellar, Sarah, Diane Paulsell, Emily Sama-Miller, Patricia Del Grosso, Lauren Akers, and Rebecca Kleinman.** 2016. *Home Visiting Evidence of Effectiveness Review: Executive Summary*. Office of Planning, Research and Evaluation. Administration for Children and Families, U.S. Department of Health and Human Services. Washington, DC.
- Avellar, Sarah A., and Lauren H. Supplee.** 2013. "Effectiveness of Home Visiting in Improving Child Health and Reducing Child Maltreatment." *Pediatrics* 132 (Supplement 2): S90.
- Bailey, Drew, Greg J. Duncan, Candice L. Odgers, and Winnie Yu.** 2017. "Persistence and Fadeout in the Impacts of Child and Adolescent Interventions". *Journal of Research on Educational Effectiveness* 10(1): 7–39.
- Bandura, Albert.** 1977. "Self-efficacy: Toward a Unifying Theory of Behavioral Change." *Psychology Review* 84: 191-215.
- Barlow, Jane, Hanna Bergman, Hege Kornør, Yinghui Wei, and Cathy Bennett.** 2016. "Group-based Parent Training Programmes for Improving Emotional and Behavioural Adjustment in Young Children". *Cochrane Database of Systematic Reviews* Issue 8: Art. No.: CD003680.
- Becker, Gary S.** 1965. "A Theory of the Allocation of Time." *The Economic Journal* 75: 493-517.
- Bennett, Cathy, Angela Underdown, and Jane Barlow.** 2013. "Massage for Promoting Mental and Physical Health in Typically Developing Infants Under the Age of Six Months. *Cochrane Database of Systematic Reviews* Issue 4. Art. No.: CD005038.

- Black, Sandra, Paul Devereux, and Kjell G. Salvanes.** 2009. "Like Father, Like Son? A Note on the Intergenerational Transmission of IQ Scores." *Economics Letters* 105 (1): 138–140.
- Blair, C., and Raver, C.C.** 2012. "Child Development in the Context of Adversity: Experiential Canalization of Brain and Behavior". *American Psychologist* 67(4), 309-318.
- Bowlby, John.** 1969. *Attachment and Loss, Vol. 1: Attachment*. New York: Basic Books.
- Bradley, Robert H., Bettye M. Caldwell, Stephen L. Rock, Craig T. Ramey, Kathryn E. Barnard, Carol Gray, ... Dale L. Johnson.** 1989. "Home Environment and Cognitive Development in the First 3 Years of Life: A collaborative Study Involving Six Sites and Three Ethnic Groups in North America." *Developmental Psychology* 25: 217–235.
- Bradley, Robert H., and Robert F. Corwyn.** 2002. "Socioeconomic Status and Child Development." *Annual Review of Psychology* 53 (1): 371-399.
- Britto, Pia R., Angelica Ponguta, Chin Reyes, and Romilla Karnati.** 2015. *A Systematic Review of Parenting Programmes for Young Children in Low- and Middle-Income Countries*. New York, NY: United Nations Children's Fund.
- Bronfenbrenner, Urie.** 1979. *The Ecology of Human Development: Experiments by Nature and design*. Cambridge, MA: Harvard University Press.
- Bus, A.G., and van Ijzendoorn, M.H.** 1999. "Phonological Awareness and Early Reading: A meta-analysis of Experimental Training Studies". *Journal of Educational Psychology* 91(3): 403–414.
- Campbell, Frances, Gabrielle Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan.** 2014. "Early Childhood Investments Substantially Boost Adult Health". *Science* 343, 1478-1485.
- Carneiro, Pedro, and James J. Heckman.** 2003. "Human Capital Policy, in Inequality in America: What Role for Human Capital Policies." *Inequality in America: What Role for Human Capital Policies*.
- Census.** 2006. Retrieved from <http://www.cso.ie/en/census/census2006reports/>.
- Chazan-Cohen, R., Raikes, H. H., and Vogel, C.** 2013. "Program Subgroups: Patterns of Impacts for Home-based, Center-based, and mixed-approach programs". *Monographs of the Society for Research in Child Development* 78(1), 93-109.
- Clements, D.H., Sarama, J., Wolfe, C.B., and Spitler, M.E.** 2013. "Longitudinal Evaluation of a Scale-up Model for Teaching Mathematics with Trajectories and

- Technologies: Persistence of Effects in the Third Year". *American Educational Research Journal* 50(4): 812–850.
- Côté, Sylvana, Massimiliano Orri, Richard E. Tremblay, and Orla Doyle.** 2018. "A Multi-Component Early Intervention Program and Trajectories of Behavior, Cognition, and Health". *Pediatrics* 141(5): e20173174.
- Council of Economic Advisors.** 2014. The Economics of Early Childhood Investments. [https://www.whitehouse.gov/sites/default/files/docs/early\\_childhood\\_report1.pdf](https://www.whitehouse.gov/sites/default/files/docs/early_childhood_report1.pdf)
- Cunha, Flávio, Irma Elo, and Jennifer Culhane.** 2013. "Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation." *NBER Working Paper* No. 19144.
- Cunha, Flávio, and James J. Heckman.** 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31-47.
- Cunha, Flávio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov.** 2006. "Interpreting the Evidence on Life Cycle Skill Formation" in *Handbook of the Economics of Education*, eds Hanushek EA, Welch F. North-Holland, Amsterdam, 697–812.
- Cunha, Flávio, James J. Heckman, and Susanne M. Schennach.** 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78: 883–931.
- De Graaf, I., Speetjens, P., Smit, F., de Wolff, M., and Tavecchio, L.** 2008. "Effectiveness of the Triple P Positive Parenting Program on Behavioral Problems in Children: A Meta-analysis". *Behavior Modification* 32(5): 714-735.
- Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall.** 2014. "Household Choices and Child Development." *The Review of Economic Studies* 81, 137–185.
- Del Bono, Emilia, Marco Francesconi, Yvonne Kelly, and Amanda Sacker.** 2016. "Early Maternal Time Investment and Early Child Outcomes." *Economic Journal* 126: 96-135.
- Deming, David.** 2009. "Early Childhood Intervention and Life-cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1 (3): 111–134.
- Derryberry, D., and Reed, M.A.** 2002. "Anxiety-related Attentional Biases and their Regulation by Attentional Control". *Journal of Abnormal Psychology* 111(2), 225.
- Diamond, A., and Lee, Kathleen.** 2011. "Interventions Shown to Aid Executive Function Development in Children 4 to 12 Years Old". *Science* 333(6045), 959-964.

- Dooley, Martin, and Jennifer Stewart.** 2007. "Family Income, Parenting Styles and Child Behavioural–Emotional Outcomes." *Health Economics* 16 (2): 145-162.
- Doyle, Orla.** 2013. "Breaking the Cycle of Deprivation: An Experimental Evaluation of an Early Childhood Intervention." *Journal of the Statistical and Social Inquiry Society of Ireland* Vol. XLI, 92-111.
- Doyle, Orla.** 2019. "The First 2000 Days and Children's Skills". Forthcoming *Journal of Political Economy*.
- Doyle, Orla, Liam Delaney, Christine O'Farrelly, Nick Fitzpatrick, and Michael Daly.** 2017b. "Can Early Intervention Policies Improve Well-being? Evidence from a randomized controlled trial." *PLoS ONE* 12 (1): e0169829.
- Doyle, Orla, Nick Fitzpatrick, Caroline Rawdon, and Judy Lovett.** 2015. "Early Intervention and Child Health: Evidence from a Dublin-based Trial". *Economics and Human Biology* 19: 224-245.
- Doyle, Orla, Colm Harmon, James J. Heckman, and Richard Tremblay.** 2009. "Investing in Early Human Development: Timing and Economic Efficiency". *Economics and Human Biology* 7 (1): 1-6.
- Doyle, Orla, Colm Harmon, James J. Heckman, Caitriona Logue, and Seong Hyeok Moon.** 2017a "Measuring Investment in Human Capital Formation: An Experimental Analysis of Early Life Outcomes." *Labour Economics* 45: 40-58.
- Doyle, Orla, and Claire Hickey.** 2013. "The Challenges of Contamination in Evaluations of Childhood Interventions." *Evaluation* 19: 180-191.
- Doyle, Orla, Louise McEntee, and Kelly A. McNamara.** 2012. "Skills, Capabilities and Inequality at School Entry in a Disadvantaged Community." *European Journal of the Psychology of Education* 27: 133-154.
- Doyle, Orla, Edel McGlanaghy, Eylin Palamaro Munsell, and Fionnuala M. McAuliffe.** 2014. "Home Based Educational Intervention to Improve Perinatal Outcomes for a Disadvantaged Community: A Randomised Control Trial". *European Journal of Obstetrics and Gynaecology* 180: 162-167.
- Doyle, Orla, and PFL Evaluation Team.** 2010. *Assessing the Impact of Preparing for Life: Baseline Report*. Report to Preparing for Life Programme. Atlantic Philanthropies & Department of Children and Youth Affairs.
- Doyle, Orla, and PFL Evaluation Team.** 2015. *Assessing the Impact of Preparing for Life at 48 Months*. Report to Preparing for Life Programme. Atlantic Philanthropies & Department of Children and Youth Affairs.

- Doyle, Orla, and PFL Evaluation Team.** 2016. *Final Report: Did Preparing for Life Improve Children's School Readiness*. Report to Preparing for Life Programme. Atlantic Philanthropies & Department of Children and Youth Affairs.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer.** 2008. *Using Randomization in Development Economics Research: A Toolkit. Handbook of Development Economics*, Elsevier.
- DuMont, K., Kirkland, K., Mitchell-Herzfeld, S., Ehrhard-Dietzel, S., Rodriguez, M. L., Lee, E., ... and Greene, R.** 2010. "A Randomized Trial of Healthy Families New York (HFNY): Does Home Visiting Prevent Child Maltreatment". *Rensselaer, NY: New York State Office of Children & Family Services and Albany, NY: The University of Albany, State University of New York*.
- Eble, Alex, Peter Boone, and Diana Elbourne.** 2016. "On Minimizing the Risk of Bias in Randomized Controlled Trials in Economics". *The World Bank Economic Review* 0: 1-12.
- Elliott, S.N., and Gresham, F.M.** 2008. *Social Skills Improvement System: Rating scales Manual*.
- Elliott, Colin D., Pauline Smith, and Kay McCulloch.** 2011. *British Ability Scales III*. Windsor, UK: NFER-Nelson.
- Evans, Gary W., and Rosenbaum, Jennifer.** 2008. "Self-regulation and the Income-Achievement Gap". *Early Childhood Research Quarterly*, 23(4), 504-514.
- Fantuzzo, J., Tighe, E., and Childs, S.** 2000. "Family Involvement Questionnaire: A Multivariate Assessment of Family Participation in Early Childhood Education". *Journal of Educational Psychology* 92(2): 367-376.
- Farah, Martha J., Laura Betancourt, David M. Shera, Jessica H. Savage, Joan M. Giannetta, Nancy L. Brodsky, ... and Hallam Hurt.** 2008. "Environmental Stimulation, Parental Nurturance and Cognitive Development in Humans." *Developmental Science* 11 (5): 793-801.
- Fernald, Anne, Virginia A. Marchman, and Adriana Weisleder.** 2013. "SES Differences in Language Processing Skill and Vocabulary are Evident at 18 Months." *Developmental Science* 16: 234-248.
- Filene, Jill H., Jennifer W. Kaminski, Linda Anne Valle, and Patrice Cachat.** 2013. "Components Associated with Home Visiting Program Outcomes: A Meta-analysis. *Pediatrics* 132(Supplement): S100-S109.



- Fiorini, Mario, and Michael Keane.** 2014. "How the Allocation of Children's Time Affects Cognitive and Non-cognitive Development." *Journal of Labor Economics* 32 (4): 787-836.
- Freedman, David, and David Lane.** 1983. "A Nonstochastic Interpretation of Reported Significance Levels." *Journal of Business Economics and Statistics* 1 (4), 292–298.
- Furlong, Mairead, Sinead McGilloway, Tracey Bywater, Judy Hutchings, Susan M. Smith, and Michael Donnelly.** 2012. "Behavioural and Cognitive-behavioural Group-based Parenting Programmes for Early-onset Conduct Problems in Children Aged 3 to 12 Years". *Cochrane Database of Systematic Reviews* Issue 2: Art. No.: CD008225.
- Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor.** 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344 (6187): 998-1001.
- Gomby, Deanna S.** 2005. *Home Visitation in 2005: Outcomes for Children and Parents* (Vol. 7). Invest in Kids Working Paper No. 7. Committee for Economic Development: Invest in Kids Working Group.
- Gomby, Deanna S., Patti L. Culross, and Richard E. Behrman.** 1999. "Home Visiting: Recent Program Evaluations: Analysis and Recommendations." *The Future of Children* 9 (1): 4.
- Good, Phillip.** 2005. *Permutation, Parametric and Bootstrap Tests of Hypotheses* (3<sup>rd</sup> ed.), New York: Springer.
- Goodman, Robert.** 1997. "The Strengths and Difficulties Questionnaire: A Research Note." *Journal of Child Psychology and Psychiatry* 38 (5): 581-586.
- Grantham-McGregor, Sally and Joanne A. Smith.** 2016. "Extending The Jamaican Early Childhood Development Intervention." *Journal of Applied Research on Children: Informing Policy for Children at Risk*: 7 (2), Article 4.
- Heckman, James J.** 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312 (5782): 1900-1902.
- Heckman, James J., Margaret L. Holland, Kevin K. Makino, Rodrigo Pinto, and Maria Rosales-Rueda,** 2017. "An Analysis of the Memphis Nurse-Family Partnership Program." NBER Working Paper No. 23610.
- Heckman, James J. and Tim Kautz.** 2014. "Fostering and Measuring Skills Interventions that Improve Character and Cognition." In *The GED Myth: Education, Achievement*

- Tests, and the Role of Character in American Life*, edited by James J. Heckman, John E. Humphries, and Tim Kautz, 341-430. Chicago, IL: University of Chicago Press.
- Heckman, James J, Seong H. Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz.** 2010. "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program." *Quantitative Economics* 1 (2): 1-46.
- Heckman, James J, and Stefano Mosso.** 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics* 6 (1): 689–733.
- Heckman James J, Rodrigo Pinto, and Peter A. Savelyev.** 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103: 2052–86.
- Kitzman, H. J., Olds, D. L., Cole, R. E., Hanks, C. A., Anson, E. A., Arcoleo, K. J., ... and Holmberg, J. R.** 2010. "Enduring Effects of Prenatal and Infancy Home Visiting by Nurses on Children: Follow-up of a Randomized Trial among Children at Age 12 years". *Archives of Pediatrics & Adolescent Medicine* 164(5): 412-418.
- Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff.** 2006. "Economic, Neurobiological, and Behavioral Perspectives on Building America's Future Workforce." *Proceedings of the National Academy of Science USA* 103 (27): 10155–10162.
- Layzer, Jean I., Barbara D. Goodson, Lawrence S. Bernstein L, and C. Price.** 2001. *National Evaluation of Family Support Programs. Volume A: The Meta-analysis. Final Report.* Cambridge, Mass: Abt Associates Inc.
- Liew, Jeffrey.** 2012. "Effortful Control, Executive Functions, and Education: Bringing Self-Regulatory and Social-Emotional Competencies to the Table". *Child Development Perspectives* 6(2): 105-111.
- Ludbrook, John, and Hugh Dudley.** 1998. "Why Permutation Tests are Superior to t and F Tests in Biomedical Research." *American Statistician* 52 (2): 127-132.
- Ludwig, Jens, and Deborah Phillips.** 2007. "The Benefits and Costs of Head Start." NBER Working Paper 12973.
- Martins, Lurdes, and Paula Veiga.** 2010. "Do Inequalities in Parents' Education Play an Important Role in PISA Students' Mathematics Achievement Test Score Disparities?". *Economics of Education Review* 29 (6): 1016-1033.

- Mewhort, Douglas J.K.** 2005. "A Comparison of the Randomization Test with the F test when Error is Skewed." *Behavior Research Methods* 37: 426–435.
- Michael, Robert T., and Gary S. Becker.** 1973. "On the New Theory of Consumer Behavior." *The Swedish Journal of Economics* 75 (4): 378-396.
- Millar, D.** 2018. *School Attendance Data from Primary and Post-Primary Schools, 2016/17, Analysis and Report to the Child and Family Agency*. Dublin: Educational Research Centre.
- Miller, Elizabeth. B., George Farkas, Deborah L. Vandell, and Greg J. Duncan.** 2014. "Do the Effects of Head Start Vary by Parental Preacademic Stimulation?" *Child Development* 85: 1385–1400.
- Miller, Sarah, Lisa K. Maguire, and Geraldine Macdonald.** 2011. "Home-based Child Development Interventions for Preschool Children from Socially Disadvantaged Families." *Cochrane Database of Systematic Reviews* 12, CD008131.
- Mingebach, Tanja, Inge Kamp-Becker, Hanna Christiansen, and Linda Weber.** 2018. "Meta-meta-analysis on the Effectiveness of Parent-based Interventions for the Treatment of Child Externalizing Behavior Problems". *PLoS One* 13(9): e0202855.
- Morawska, A., Mitchell, A. E., Burgess, S., and Fraser, J.** 2011. "Effects of Triple P parenting intervention on child health outcomes for childhood asthma and eczema: Randomised controlled trial". *Behaviour Research and Therapy* 83, 35-44.
- Neuenschwander, R., Röthlisberger, M., Cimeli, P., and Roebbers, C.M.** 2012. "How do Different Aspects of Self-regulation Predict Successful Adaptation to School?" *Journal of Experimental Child Psychology* 113(3): 353-371.
- Nisbett, Richard E., Joshua Aronson, Clancy Blair, William Dickens, James Flynn, Diane F. Halpern, and Eric Turkheimer.** 2012. "Intelligence: New Findings and Theoretical Developments." *American Psychologist* 67(2): 130–159.
- O’Sullivan, Aifric, Nick Fitzpatrick, and Orla Doyle.** 2017. "Effects of Dietary Recommendations During Early Childhood on Cognitive Functioning: A Randomized Controlled Trial." *Public Health Nutrition* 20 (1): 154-164.
- OECD.** 2016. *Enhancing Child Well-being to Promote Inclusive Growth*. OECD Publishing, Paris.
- Office of the Minister for Children and Youth Affairs.** 2008. *Forum on Prevention and Early Intervention for Children and Youth*. The Stationery Office, Dublin.
- Olds, David. L.** 2002. "Prenatal and Infancy Home Visiting by Nurses: From Randomized Trials Community Replication". *Prevention Science* 3: 153–172.

- Olds, David L., Harriet Kitzman, Robert Cole, JoAnn Robinson, Kimberly Sidora, Dennis W. Luckey, Charles R. Henderson Jr, et al.** 2004. "Effects of Nurse Home-visiting on Maternal Life Course and Child Development: Age 6 Follow-up Results of a Randomized Trial." *Pediatrics* 114 (6): 1550-1559.
- Olds, D. L., Kitzman, H., Hanks, C., Cole, R., Anson, E., Sidora-Arcoleo, K., ... and Stevenson, A. J.** 2007. "Effects of nurse home visiting on maternal and child functioning: age-9 follow-up of a randomized trial". *Pediatrics* 120(4), e832-e845.
- Olds, David. L., Charles R. Henderson Jr, and Harriet Kitzman.** 1994. "Does Prenatal and Infancy Nurse Home Visitation have Enduring Effects on Qualities of Parental Caregiving and Child Health at 25 to 50 Months of Life?" *Pediatrics* 93 (1): 89-98.
- Olds, David L., Charles R. Henderson Jr, Harriet J. Kitzman, John J. Eckenrode, Robert E. Cole, and Robert C. Tatelbaum.** 1999. "Prenatal and Infancy Home Visitation by Nurses: Recent Findings." *The Future of Our Children* 9 (1): 44–65.
- Olds, D. L., Holmberg, J. R., Donelan-McCall, N., Luckey, D. W., Knudtson, M. D., and Robinson, J.** 2014. "Effects of home visits by paraprofessionals and by nurses on children: follow-up of a randomized trial at ages 6 and 9 years". *JAMA Pediatrics* 168(2), 114-121.
- Peacock, Shelley, Stephanie Konrad, Erin Watson, Darren Nickel, and Nazeem Muhajarine.** 2013. "Effectiveness of Home Visiting Programs on Child Outcomes: A Systematic Review." *BMC Public Health* 13 (1): 17.
- Rayce, Signe B., Ida S. Rasmussen, Sihy K. Klest, Joshua Patras, and Maiken Pontoppidan.** 2017. "Effects of Parenting Interventions for At-risk Parents with Infants: A Systematic Review and Meta-analyses". *BMJ Open* e015707.
- Robins, James M., Andrea Rotnitzky, and Lue P. Zhao.** 1994. "Estimation of Regression Coefficients when Some Regressors are Not Always Observed." *Journal of the American Statistical Association* 89 (427): 846-866.
- Roggman Lori A., Gina A. Cook, Carla A. Peterson, and Helen H. Raikes.** 2008. "Who Drops Out of Early Head Start Home Visiting Programs?" *Early Education And Development* 19 (4).
- Romano, Joseph P., and Michael Wolf.** 2005. "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing." *Journal of the American Statistical Association* 100 (469): 94-108.

- Sanders, Matthew R.** 2012. "Development, Evaluation, and Multinational Dissemination of the Triple P-Positive Parenting Program." *Annual Review of Clinical Psychology* 8(1): 345–379.
- Sanders, Matthews R., James N. Kirby, Cassandra L Tellegen, and Jamin J. Day.** 2014. "The Triple P-Positive Parenting Program: A Systematic Review and Meta-analysis of a Multi-level System of Parenting Support." *Clinical Psychology Review* 34: 337-357.
- Sanders, Matthews R., Carol Markie-Dadds, and Karen Turner.** 2003. "Theoretical, Scientific and Clinical Foundations of the Triple P-Positive Parenting Program: A Population Approach to the Promotion of Parenting Competence." *Parenting Research and Practice Monograph* 1: 1-21.
- Sandner, Malte, and Tanja Jungmann.** 2017. "Gender-specific Effects of Early Childhood Intervention: Evidence from a Randomized Controlled Trial." *Labour Economics* 45: 59-78.
- Sarganas, G., Schaffrath Rosario, A., and Neuhauser, H.** 2017. "Resting Heart Rate Percentiles and Associated Factors in Children and Adolescents". *The Journal of Pediatrics* 187: 174-181
- Schwarz, Gideon.** 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6: 461-464.
- Sidora-Arcoleo, K., Anson, E., Lorber, M., Cole, R., Olds, D., and Kitzman, H.** 2010. "Differential Effects of a Nurse Home-visiting Intervention on Physically Aggressive Behavior in Children". *Journal of Pediatric Nursing* 25(1): 35-45.
- Sitnick, S. L., Shaw, D. S., Gill, A., Dishion, T., Winter, C., Waller, R., ... and Wilson, M.** 2015. "Parenting and the Family Check-Up: Changes in Observed Parent-Child Interaction Following Early Childhood Intervention". *Journal of Clinical Child & Adolescent Psychology* 44(6): 970-984.
- Smith, J., Dishion, T., Shaw, D., and Wilson M.** 2014. "Coercive Family Process and Early-Onset Conduct Problems from Age 2 to School Entry." *Development and Psychopathology* 26: 917-932.
- Smith, T.M., Cobb, P., Farran, D.C., Cordray, D.S., and Munter, C.** 2013. "Evaluating Math Recovery: Assessing the Causal Impact of a Diagnostic Tutoring Program on Student Achievement". *American Educational Research Journal* 50(2): 1–32.
- Sweet, Monica A., and Mark I. Appelbaum.** 2004. "Is Home Visiting an Effective Strategy? A Meta-Analytic Review of Home Visiting Programs for Families with Young Children." *Child Development* 75: 1435-1456.

- Thompson, Ross A., and Charles A. Nelson.** 2001. "Developmental Science and The Media: Early Brain Development." *American Psychologist* 56 (1): 5–15.
- Todd, Petra E., and Kenneth I. Wolpin.** 2007. "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps." *Journal of Human Capital* 1 (1): 91-136.
- Washington State Institute for Public Policy.** 2016. *Benefit-Cost Results – Public Health and Prevention*. <http://www.wsipp.wa.gov/BenefitCost>
- Weaver, Ian C.G., Nadia Cervoni, Frances A. Champagne, Ana C. D'Alessio, Shakti Sharma, Jonathan R. Seckl, Sergiy Dymov, Moshe Szyf, and Michael J. Meaney.** 2004. "Epigenetic Programming by Maternal Behavior" *Nature Neuroscience* 7: 847–854.
- Wesley, Patricia W., Virginia Buysse, and Sabrina Tyndall.** 1997. "Family and Professional Perspectives on Early Intervention: An Exploration Using Focus Groups." *Topics in Early Childhood Special Education* 17 (4): 435-456.
- Williams, James, Sheila Greene, Sinéad McNally, Aisling Murray, and Amanda Quail.** 2010. *Growing up in Ireland National Longitudinal Study of Children. The Infants and their Families*. The Stationery Office, Ireland.
- Wilson, P., Rush, R., Hussey, S., Puckering, C., Sim, F., Allely, C. S., ... and Gillberg, C.** 2012. "How Evidence-based is an 'Evidence-based Parenting Program'? A PRISMA Systematic Review and Meta-analysis of Triple P". *BMC medicine* 10(1): 130.
- Yap, Marie B.H., Amy J. Morgan, Kathryn Cairns, Anthony F. Jorm, Sarah E. Hetrick and Sally Merry.** 2016. "Parents in Prevention: A Meta-analysis of Randomized Controlled Trials of Parenting Interventions to Prevent Internalizing Problems in Children from Birth to Age 18". *Clinical Psychology Review* 50: 138–158.
- Zelazo, P.D., and Bauer, P.J.** 2013. "National Institutes of Health Toolbox—Cognitive Function Battery (NIH Toolbox CFB): Validation for Children Between 3 and 15 Years". *Monographs of the Society for Research in Child Development* 78(4).

## Appendix A Measure Description

### *Parent-reported Instruments*

#### Socio-emotional Development

The *Brief Problems Monitor* (BPM; Achenbach, McConaughy, Ivanovaa, & Rescorla, 2011) is a parent report measure for children aged 6-18 years to monitor children's functioning and responses to interventions. The BPM is based on items from the Child Behaviour Checklist (CBCL), Teachers Report Form (TRF) and Youth Self Report (YSR). The measure consists of 19 items with the response options *not true*, *somewhat true*, or *very true*. These were scored as zero, 1 and 2 respectively. The measure yields scores across three subscales: internalising (6 items;  $\alpha = 0.76$ ), attention (6 items;  $\alpha = 0.89$ ) and externalising (7 items;  $\alpha = 0.80$ ) problems. The scores for each of the three problems subscales were summed to create a Total Problems score. Scores were then converted to standard scores based on the child's age and gender, and binary indicators of concerning problem behaviour were created based on standard scores exceeding 60.

The *Strengths and Difficulties Questionnaire* (SDQ; Goodman, 1997) is a 25-item questionnaire assessing behaviours, emotions, and relationships of 4- to 16-year-olds. The questionnaire covers five dimensions: conduct problems, emotional symptoms, hyperactivity, peer problems, and pro-social behaviour. The 5-item Peer Problems ( $\alpha = 0.62$ ) and 5-item Pro-Social ( $\alpha = 0.72$ ) subscales were used in this report. Items were scored 0 for *not true*, 1 for *somewhat true*, and 2 for *certainly true*. Two items from the Peer Problems subscale were reverse scored. The five items for each subscale were summed giving a total score of 0 to 10 for each subscale ( $\alpha = 0.90$ ). Cutoff scores were also created to indicate scores that were of clinical concern.

Binary indicators for whether or not the child has a diagnosis of Autism Spectrum Disorder/Attention Deficit Hyperactivity Disorder (ASD-ADHD) or whether or not the child has a diagnosed *learning difficulty* were created from parents responses to a question asking them if their child has a developmental delay/disorder or any diagnosed learning or physical disability, and if so to list the diagnosis.

#### Parenting

The *Family Involvement Questionnaire* (FIQ) home-based subscale (Fantuzzo, Tighe & Childs, 2000) measures family members' involvement in children's education. This instrument is a multidimensional rating scale that asks primary care providers of young children (i.e., parents, other family members, or legal guardians) to indicate the nature and extent of their involvement in their children's early educational experiences. The home-based involvement dimension was used in this report. This dimension consists of 13 items ( $\alpha = 0.80$ ) with the response options *rarely*, *sometimes*, *often* and *always*. These were scored as 1, 2, 3 and 4 respectively and summed to create a home-based involvement score ranging from 13 to 52.

The *Attentional Control Scale* (ACS; Derryberry & Reed 2002) is a 20-item self-report questionnaire that assesses individual differences in attentional control. The 9-item subscale of attentional focusing ( $\alpha = 0.64$ ) was used in this report. Items were scored on a 4-point scale with the response options *almost never*, *sometimes*, *often* or *always*. These were scored as 4, 3, 2 and 1 respectively. Scores were summed to create an attentional focusing variable ranging from 9 to 36, with higher scores indicating better attentional control.

Parents were also asked about their child's *internet access*. They were asked to report whether their child has access to the internet (e.g. through a phone, computer, tablet, PlayStation, or internet enabled TV) and to select from a list of rules and restrictions any strategies they use to restrict the content viewed or time spent by their child on the internet. Parents were also asked whether their child was supervised *Never*, *Sometimes*, or *Always* when they accessed the internet. Answers to this combination of variables were used to create binary indicators for whether or not the parent had rules in place to restrict internet access, restrictions in place to restrict internet access, and whether the child was always supervised when accessing the internet.

### Child Health

Parents were asked several questions about their *child's health* in the last 12 months. They were asked to rate their child's health in the last 12 months on a 5-point scale ranging from *excellent* to *poor*. Responses to this question were dichotomised to indicate whether the child had good health (*good*, *very good*, *excellent*) or not (*poor*, *fair*). Parents were also asked if their child had any of 11 ongoing chronic illnesses presented on a list. Answers to this question were used to create binary indicators of whether the child was *diagnosed with asthma* or *had an ongoing diagnosed chronic illness other than asthma*.

A series of questions asking parent to report on medical and dental visits were used to create seven binary variables representing whether or not the child had attended a *GP*, the *Emergency Department*, or a routine dental check-up, had attended a health facility due to an *accident*, *chest infection*, or *wheezing or asthma*, and whether the child had *received dental treatment*. Reported bedtime and wake time on a typical day were used to calculate *number of hours sleep on average night* and parents were also asked to indicate whether their child had any, of a list of 10, sleep problems.

Finally, parents responded to food frequency questions asking them to indicate the frequency that their child consumed food from each of the six main food groups: grains, dairy, protein, fruits, vegetables, and other foods (such as sugars and fats, sweets, crisps, etc). These questions were a collapsed version of the 52-individual item Lifeways Cross-Generation Cohort Study (Shrivastava *et al.*, 2013) and reflected national health eating guidelines and the food pyramid groups. Parents responded to each food group on a 9-point scale ranging from *Never* to *More than six times per day*. Four binary variables were created indicating whether the child consumed the recommended daily number of servings according to the Food and Nutrition Guidelines for Pre-school Services (Department of Health and Children, 2004) and endorsed by the PFL Tip Sheets.

### *Child-reported/assessed Instruments*

#### Cognitive development

Children's cognitive ability was assessed using the School Age *British Ability Scales III* (Elliot & Smith, 2011). The BAS III School Age battery was designed as an assessment of children's abilities in clinical, educational, and research settings for children and students aged from 5 years 0 months to 17 years 11 months. The BAS III consists of six subscales: word definitions, verbal similarities, matrices, quantitative reasoning, recognition of designs, and pattern construction. These sub-scales yield an overall score reflecting general cognitive ability (General Conceptual Ability, GCA), as well as three cluster scores for Verbal Ability, Non-Verbal Ability, and Spatial Ability. The GCA score assesses overall cognitive ability



such as thinking logically, making decisions, and learning. The Spatial Ability score assesses problem solving, spatial visualisation, and short-term visual memory. The Nonverbal Reasoning score assesses inductive reasoning. The Verbal ability score assesses children's verbal reasoning, verbal knowledge, and expressive language. Age-based *T* scores are calculated for each domain that are standardised to have a mean of 100 and a standard deviation of 15, as well as cutoff scores indicating whether the child scores above or below average for the GCA and cluster scores.

The *National Institutes of Health Toolbox for Assessment of Neurological and Behavioral Function Cognition Battery* (NIH Toolbox; Zelazo & Bauer, 2013) is a battery of neuropsychological tests administered using an iPad that assess cognitive ability and executive functions for ages 7+. Three NIH Toolbox measures were used in this report to assess children's executive functions. Executive functions are higher order meta-cognitive processes involved in concentration, reasoning, problem solving, and planning<sup>23</sup>. The *Flanker Task* was used to assess inhibitory control. Children were asked to indicate the left-right orientation of a centrally presented stimulus arrow surrounded by congruent or incongruent stimuli arrows. The *Dimensional Change Card Sort* task was used to assess attention flexibility. Children were asked to match test pictures to a target picture that varied along two dimensions, colour and shape. Finally, the *List Sorting* task was used to assess working memory. For the first part of the task, children were presented with a series of stimuli (either food or animals) on screen and orally and asked to order the list of items from smallest to largest. In the second part of the task children were similarly presented with a series of stimuli and asked to recall the food items in size order followed by the animals in size order from smallest to largest. Age-corrected scores for each of the three NIH toolbox measures was then standardised and summed to create a composite indicator of executive functions.

### Socio-emotional development

The *Social Skills Improvement System Rating Scales* (SSIS-RS; Elliot & Gresham, 2008) is a child report measure ( $\alpha = 0.87$ ) that assesses children's social skills and problem behaviour with two subscales by those names. The current study used the behaviour problems subscale, which consists of 29-items measuring: externalising (x items;  $\alpha = 0.79$ ), internalising (x items;  $\alpha = 0.84$ ), bullying (x items;  $\alpha = 0.44$ ), and hyperactivity/inattention (x items;  $\alpha = 0.70$ ) for children aged 3 to 18 years. Children indicated how true a statement about each social skill and problem behaviour was for them using a 4-point scale of *not true*, *a little true*, *a lot true*, and *very true*. These were scored as 0, 1, 2, 3 respectively. The relevant items were summed to create subscale scores. A total score was also computed and converted to a standard score using the scale norms for age and gender. Cutoff scores were also created for each of the subscales and the standardised total score to indicate whether or not the child scored above average for each score.

### Child Health

---

<sup>23</sup> Executive functions are comprised of three core abilities. 1. Inhibitory control which involves the ability to override impulse responses and ideally replace them with a more adaptive behaviour. For example, delaying eating a treat to receive a reward. 2. Attention flexibility which involves being able to deliberately focus and maintain attention or to divert attention to a new task if required. For example, blocking out distraction to complete a task or shifting attention from one task to another. And 3, working memory which involves the ability to retain and manipulate information over brief periods of time. Working memory is central to remembering instructions or rules or pieces of information that are necessary to solve a problem.

Children's height in centimetres and weight in kilograms was recorded by the researchers either at the recruitment meeting or during the direct assessments in schools. Height and weight were used to calculate the child's *Body Mass Index* (BMI). Indicators of whether the child was *Overweight/obese* or *Obese* were created based on the Centers for Disease Control and Prevention (CDC) data table of BMI for age charts (CDC, 2001). Children in the 85<sup>th</sup> percentile or above were categorised as overweight/obese and children in the 95<sup>th</sup> percentile or above were categorised as obese.

Children's resting pulse and blood pressure (systolic and diastolic) were recorded using a Kinetik Advanced Wrist Blood Pressure Monitor. Measurements were taken twice and the average value was used to create binary indicators of whether the child's pulse (Sarganas *et al.*, 2017), systolic blood pressure, and diastolic blood pressure (National High Blood Pressure Education Programme, 2004) were elevated, i.e. at or above the 95<sup>th</sup> percentile, according to their age and gender. Pulse and blood pressure are important vital signs and elevated recordings are indicators of cardiovascular risk (Sarganas *et al.*, 2017).

### *School Information*

#### Academic Scores

All primary schools in Ireland are required to administer standardised achievement tests in English reading and Mathematics at the end of second class when children are approximately 7/8 years of age, some school also continue to administer standardised achievement tests at the end of each school class. There are two published sets of standardised achievement tests normed for the Irish population that primary schools can use: Drumcondra Tests for Reading and Mathematics, produced by the Educational Research Centre or the Micra-T for English Reading and Sigma-T for Mathematics produced by Folens. These norm-referenced tests provide a standard score (mean = 100, SD = 15), with higher scores indicative of better performance. Standard scores are also converted to STen scores ranging from 1 to 10 indicating the child's approximate position with respect to the reference population. Standard and STen scores for participating children were obtained from the child's school. Scores for 2<sup>nd</sup> and 3<sup>rd</sup> class, where available (equivalent to US 2<sup>nd</sup> and 3<sup>rd</sup> grade and UK Year 3 and 4) were used in this report. STen scores were converted to indicators of above and below average performance based on National Council for Curriculum and Assessment (NCCA) STen score descriptors.

#### Absenteeism and School Resources

Information on *daily school attendance* was gathered from participating children's schools in the form of days attended and days absent for the previous school year (2017/2018) and up to the point of data collection for the year of data collection (2018/2019). This information was used to calculate the proportion of days attended in the school year for the current and previous school year.

Children with special education needs in Primary Schools in Ireland receive educational support to assist their integration into ordinary mainstream schools or to support their education in specialised schools. These supports normally take the form of in/out of class resource/learning support, or Special Needs Assistant depending on the child's assessed level of need. Schools were asked to indicate, yes or no, whether each child received education supports in school in the form of in-class support, out of class support, Special Needs Assistant, or other educational supports.

## Appendix References

National Council for Curriculum and Assessment (NCCA) (n.d.). Understanding STen scores: information for parents, Retrieved from [www.ncca.ie/en/primary/resources-for-parents](http://www.ncca.ie/en/primary/resources-for-parents)

Shrivastava, A., Murrin, C., Sweeney, M., Heavey, P., Kelleher, C., & Lifeways Cohort Study Steering Group (2013). Familial intergenerational and maternal aggregation patterns in nutrient intakes in the Lifeways Cross-Generation Cohort Study. *Public Health and Nutrition*, 16(8), 1476-1486.

Department of Health and Children (2004). *Food and Nutrition Guidelines for Pre-School Services*. Dublin: Department of Health and Children.

CDC, Centre for Disease Control and Prevention (2001). *Data Table of BMI-for-age Charts*. Retrieved from [https://www.cdc.gov/growthcharts/html\\_charts/bmiagerev.htm](https://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm)

Sarganas, G., Schaffrath Rosario, A., & Neuhauser, H. (2017). Resting heart rate%iles and associated factors in children and adolescents. *The Journal of Pediatrics*, 187, 174-181. <https://doi.org/10.1016/j.jpeds.2017.05.021>

National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents (2004). *The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents*, *Pediatrics*, 114(Suppl 4<sup>th</sup> Report), 555-576.

## Appendix B Robustness Tests

**Table B1** Comparison of high and low treatment groups: School outcomes conditional on DEIS status and type of test

	N (HIGH/LOW)	$M_{\text{HIGH}}$ (SD)	$M_{\text{LOW}}$ (SD)	$p^1$	$p^2$	ES
<i>Academic Standardised Scores</i>						
2 <sup>nd</sup> class reading standardised score	118 (66/52)	99.55 (15.10)	94.93 (12.63)	<b>0.044</b>	<b>0.044</b>	0.33
3 <sup>rd</sup> class reading standardised score	69 (40/29)	97.51 (12.09)	89.67 (9.30)	<b>0.001</b>	<b>0.011</b>	0.73
2 <sup>nd</sup> class maths standardised score	117 (66/51)	97.42 (14.62)	89.73 (12.89)	<b>0.003</b>	<b>0.011</b>	0.56
3 <sup>rd</sup> class maths standardised score	69 (40/29)	94.47 (14.19)	88.05 (12.96)	<b>0.058</b>	<b>0.081</b>	0.47
<i>Academic Above the Norm Cutoff Scores</i>						
2 <sup>nd</sup> class above average reading score %	118 (66/52)	0.26 (0.44)	0.19 (0.39)	0.127	0.127	0.19
3 <sup>rd</sup> class above average reading score %	69 (40/29)	0.26 (0.44)	0.05 (0.22)	<b>0.006</b>	<b>0.014</b>	0.65
2 <sup>nd</sup> class above average maths score %	117 (66/51)	0.22 (0.42)	0.05 (0.21)	<b>0.000</b>	<b>0.006</b>	0.55
3 <sup>rd</sup> class above average maths score %	69 (40/29)	0.15 (0.36)	0.05 (0.22)	<b>0.059</b>	0.119	0.35
<i>Academic Below the Norm Cutoff Scores</i>						
2 <sup>nd</sup> class below average reading score %	118 (66/52)	0.31 (0.47)	0.47 (0.50)	<b>0.044</b>	<b>0.044</b>	0.34
3 <sup>rd</sup> class below average reading score %	69 (40/29)	0.35 (0.48)	0.62 (0.49)	<b>0.033</b>	<b>0.047</b>	0.56
2 <sup>nd</sup> class below average maths score %	117 (66/51)	0.34 (0.48)	0.55 (0.50)	<b>0.030</b>	<b>0.075</b>	0.43
3 <sup>rd</sup> class below average maths score %	69 (40/29)	0.44 (0.50)	0.72 (0.46)	<b>0.035</b>	<b>0.091</b>	0.57
<i>Absenteeism &amp; School Resources</i>						
Proportion of days present in last school year %	100 (55/45)	0.93 (0.05)	0.93 (0.06)	0.663	0.663	0.11
Proportion of days present in current school year %	117 (66/51)	0.94 (0.04)	0.94 (0.06)	0.626	0.746	0.04
In class supports %	118 (66/52)	0.39 (0.49)	0.53 (0.50)	<b>0.092</b>	0.400	0.28
Out of class supports %	118 (66/52)	0.24 (0.43)	0.37 (0.49)	0.121	0.359	0.28
SNA supports %	118 (66/52)	0.04 (0.19)	0.03 (0.17)	0.279	0.709	0.04
Other supports %	118 (66/52)	0.09 (0.29)	0.14 (0.35)	0.158	0.556	0.16

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional  $p$ -value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional  $p$ -value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the standard deviation of the low treatment group. The conditional set includes gender, DEIS status, and type of test conducted.

**Table B2** Comparison of high and low treatment groups: Child cognitive outcomes conditional on baseline differences

	N (HIGH/LOW)	$M_{\text{HIGH}}$ (SD)	$M_{\text{LOW}}$ (SD)	$p^1$	$p^2$	ES
<i>BAS Composite Scores</i>						
General Conceptual Ability	110 (68/42)	88.18 (11.95)	79.65 (12.85)	<b>0.008</b>	<b>0.022</b>	0.61
Spatial Ability	111 (68/43)	94.39 (14.22)	86.72 (17.08)	<b>0.059</b>	0.129	0.42
Non-Verbal Ability	111 (68/43)	84.53 (11.76)	76.47 (10.22)	<b>0.000</b>	<b>0.004</b>	0.73
Verbal Ability	110 (68/42)	92.19 (11.81)	86.19 (13.79)	<b>0.054</b>	0.114	0.39
<i>BAS Above the Norm %</i>						
General Conceptual Ability	110 (68/42)	0.02 (0.14)	0.00 (0.00)	0.190	0.473	0.31
Spatial Ability	111 (68/43)	0.11 (0.31)	0.12 (0.33)	0.521	0.789	0.03
Non-Verbal Ability	111 (68/43)	0.03 (0.18)	0.00 (0.00)	0.241	0.460	0.25
Verbal Ability	110 (68/42)	0.05 (0.22)	0.08 (0.28)	0.845	0.845	0.19
<i>BAS Below the Norm %</i>						
General Conceptual Ability	110 (68/42)	0.57 (0.50)	0.74 (0.44)	<b>0.041</b>	<b>0.041</b>	0.38
Spatial Ability	111 (68/43)	0.33 (0.47)	0.68 (0.47)	<b>0.008</b>	<b>0.015</b>	0.68
Non-Verbal Ability	111 (68/43)	0.71 (0.46)	0.88 (0.33)	<b>0.006</b>	<b>0.023</b>	0.52
Verbal Ability	110 (68/42)	0.33 (0.47)	0.66 (0.48)	<b>0.024</b>	<b>0.030</b>	0.54
<i>NIH Toolbox Executive Functioning</i>						
Flanker Task - Inhibitory control	110 (68/42)	97.89 (16.77)	90.11 (11.90)	<b>0.061</b>	<b>0.061</b>	0.39
Dimensional Change Card Sort Task - Attention flexibility	109 (68/41)	102.26 (21.88)	91.76 (12.57)	<b>0.046</b>	<b>0.065</b>	0.44
List Sorting Task - Working memory	107 (68/39)	96.05 (13.46)	89.51 (9.46)	<b>0.011</b>	<b>0.069</b>	0.49
<i>Other</i>						
Composite Executive Function score	107 (68/39)	0.38 (1.03)	0.18 (0.69)	<b>0.033</b>	<b>0.033</b>	0.49

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional  $p$ -value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional  $p$ -value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the pooled standard deviation. The conditional set includes gender maternal knowledge of child development, parenting self-efficacy, maternal attachment, and maternal consideration of future consequences.

**Table B3** Comparison of high and low treatment groups: School outcomes conditional on baseline differences

	N (HIGH/LOW)	$M_{\text{HIGH}}$ (SD)	$M_{\text{LOW}}$ (SD)	$p^1$	$p^2$	ES
<i>Academic Standardised Scores</i>						
2 <sup>nd</sup> class reading standardised score	112 (65/47)	99.89 (15.06)	95.10 (13.08)	<b>0.061</b>	<b>0.061</b>	0.32
3 <sup>rd</sup> class reading standardised score	66 (41/25)	97.51 (11.97)	89.13 (9.58)	<b>0.009</b>	<b>0.028</b>	0.65
2 <sup>nd</sup> class maths standardised score	111 (65/46)	97.49 (14.77)	89.99 (13.36)	<b>0.011</b>	<b>0.026</b>	0.50
3 <sup>rd</sup> class maths standardised score	66 (41/25)	94.42 (14.05)	86.22 (12.98)	<b>0.037</b>	<b>0.054</b>	0.53
<i>Academic Above the Norm Cutoff Scores</i>						
2 <sup>nd</sup> class above average reading score %	112 (65/47)	0.27 (0.45)	0.21 (0.41)	0.283	0.283	0.12
3 <sup>rd</sup> class above average reading score %	66 (41/25)	0.26 (0.44)	0.05 (0.23)	<b>0.014</b>	<b>0.052</b>	0.52
2 <sup>nd</sup> class above average maths score %	111 (65/46)	0.22 (0.42)	0.05 (0.22)	<b>0.005</b>	<b>0.026</b>	0.49
3 <sup>rd</sup> class above average maths score %	66 (41/25)	0.14 (0.36)	0.05 (0.23)	0.128	0.277	0.27
<i>Academic Below the Norm Cutoff Scores</i>						
2 <sup>nd</sup> class below average reading score %	112 (65/47)	0.30 (0.46)	0.46 (0.50)	<b>0.067</b>	<b>0.067</b>	0.30
3 <sup>rd</sup> class below average reading score %	66 (41/25)	0.34 (0.48)	0.63 (0.49)	<b>0.020</b>	<b>0.049</b>	0.61
2 <sup>nd</sup> class below average maths score %	111 (65/46)	0.35 (0.48)	0.52 (0.50)	<b>0.063</b>	<b>0.099</b>	0.34
3 <sup>rd</sup> class below average maths score %	66 (41/25)	0.45 (0.50)	0.80 (0.41)	<b>0.015</b>	<b>0.044</b>	0.65
<i>Absenteeism &amp; School Resources</i>						
Proportion of days present in last school year %	99 (59/05)	0.92 (0.05)	0.94 (0.06)	0.708	0.708	0.15
Proportion of days present in current school year %	116 (69/47)	0.94 (0.04)	0.94 (0.06)	0.401	0.579	0.10
In class supports %	117 (69/48)	0.39 (0.49)	0.56 (0.50)	<b>0.090</b>	0.276	0.31
Out of class supports %	117 (69/48))	0.24 (0.43)	0.40 (0.49)	0.076	0.273	0.34
SNA supports %	117 (69/48)	0.07 (0.25)	0.03 (0.18)	0.618	0.871	0.13
Other supports %	117 (69/48)	0.09 (0.29)	0.13 (0.34)	0.237	0.625	0.11

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional  $p$ -value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional  $p$ -value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the standard deviation of the low treatment group. The conditional set includes gender maternal knowledge of child development, parenting self-efficacy, maternal attachment, and maternal consideration of future consequences.

**Table B4** Comparison of high and low treatment groups: Child socio-emotional & health outcomes conditional on baseline differences

	N (HIGH/LOW)	$M_{\text{HIGH}}$ (SD)	$M_{\text{LOW}}$ (SD)	$p^1$	$p^2$	ES
<i>Child Socio-emotional Skills</i>						
SSIS Internalising Problems	111 (68/43)	7.54 (5.83)	8.36 (5.58)	0.304	0.552	0.09
SSIS Externalising Problems	111 (68/43)	6.43 (3.92)	4.89 (5.61)	0.830	0.887	0.25
SSIS Bullying	111 (68/43)	1.11 (1.24)	0.60 (1.68)	0.818	0.818	0.27
SSIS Hyperactivity/Inattention	111 (68/43)	6.45 (3.48)	6.17 (4.40)	0.585	0.789	0.08
<i>Child Socio-emotional Skills Cutoff</i>						
SSIS Internalising Problems %	111 (68/43)	0.11 (0.31)	0.11 (0.32)	0.376	0.376	0.04
SSIS Externalising Problems %	111 (68/43)	0.08 (0.27)	0.09 (0.29)	0.467	0.559	0.04
SSIS Bullying %	111 (68/43)	0.01 (0.10)	0.04 (0.21)	0.178	0.456	0.17
SSIS Hyperactivity/Inattention %	111 (68/43)	0.11 (0.31)	0.13 (0.34)	0.346	0.521	0.09
<i>Child Health</i>						
% Overweight/obese	110 (66/44)	0.31 (0.46)	0.29 (0.46)	0.373	0.769	0.02
Elevated heart rate	107 (66/41)	0.07 (0.25)	0.04 (0.20)	0.596	0.596	0.08
Elevated systolic blood pressure	111 (68/43)	0.12 (0.33)	0.11 (0.32)	0.627	0.829	0.06
Elevated diastolic blood pressure	107 (66/41)	0.11 (0.32)	0.12 (0.33)	0.312	0.654	0.10
<i>Other</i>						
SISS Problem Behaviours Standardised Total Score	111 (68/43)	96.73 (10.03)	96.63 (14.85)	0.506	0.506	0.02
SISS Problem behaviours total score - ~ 85%ile	111 (68/43)	0.04 (0.20)	0.12 (0.32)	<b>0.072</b>	0.205	0.32
BMI	110 (66/44)	18.26 (3.48)	18.53 (3.74)	0.288	0.478	0.10
% Obese	110 (66/44)	0.15 (0.36)	0.20 (0.41)	0.182	0.366	0.18

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional  $p$ -value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional  $p$ -value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the pooled standard deviation. The conditional set includes gender maternal knowledge of child development, parenting self-efficacy, maternal attachment, and maternal consideration of future consequences.

**Table B5** Comparison of high and low treatment groups: Parent reported child socio-emotional & parenting outcomes conditional on baseline differences

	N (HIGH/LOW)	M <sub>HIGH</sub> (SD)	M <sub>LOW</sub> (SD)	p <sup>1</sup>	p <sup>2</sup>	ES
<i>Child Socio-emotional Standardised Scores</i>						
SDQ Peer problems	107 (64/43)	1.59 (1.61)	1.93 (1.49)	0.115	0.385	0.20
SDQ Prosocial behaviour	107 (64/43)	8.85 (1.71)	8.80 (1.29)	0.465	0.456	0.05
BPM Internalising problems	106 (63/43)	57.54 (6.34)	57.23 (6.14)	0.452	0.663	0.04
BPM Externalising problems	106 (63/43)	53.79 (5.74)	54.70 (6.08)	0.261	0.398	0.17
BPM Attention problems	106 (63/43)	53.89 (7.46)	56.47 (7.05)	0.133	0.370	0.23
<i>Child Socio-emotional Cutoff Scores</i>						
SDQ Peer problems	107 (64/43)	0.13 (0.34)	0.14 (0.35)	0.384	0.762	0.03
SDQ Prosocial behaviour	107 (64/43)	0.08 (0.27)	0.02 (0.12)	0.918	0.918	0.37
BPM Internalising problems	106 (63/43)	0.21 (0.41)	0.22 (0.42)	0.344	0.695	0.02
BPM Externalising problems	106 (63/43)	0.07 (0.27)	0.16 (0.37)	0.262	0.491	0.23
BPM Attention problems	106 (63/43)	0.13 (0.33)	0.14 (0.35)	0.531	0.657	0.03
<i>Parenting Outcomes</i>						
Family Involvement Questionnaire (Home-based subscale)	107 (64/43)	38.53 (6.28)	39.73 (6.25)	0.711	0.971	0.18
Attentional Control Scale (Attentional focusing subscale)	106 (63/43)	26.94 (4.29)	27.68 (4.34)	0.881	0.881	0.34
Always supervises internet use	101 (60/41)	0.54 (0.50)	0.60 (0.50)	0.561	0.953	0.06
Rules in place to restrict internet access	99 (59/40)	0.90 (0.31)	0.97 (0.17)	0.855	0.976	0.28
Restrictions in place to restrict internet access	99 (59/40)	0.79 (0.41)	0.86 (0.35)	0.670	0.971	0.16
<i>Other</i>						
BPM Total problems standardised score	106 (63/43)	54.60 (6.87)	56.48 (6.44) <sup>3</sup>	0.138	0.256	0.24
Child has ASD-ADHD	107 (64/43)	0.08 (0.27)	0.05 (0.21)	0.561	0.561	0.10
Child has learning difficulty	107 (64/43)	0.06 (0.23)	0.11 (0.32)	0.204	0.260	0.19

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional *p*-value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional *p*-value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the pooled standard deviation. The conditional set includes gender maternal knowledge of child development, parenting self-efficacy, maternal attachment, and maternal consideration of future consequences.



**Table B6** Comparison of high and low treatment groups: Parent reported child health outcomes conditional on baseline differences

	N (HIGH/LOW)	M <sub>HIGH</sub> (SD)	M <sub>LOW</sub> (SD)	p <sup>1</sup>	p <sup>2</sup>	ES
<i>Health</i>						
Child has good health	107 (64/43)	0.97 (0.16)	0.93 (0.26)	0.361	0.636	0.18
Child is diagnosed with asthma	107 (64/43)	0.13 (0.34)	0.20 (0.40)	0.165	0.642	0.16
Child has ongoing diagnosed chronic illness (other than asthma)	107 (64/43)	0.18 (0.39)	0.23 (0.42)	0.199	0.511	0.13
No. of hours sleep on average night	107 (64/43)	10.58 (0.80)	10.77 (0.71)	0.730	0.730	0.15
Child has sleep problems	107 (64/43)	0.45 (0.50)	0.54 (0.50)	0.394	0.590	0.05
<i>Health Service Use in last 12 months</i>						
Attended GP	107 (64/43)	0.55 (0.50)	0.67 (0.47)	0.186	0.668	0.20
Attended Emergency Department	107 (64/43)	0.21 (0.41)	0.25 (0.44)	0.338	0.793	0.13
Attended routine dental check-up	107 (64/43)	0.53 (0.50)	0.65 (0.48)	0.743	0.743	0.19
Attended health facility due to accident	107 (64/43)	0.15 (0.36)	0.11 (0.31)	0.619	0.881	0.01
Attended health facility due to chest infection	107 (64/43)	0.14 (0.35)	0.19 (0.39)	0.263	0.798	0.11
Attended health facility due to wheezing or asthma	107 (64/43)	0.08 (0.28)	0.12 (0.33)	0.234	0.798	0.12
Received dental treatment	107 (64/43)	0.44 (0.50)	0.46 (0.50)	0.793	0.919	0.10
<i>Meets dietary recommendations</i>						
Grains	107 (64/43)	0.59 (0.50)	0.62 (0.49)	0.673	0.917	0.02
Protein	107 (64/43)	0.28 (0.45)	0.29 (0.46)	0.787	0.929	0.11
Dairy	107 (64/43)	0.11 (0.31)	0.12 (0.33)	0.510	0.944	0.00
Fruit & Vegetables	107 (64/43)	0.06 (0.24)	0.10 (0.31)	0.810	0.810	0.15

**Notes:** N' indicates the sample size. 'M' indicates the IPW-adjusted mean. 'SD' indicates the IPW-adjusted standard deviation. <sup>1</sup> one-tailed (right-sided) conditional *p*-value from individual IPW-adjusted permutation test with 100,000 replications. <sup>2</sup> one-tailed (right-sided) conditional *p*-value from IPW-adjusted stepdown permutation test with 100,000 replications. 'Effect size' is the ratio of the treatment effect to the pooled standard deviation. The conditional set includes gender maternal knowledge of child development, parenting self-efficacy, maternal attachment, and maternal consideration of future consequences.