Changing the Life Trajectories of Australia's Most Vulnerable Children

Report No. 2

The first twelve months in the Early Years Education Program: An initial assessment of the impact on children and their primary caregivers

March 2018

THE UNIVERSITY OF MELBOURNE

MELBOURNE INSTITUTE
APPLIED ECONOMIC & SOCIAL RESEARCH

children's protection society

The first twelve months in the Early Years Education Program: An initial assessment of the impact on children and their primary caregivers

Yi-Ping Tseng	Melbourne Institute of Applied Economic and Social Research, University of Melbourne
Brigid Jordan	Department of Paediatrics, University of Melbourne; Royal Children's Hospital; Murdoch Children's Research Institute
Jeff Borland	Department of Economics, University of Melbourne
Nichola Coombs	Department of Economics and Melbourne Institute of Applied Economic and Social Research, University of Melbourne
Kate Cotter	
Alice Hill	
Anne Kennedy	Melbourne Graduate School of Education, University of Melbourne; Faculty of Arts and Education, Charles Sturt University
Corresponding author	**Jeff Borland (jib@unimelb.edu.au)**

Acknowledgements

Our greatest debt is to the children and families who have been willing to participate in the research trial of the Early Years Education Program. We are grateful for the expertise, care and professionalism of the many staff at the Children's Protection Society who have been involved in providing the Early Years Education Program since it commenced in 2010. The Early Years Education Program was initiated and has been substantially funded by the Children's Protection Society. We have appreciated the opportunity to work in partnership with, and the support we have received from, the Children's Protection Society Board and its senior management. A complete list of staff from the Children's Protection Society who have been involved with delivery of the Early Years Education Program and the research trial is in Appendix 1. Funding for the trial of the Early Years Education Program has been provided by the Children's Protection Society, Commonwealth Department of Education, Employment and Workplace Relations (now Department of Education), Commonwealth Department of Families, Housing, Community Services and Indigenous Affairs (now Department of Social Services), the Victorian Department of Human Services (now Victorian Department of Health and Human Services), the Victorian Department of Education and Training (now Victorian Department of Education), the Ian Potter Foundation, the RE Ross Trust, the Pratt Foundation, the Barr Family Foundation, the Sidney Myer Fund, Vic Health, the Antipodean Family Foundation, the Murphy-McNicol Family, the Crawford Foundation, and Australian Research Council Linkage Grant LP140100897. We have benefited from excellent research assistance from Tamera Clancy and Lauren McCabe in collecting the baseline data, Andrew Bevitt in setting up the database, Kerry Ware, Steph Brophy and Jonathan Reyes in data collation, and Penny Hartmann, Jane Sheehan, Xuan Vu and especially Mael Guillou in preparing this report.

Contents

Executive summary

Background

This report presents initial findings on the impact on children and their primary caregivers after twelve months of enrolment in the Early Years Education Program (EYEP).

EYEP is a model of early years care and education targeted at the particular needs of children who are exposed to significant family stress and social disadvantage. Children who participate in EYEP are offered three years of care and education (50 weeks per year, five hours per day each week). Key features of EYEP are high staff/child ratios, qualified and experienced staff, an infant mental health consultant in the team and a rigorously developed curriculum. A relationship-based pedagogy is used to ensure that children are ready for learning. The ultimate objective of EYEP is to ensure that at-risk and vulnerable children realise their full potential and arrive at school developmentally and educationally equal to their peers.

The impact of EYEP is being evaluated through a Randomised Controlled Trial (RCT) as part of the Early Years Education Research Program (EYERP); otherwise referred to in this report as the 'EYEP trial'. Children for whom consent was given to participate in the EYEP trial were randomly assigned into either an intervention group who were enrolled in EYEP or to a control group. Estimates of the impact of EYEP on children and their primary caregivers are derived from comparisons of outcomes between the intervention group and the control group. Measurement of outcomes described in this report took place twelve months after entry to the trial.

Children and primary caregivers in the EYEP trial

To be eligible for the EYEP trial, children had to be aged less than 36 months at the time of entry to the trial, assessed as having two or more risk factors as defined in the Department of Human Services 2007 *Best Interest Case Practice Model*, and be currently engaged with family services or child protection services and have early education as part of their care plan.

The eligibility criteria enabled the selection of a group of participants in the EYEP trial for whom the program was designed – children with substantial development delay living in families experiencing high levels of stress. Even relative to children living in low SES households, children participating in the EYEP trial are highly disadvantaged. EYEP trial participants had lower birth weight, and at the time of entry to the trial, had compromised development of language, motor skills and adaptive behaviour.
The primary caregivers of children in the EYEP trial have fewer personal and social resources available to face the challenges of parenting compared with caregivers in low SES households in Australia.

Outcome measures

Findings on the impact of EYEP on six main outcomes are presented in this report:

- Child development – IQ
- Child development – Language skills
- Child development – Within-child protective factors related to resilience
- Child social and emotional development
- Parent psychological distress
- Home environment

Sample size

In total, 145 children were recruited to the EYEP trial. These children come from 99 families. The intervention group comprises 72 children from 50 families, and the control group has 73 children from 49 families. There are 64 girls and 81 boys in the trial.

Over time, there has been attrition from the group of 145 children who were originally recruited to the EYEP trial. The data collection at twelve months, on which the analysis in this report primarily relies, obtained responses from 114 children and their primary caregivers. In addition, it was not possible to collect a complete set of data on outcomes for this sample of 114 children and primary

caregivers. Hence, the analysis of the impact of EYEP on outcomes at twelve months is based on samples that usually consist of 90 to 100 children or primary caregivers, depending on the outcome.

Assessment of randomisation

Randomisation of children to enrolment in EYEP or to the control group was implemented successfully. Balance was achieved for almost all characteristics of children and primary caregivers on which data were collected at the time of entry to the trial. Where balance was not achieved between the intervention and control groups – predominantly for several measures of child development – this seems primarily to be due to attrition that occurred between the time of consent being given for a child to participate in the EYEP trial and the main stage of data collection at the time of entry to the trial.

Empirical method and definition of sample

Where random assignment of participants between an intervention group and a control group achieves balance in the characteristics of those groups, analysing the impact of a program is straightforward. The causal impact of the program on an outcome can be evaluated by a simple comparison of the average values of that outcome between the intervention and control groups. In this study however, some characteristics are found to be imbalanced between the intervention and control groups. Hence, it is necessary to choose a method for estimating the impact of EYEP that can account for the imbalance. In our analysis, two estimation methods are applied – a regression method and a propensity score matching method. These approaches are alternative ways in which it is possible to control for differences between the characteristics of the intervention and control groups. By adjusting for those differences, it follows that estimates of the impact of EYEP should reflect only the effect of participation in the program.

Defining the sample of children (or primary caregivers) used to estimate the impact of EYEP on each outcome involves two steps. In the first step, for all outcomes the sample is restricted to observations where data are available on children's IQ and language at the time of entry to the trial. This is because these measures are used as covariates in the regression and matching analysis for all outcomes. In the second step, which is done separately for each outcome, the sample is further restricted to those children (or primary caregivers) for whom data on that outcome were collected at twelve months.

Usage of early years care and education service by children enrolled in EYEP and the control group

The estimated impact of EYEP will depend on two aspects of the early years care and education services children received: first, on differences in the extent to which children in the intervention and control groups attended early years care and education programs in the first twelve months in the trial; and second, on differences between the quality and features of EYEP and the programs attended by the control group.

Children enrolled in EYEP received a larger number of hours of early years care and education than children in the control group in the first twelve months of the trial. For example, for the sample of children on whom IQ data were collected at the time of entry to the trial and at twelve months, children enrolled in EYEP received an average of 21.6 hours per week of formal early years care and education compared to 15.0 hours per week for children in the control group. Hence, it seems reasonable to regard any impact of EYEP on outcomes for children in the first twelve months of the trial as potentially deriving both from receiving a larger number of hours of early years care and education services, and from differences between the quality and attributes of EYEP and the services received by the control group.

Attendance at EYEP and type of impact of EYEP estimated

A large proportion of children enrolled in EYEP had high rates of attendance in the first twelve months of the trial, but there was also a sizable group who attended for a small fraction of the available time. For example, about two-thirds of children in the intervention group attended EYEP for at least 60 days; with the other one-third either not commencing or attending for less than 60 days. Most children who attended for more than 60 days were present for relatively large fractions of the available days. More than 90 percent of this group of children attended for at least 70 percent of the available days in the first twelve months. Where children had low rates of attendance in the first twelve months, this was most often due to a

delay in commencing in the program following consent to participate in the trial being given.

In this report, the impact of EYEP after twelve months of enrolment is estimated for children who attended the program for at least 60 days in the first twelve months. We choose to restrict attention to this sample in order that all children in the intervention group have had a level of exposure to EYEP that could conceivably be expected to affect their development.

Main findings

Estimates of the impact of EYEP on child development after twelve months of enrolment are encouraging, but not as yet conclusive.

Children who attended EYEP had an estimated increase in IQ relative to the control group of about 4 points. On the scale used for measuring IQ, this is a relatively large increase, representing about one-quarter of a standard deviation in a norm population. For a normal distribution of IQ scores, this would imply that the mean IQ for children enrolled in EYEP is better than the outcomes of about 60 percent of the control group; whereas if EYEP had no effect the mean IQ for children in EYEP would be better than only 50 percent of the control group. The estimated impact size of EYEP on IQ seems comparable to estimates from early years demonstration programs in the United States; as evidenced, for example, by the RAND Corporation's 2005 review, *Early Childhood Interventions: Proven Results, Future Promise*.

The estimated effect of EYEP on IQ is, however, only marginally statistically significant. The marginal significance of the estimated impact of EYEP on IQ appears to be partly explained by large changes in IQ for many children between the time of entry to the trial and at twelve months – probably associated with extra variability when developmental assessments are carried out for children at very young ages.

The positive impact of EYEP on IQ is primarily observed for boys. Compared to the overall impact on IQ of about 4 points, the estimated impact for boys is 5.8 points and for girls is 1.6 points – with only the impact for boys being statistically significant. The most likely explanation for the different impact on IQ by gender is that the IQ scores for boys at the time of entry to the trial were lower than for girls, hence offering greater scope for improvement.

No significant impact on other development outcomes for children who attended EYEP was detected. The estimated impact of EYEP on children's language development and resilience during the first twelve months of enrolment is small. There is a relatively large estimated impact on children's social and emotional development, but the effect is not statistically significant. In addition, no significant impacts of EYEP during the first twelve months of the program are found for the psychological well-being of primary caregivers or the home environment.

These estimates of the impact of EYEP after the first twelve months of enrolment are robust to alternative estimation methods (regression or matching) and to controlling for different sets of characteristics of children and their primary caregivers. The results are also not affected when children who attended EYEP for less than 60 days are included in the intervention group.

1. Introduction

This report presents initial findings on the impact on children and their primary caregivers after twelve months of enrolment in the Early Years Education Program (EYEP). EYEP is a model of early years care and education targeted at the particular needs of children who are exposed to significant family stress and social disadvantage.

The impact of EYEP is being evaluated through a Randomised Controlled Trial (RCT) as part of the Early Years Education Research Program (EYERP); otherwise referred to in this report as the 'EYEP trial'.

Children in the EYEP trial have been randomly assigned to be enrolled in EYEP or to a control group. Children enrolled in EYEP are allocated to receive three years of a centre-based early years education and care program (five hours per day from Monday to Friday for 50 weeks each year). The control group receive 'usual care', a mix of parental and guardian care as well as education and care provided by other local childcare centres or kindergartens, chosen by parents without direction from the trial.

Estimates of the impact of EYEP on children's development are derived from comparisons of outcomes between the children who are enrolled in EYEP and the control group of children. Measurement of outcomes for both groups of children took place twelve months after their time of entry to the trial. Findings from comparisons of the well-being of primary caregivers and of the home environment of children enrolled in EYEP and in the control group are also presented.

EYEP was initiated by the Children's Protection Society (CPS), an independent not-for-profit child welfare organisation based in the north-east of Melbourne. The program was designed and is being implemented by CPS in collaboration with Associate Professor Brigid Jordan and Dr Anne Kennedy.

The EYEP trial is being undertaken by a consortium of researchers (who are authors of this report) with support from their institutions and in partnership with CPS. Funding for the research trial has come from CPS, government departments at the Commonwealth and State levels, philanthropic organisations, individual donors and the Australian Research Council.

Section 2 describes the motivation for the EYEP trial. Section 3 presents an overview of EYEP. Section 4 provides background information about the EYEP trial, and details on the characteristics of participants in the trial. Section 5 introduces the outcome variables that are examined in this report. Sections 6 and 7 present preliminary information relevant to interpreting the main findings. Section 8 describes the empirical methods used to estimate the impact of EYEP. Section 9 presents the main findings on the effect of enrolment in EYEP for twelve months.

2. Background to the EYEP trial

Children's experiences in the years immediately after birth are a major determinant of their lifetime circumstances and well-being. Early life experiences have a fundamental influence on brain architecture, genetic development and physiology. Critical aspects of children's early experiences are the interactions they have with the people around them and the degree of stress they live with. Having relations with adults that are 'reciprocal and dynamic' and avoiding excessive stress are regarded as essential to healthy development (Center on the Developing Child at Harvard University, 2016b, pp.7-8).

The impact of the early years is especially pronounced for children who experience neglect and toxic stress. Prolonged exposure to physical, emotional and/or sexual abuse and traumatic experiences early in life have been established to cause profound long-term adverse effects on brain development, including self-regulation capacities and the ability to cope with stress (Perry, 2002; Shonkoff, 2012; Center on the Developing Child at Harvard University, 2016a, pp.7-12). Disruption to brain development in turn affects the ability to learn, with recent studies, for example, showing that self-regulation is linked to the development of literacy and numeracy skills (Raver et al., 2011).

When children fall behind in their development of cognitive and social skills early in life, this disadvantage can become entrenched in later years. This happens because skill development is dynamic and hierarchical. Missing out at an early age means that children lack the necessary building blocks and foundation for subsequent learning (Heckman and Mosso, 2014; and for an overview see Tough, 2016, pp.48-52). Deficiencies in cognitive and social skills before the age of five therefore are likely to persist into later life, and become the basis of problems such as low education attainment, unemployment, teenage pregnancy and involvement in crime (Caspi et al., 2016).

Early adversity has also been linked to physiological disruptions such as alterations in immune function (for example, Bierhaus et al., 2003; Currie and Spatz-Widom, 2010; Nicholson et al., 2010), to an increased risk of lifelong physical and mental health problems, including major depression, heart disease and diabetes (Center on the Developing Child, 2016b, p.6; Campbell et al., 2014), and to a variety of health-threatening behaviours in adolescence and adulthood (for example, Rothman et al., 2008; Ford et al., 2011; Caspi et al., 2016).

Addressing the problem of inequality in skill development for children who are exposed to significant family stress is widely agreed to require a different type of education and care than is available from mainstream early childhood services. In a review article in Science, the renowned educationalist Jack Shonkoff (2011, p.982), argued that whereas most current programs for children from disadvantaged backgrounds focus on providing enriched learning experiences for children and parenting education for mothers, a better approach for redressing inequalities in skill development is likely to be 'by linking high-quality pedagogy to interventions that prevent, reduce, or mitigate the disruptive effects of toxic stress on the developing brain.'

Having a model that addresses the developmental delay of at-risk children is a critical policy issue in Australia. First, the size of the at-risk population of children in Australia is substantial. It has been estimated, for example, that in 2015-16 there were 52,300 pre-school children receiving child protection services (Australian Institute of Health and Welfare, 2017, table S4). Second, at-risk children in Australia currently seem the group least likely to be able to access early years education and care (Biddle et al., 2017). Third, while evidence from trials of demonstration programs such as Perry Preschool and Abecedarian provide insights into the potential impact of early years programs, those trials were undertaken in the United States with populations who were largely African-American and living in small cities in the 1960s (Schweinhart et al., 2005; Campbell and Ramey, 1994). The relevance of this existing evidence to Australia is uncertain – causing, for example, the Productivity Commission to argue (2014, p.155): '… it is unclear whether or not such programs would generate as significant benefits in a different cultural context and where the general quality of ECEC services and schooling is different from that of the United States'. Australian

policy-makers are therefore requesting evidence which is both current and derived from practice in Australia.

It was this set of influences that provided the motivation for CPS to create and trial a new early years program, EYEP. CPS brought together a multi-disciplinary team of researchers in 2009 to undertake the EYEP trial. A pilot was conducted in 2010 in order to refine the service model, the survey and measurement methods, and the research process. Enrolment of children into the EYEP trial commenced in early 2011 and concluded in early 2016. The EYEP trial is approved by the University of Melbourne Human Research Ethics Committee (HREC 1034236). This is the first RCT of an early years care and education intervention in Australia (Tapper and Phillimore, 2012).

3. The Early Years Education Program

EYEP is an innovative Australian inter-disciplinary centre-based early years care and education program designed to meet the particular educational and developmental needs of infants and young children who are exposed to significant family stress and social disadvantage. The program has a dual focus: first, addressing the consequences of significant family stress on children's brain development and emotional and behavioural regulation; and second, redressing learning deficiencies. It involves direct intervention with a child to address his or her identified needs, reverse developmental delays, and reduce the impact of risk factors and adverse events. The ultimate objective of EYEP is to ensure that at-risk and vulnerable children realise their full potential and arrive at school developmentally and educationally equal to their peers.

The foundation of EYEP is a holistic model of care and education. The basis for **care** in EYEP is an attachment-focused, trauma informed, primary-care model which recognises the significance of respectful and responsive relationships for every child's learning and development. The purpose of the primary care model is to encourage the fostering of supplementary significant and secure attachment relationships for children who are likely to be experiencing disrupted and compromised attachment relationships in their home environments. The **education** model in EYEP is a pedagogically-driven reflective teaching model that is child-focused and built on the *National Early Years Learning Framework* of 'Belonging, Being and Becoming' (DEEWR, 2009). Each child has individual learning goals developed in partnership with families. Educators plan a curriculum using play-based approaches and intentional teaching to support each child's learning and development across learning outcomes in the *Early Years Learning Framework*.

Children who participate in EYEP are offered three years of care and education (50 weeks per year, five hours per day each week). Key features of EYEP are high staff/child ratios (1:3 for children under three years, and 1:6 for children over three years), qualified and experienced staff, a rigorously developed curriculum and the use of relationship-based pedagogy. An innovative feature of the program is a trans-disciplinary model with an in-house infant mental health consultant as an integral team member, and family support and early childhood curriculum consultants. The infant mental health consultant conducts an assessment with each child as they commence in EYEP. This provides an understanding of the individual child's emotional functioning and behavioural regulation, and the parent-child attachment relationship, which contribute to the individualised learning plan and the relational pedagogical strategies developed for the child.

The EYEP model actively engages with parents to encourage their continued participation in the program, as well as to enhance their usage of all health, educational and social services available in the community, in order to improve outcomes for their children. Care team meetings with parents and family support/child protection workers and the Early Years Educators (primary care worker for the child) take place every 12 weeks.

EYEP also addresses a variety of barriers that might otherwise exist for families taking advantage of support services – such as affordability, where families' beliefs place low priority on early education services, and inter-personal barriers including attitudes on the part of service providers that might compromise engagement (Centre for Community Child Health, 2011; see also Turnbull et al., 2000).

4. The Randomised Controlled Trial and participants in the EYEP trial

Assignment and sample size

The impact of EYEP is being evaluated through a RCT (see Jordan et al., 2014). Families with children who were eligible and consented to participate in the EYEP trial were randomly assigned into either an intervention group who were enrolled in EYEP or to a control group. Random assignment was implemented with blocked randomisation (variable block size) using the STATA *ralloc* command.

Information on the children for whom consent was given to participate in the EYEP trial is presented in Table 1. There are 145 children who were recruited into the EYEP trial when aged less than 36 months. There are 64 girls and 81 boys, and the children come from 99 families. The intervention group comprises 72 children from 50 families, and the control group has 73 children from 49 families.

It is usual for trials of intensive and long-duration early years programs, such as EYEP, to have relatively small sample sizes. For example, a review of 115 trials of early years programs in the United States found that about 40 percent involved less than 200 children (Cannon et al., 2017, pp.44-46).

Table 1: Key descriptive information on children in the EYEP trial

	Number	Percent
Children – By group		
EYEP	72	49.7
Control	73	50.3
Families – By group		
EYEP	50	50.5
Control	49	49.5
Children – By gender		
Girls	64	44.1
Boys	81	55.9

*Note: In the initial report on the EYEP trial (Tseng et al., 2017) it was incorrectly stated that 97 families were included in the trial. This error did not affect any other information presented in that report.

The intervention group remain enrolled in EYEP for three years, or until school entry if that time is reached before completion of three years in EYEP. At the time of consent to participate in the trial children needed to be young enough to be able to attend EYEP for three years prior to school commencement age. For some children, however, factors such as delay in commencing attendance at EYEP have meant that school entry has occurred without completing three years of attendance at EYEP. The control group receive 'usual care', a mix of parental and guardian care as well as care and education provided by other childcare centres or kindergartens. The usual care is determined by the choice of the child's primary caregiver(s) without any direction from the research trial.

In families with multiple children participating in the trial, all those children were assigned to either the intervention group or control group. In addition, when a new sibling of a participant was born during the recruitment phase, and eligibility conditions for participation were met, then the primary caregiver was invited to consent for the new infant to participate in the EYEP trial in the same group to which their older sibling had been assigned.

While children in the control group do not participate in EYEP, it is still possible they may benefit from being in the trial. There are a variety of ways in which this might happen. A report on each child's development (cognitive, language and motor skills) is provided to their primary caregiver(s) at an annual meeting after assessments have taken place, and the caregivers are encouraged to share the report with any professional involved with their child or family, particularly if there is evidence of developmental delay. Questions about family life and parenting practices in the assessment instruments may prompt primary caregivers to recognise and alter patterns of interaction with their children. Many parents in the control group looked forward to contact with researchers from the EYEP trial, and reported that the researcher was the most consistent professional involved in their life. Hence, it is possible that outcomes for children and their primary caregivers in the control group are superior to if they had not been involved in the trial. In future research,

we plan to investigate this issue by comparing outcomes for the control group against a matched sample from the Longitudinal Survey of Australian Children (LSAC).

Data collection

Data are being collected on an extensive set of outcome measures for participants in the EYEP trial at five points in time: at entry to the trial, at yearly intervals for three years after entry to the trial (at 12, 24 and 36 months), and six months after beginning the first year of school. In this report, analysis is based on data from the time of entry to the trial and at twelve months after entry to the trial.

Eligibility criteria and characteristics of EYEP trial participants

Conditions for eligibility for the EYEP trial were chosen with the aim of evaluating its impact on children exposed to significant family stress and social disadvantage. Children were required to be aged less than 36 months, assessed as having two or more risk factors as defined in the Department of Human Services 2007 *Best Interest Case Practice Model*, and be currently engaged with family services or child protection services and have early education as part of their care plan. The list of risk factors consists of 24 'Child and family risk factors' and nine 'Parent risk factors'. Risk factors include having teenage parents, parental substance abuse, parental mental health difficulties and the presence of family violence. A full list of risk factors is included in Appendix 2. Referrals of potential EYEP participants were made by caseworkers from clients of child welfare services including (but not exclusively from) Child FIRST and Child Protection within the Victorian Department of Health and Human Services.

In an earlier report, we presented a detailed overview of the main characteristics of children in the EYEP trial and their primary caregivers (Tseng et al., 2017). That report confirms that the eligibility criteria achieved the selection of a group of participants in the EYEP trial for whom the program was designed – children with substantial delays in development living in families experiencing high levels of stress. This was evident in several ways.

First, at the time of entry to the EYEP trial, most children had many more than the minimum number of two risk factors. About 30 percent of children had two or three risk factors, 35 percent had four or five risk factors, and 35 percent had six to nine risk factors. The most frequent 'Child and family risk factors' for participants were 'attachment/relationship issues', 'mental health issues' and 'family violence, current or past'; and the most frequent 'Parent risk factor' was 'harsh, inconsistent discipline, neglect or abuse'.

The existence of multiple risk factors for children in the EYEP trial is noteworthy – being consistent with evidence that it is this feature which primarily identifies children who are living in environments likely to adversely affect their long-term development. A review of evidence on resilience to childhood adversity by Fergusson and Horwood (2003, p.130) concludes that:

'…the effects of specific risk factors in isolation on later outcomes often tend to be modest…What distinguishes the high-risk child from other children is not so much exposure to a specific risk factor but rather a familial history characterised by multiple familial disadvantages that span economic and social disadvantages; impaired parenting; a neglectful and abusive home environment; marital conflict, family instability; family violence; and high exposure to adverse life events.'

Second, at their time of entry to the EYEP trial, children had relatively low birth weights, even compared to children of the same age living in the bottom quartile of households in Australia ranked by socio-economic status (SES). They also exhibited compromised development in the areas of language, motor skills and adaptive behaviour. This can be seen in Table 2 (Panel A) which presents summary

Table 2: Characteristics of children in the EYEP trial and their primary caregivers

Panel A: Children in the EYEP trial

	EYEP	LSAC – Low SES households	General population
Low birth weight (Less than 2500g) (percent)	19.7	6.4	4.5
Delay or significant delay in development at time of entry to EYEP trial (percent):			
Cognitive development	15.3		15.9
Language	33.9		15.9
Motor skills	29.0		15.9
Social and emotional development	13.4		15.9
Adaptive behaviour	39.5		15.9

Panel B: Primary caregivers of children in the EYEP trial

	EYEP	LSAC – Low SES households
Severe psychological stress (K6 equal to 19 or greater) (percent)	25.8	4.4
Had a major financial crisis - Past 12 months (percent)	32	18.8
Had problems with the police and a court appearance – Past 12 months (percent)	15.3	4.0
Labour force status: Unemployed and not in the labour force (percent)	89.0	70.7
Equalized disposable family income: Percent less than $250 per week ($2016 qtr. 1)	27.4	12.9

information on the birth weights and development of children in the EYEP trial.

Third, primary caregivers for children in the EYEP trial are more likely to be young parents, have fewer financial resources and not be participating in the labour force. The number of stressful life events beyond their control at the time of entry to the trial was extraordinarily high. Many primary caregivers of children in the EYEP trial had severe levels of psychological distress. Summary information on primary caregivers of children in the EYEP trial is shown in Table 2 (Panel B).

5. Overview of outcome variables

The starting point for the analysis undertaken for this report was to pre-commit to a list of outcomes that would be examined. Six main outcomes were selected. The outcomes relate to children's cognitive and social development, to the psychological well-being of their primary caregivers, and to home environment. A list of the outcomes is in Table 3, and a description of each of the outcomes is provided below (see also Appendix 3 for full technical details).

The outcomes were chosen to allow testing of the main hypotheses for the impact of EYEP that were specified in the protocol paper for the trial (Jordan et al., 2014, p.4); and follow closely the types of outcomes examined in trials of early years programs in the United States (Cannon et al., 2017, p.53). As well as estimating the impact of EYEP for all participants, a pre-commitment was made to undertake an analysis of impacts by gender, as this is a major dimension where the impact of early years interventions has been found to differ (Anderson, 2008; Elango et al., 2015).

- **Child development – IQ and language skills:** These aspects of child development are measured using standardized tests: the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley 2006); and the Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI) (Wechsler, 2002). These are the most widely applied measures of the development of infants and toddlers in clinical and research settings.

Our analysis uses the Bayley Scales for children aged up to 42 months, and WPPSI for children aged 43 months and above. Age-adjusted composite scores can be calculated for the IQ and language domains of development for both measures. Both measures are scaled with a range of 40 to 160, a mean of 100 and standard deviation (SD) of 15. A score of 100 defines the average performance of a given age group, and scores of 85 and 115 are one standard deviation below and above the mean respectively. A score between 70 and 85 is defined to identify a delay in child development, and a score below 70 a significant delay in development.

Since the Bayley Scales and the WPPSI are scaled equivalently against population norms, in our analysis we simply integrate the scores from these measures. This means that if a child was assessed using the Bayley Scales at the time of entry to the trial and WPPSI at twelve months, the scores from each test are treated as being directly comparable. Whether combining the different measures in this way affects the findings on the impact of EYEP is tested as part of the empirical analysis.

- **Child development – Within-child protective factors related to resilience:** This aspect of development is measured by the Devereux Early Childhood Assessment (DECA) (Mackrain et al., 2007; LeBuffe and Naglieri, 2012). It is a parent response measure.

Table 3: Outcomes and measures of the impact of EYEP

	Variable	Measure
1	Child development - IQ	Bayley Scales of Infant and Toddler Development III (BSID); Wechsler Preschool and Primary Scale of Intelligence (WPPSI)
2	Child development – Language skills	Bayley Scales of Infant and Toddler Development III (BSID); Weschler Preschool and Primary Scale of Intelligence (WPPSI)
3	Child development – initiative, self-regulation, attachment/relationships, behavioural concerns	Devereux Early Childhood Assessment Program (DECA)
4	Child social and emotional development	Brief Infant Toddler Social Emotional Assessment (BITSEA); Child Behaviour Checklist (CBCL)
5	Parent psychological distress	K6; The Parenting Daily Hassles Scale
6	Home environment	Home Observation and Measurement of Environment (HOME)

DECA-I is used to assess infants aged from 1 month to less than 18 months, DECA-T is used for toddlers from ages 18 to less than 36 months and DECA-P2 is used for children aged 3 to 5 years. Reponses from each instrument on items relating to children's attachment/relationships, initiative and self-regulation are integrated into a Total Protective Factors Scale. This Scale is reported as age normalised T scores and percentile rankings against a norm population. The T score has mean of 50 and SD of 10, and ranges from 28 to 72. A score of 40 or below is defined as signifying an area of need.

- **Child social and emotional development:**
  These aspects of child development are measured using the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) (Briggs-Gowan and Carter, 2006); and the Child Behavior Checklist (CBCL) (Achenbach and Rescorla, 2000). Both are parent response measures. BITSEA is used for children up to three years, and the CBCL for children older than three years.

The BITSEA Parent Response Form is a tool for identifying children aged less than 36 months who may have social-emotional and behavioural problems and/or delays, or deficits in social-emotional competence. In this report, we focus on the instrument for identifying socio-emotional and behavioural problems. The problem score from BITSEA ranges from 0 to 62. A percentile ranking based on age-based population norms can be assigned to each score.

The CBCL is a parent response index of behavioural, social and emotional functioning intended for children from 18 months up to 5 years. The total score on the CBCL ranges from 0 to 200. A percentile ranking based on age-based population norms can be assigned to each score (although scores below the 50th percentile are aggregated).

The BITSEA and CBCL instruments are integrated to obtain a consistent measure of problems with child social and emotional development. This is done by using the proportion of children classified as having development problems in the clinical range as the outcome measure from each instrument. The clinical range is defined as scores below the population norm age-based 10th percentile cut-off. The robustness of the findings to integrating the BITSEA and CBCL measures is tested in the empirical analysis.

- **Parent psychological distress:**
  Parent stress is measured using the Kessler K6 screening scale (Kessler et al., 2002); and the Parenting Daily Hassles Scale (Crnic and Greenberg, 1990).

The K6 scale is a widely used measure of psychological distress, including in the 1997 Australian National Survey of Mental Health and Wellbeing (Furukawa et al., 2003). The scale has six questions about feelings over the last four weeks. A K6 score is derived from summing the responses of the primary caregiver to these questions. The score can range from 6 to 30, with individuals scoring 6 to 13 being classified as exhibiting 'low' psychological distress, 14 to 18 classified as 'medium' psychological distress, and 19 to 30 classified as 'severe' psychological distress.

The Parenting Daily Hassles scale aims to assess the frequency and intensity/impact of 20 experiences that can be a 'hassle' to parents. The frequency score can range from 0 to 80 and the intensity score from 0 to 100. Scores above (respectively) 50 and 70 are considered to show high frequency and significant intensity of pressure on parents.

- **Home environment:**
  Home environment is assessed using the Home Observation and Measurement of Environment (HOME) (Caldwell and Bradley, 2003).

HOME is a home-based rating of the home environment by an assessor/observer. It is designed to achieve systematic measurement of the environment based on observation of interaction between the primary caregiver and their child, and interview data on significant aspects of the family's interpersonal and physical environment. The Infant-Toddler instrument is used for children aged up to three years; and the Early Childhood instrument for children aged above three years. For our report, we rescale the scores from the instruments so that both have scales from 0 to 100. Higher scores signify a higher rated home environment.

6. Description of data and analysis of randomisation

Background

Analysis in this report draws on data collected from children and their primary caregivers at the time of entry to and after twelve months in the EYEP trial. The data collected at the time of entry to the trial are used primarily to evaluate whether the samples of children and primary caregivers in the intervention group and the control group are balanced in their characteristics. Analysis of the impact of enrolment in EYEP on outcomes for children and their primary caregivers principally relies on the data collected twelve months after entry to the trial.

What we refer to as the time of entry to the trial, in fact, encompasses two stages of data collection. In the first stage, for all children for whom consent was given to participate in the EYEP trial, data on a small set of variables were collected; mainly on basic demographics and the risk factors for eligibility for the trial. In the second stage, at research data collection appointments with the primary caregivers and their children, detailed baseline data were collected on child development, as well as on demographic and other characteristics of children and their primary caregivers.

From the first stage, data are available for all 145 children for whom consent to participate in the trial was given. Data are available from the second stage, however, only for a maximum of 136 children. There were a variety of reasons why data could not be collected for nine children who consented to participate in the EYEP trial – such as children dropping out of the trial due to being placed in out of home care or with other carers who were unwilling to participate in the trial.

By the time of the data collection at twelve months after entry to the trial, there had been further drop-out of children and their primary caregivers from the sample of 136. From that collection at twelve months, data are available for a maximum of 114 children and their primary caregivers.

For children from whom data were collected at the time of entry to the trial and at twelve months, there is also some extra non-response. For example, data might have been collected for a child on their IQ and language development, but not collected on the child's home environment. The main reasons for non-response on specific variables were scheduling issues and time constraints – for example, where at an initial meeting with the primary caregiver it was not possible to complete all data collection and a follow-up meeting could not be scheduled; or was scheduled and the family failed to attend. Other studies on populations of children and families with high levels of disadvantage have experienced similar difficulties in collecting complete information (for example, St.Pierre et al., 2005; US Department of Health and Human Services, 2010, pp.2-19).

Table 4 provides information on the number of responses available for each outcome measure and for the various samples used in the analysis in this report. Numbers of responses are shown for the time of entry to the trial and at twelve months; and separately for the intervention and control groups. Data on outcomes to be examined in this report (apart from the home observation measure) are available for 50 to 57 children who were enrolled in EYEP, and for 47 to 50 children in the control group.

Timing of data collection

Figures 1a to 1c present summary information on the distribution of times of data collection for the sample of children studied in this report. The summary measures are for data collection at the time of entry to the trial and at twelve months, and on the time interval between those phases of data collection. Each phase of data collection usually took some time to complete. In Figures 1a to 1c the timing of data collection is defined to be the date on which assessments for the children's IQ were conducted.

The summary information shows that data for most children were collected in a timely manner. Data collection at the time of entry to the trial took place for about 95 percent of children within three months of

Table 4: Outcome variables – Sample sizes

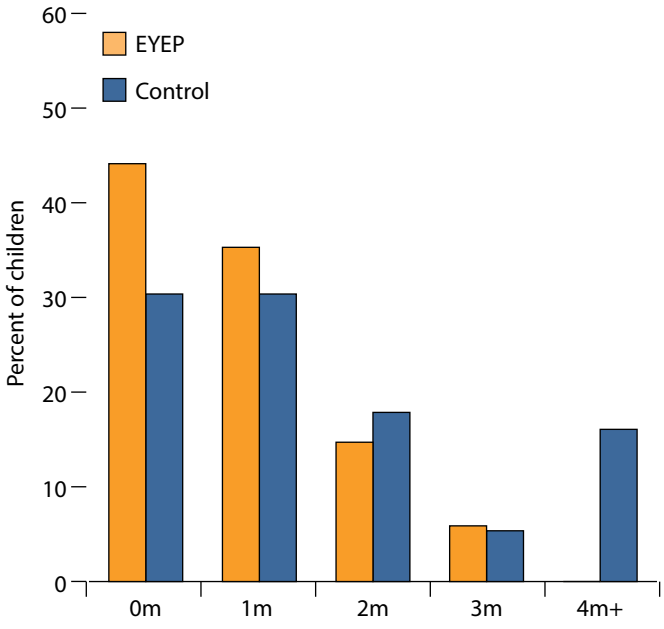
	At time of entry to the EYEP trial			At twelve months after entry to the EYEP trial		
	EYEP group		Control group	EYEP group		Control group
	All	Attendance at least 60 days	All	All	Attendance at least 60 days	All
Bayleys/WPPSI – IQ	68	50	56	57	50	47
Bayleys/WPPSI – Language	68	50	56	57	50	47
DECA	67	49	55	55	49	50
BITSEA/CBCL				54	47	48
K6	68	49	61	53	48	49
Parenting daily hassles	63	44	57	50	46	48
HOME	64	45	55	40	36	43

consent to participate. Data collection at twelve months occurred for over 60 percent of children in the four-month window around their one-year anniversary of consent. For over 80 percent of children the gap between data collection at time of entry to the trial and at twelve months was between ten and fourteen months. Delays in data collection, where they have occurred, have been concentrated among the control group. For example, the control group accounts for all data collection at time of entry to the trial that took place four months or more after consent to participate. This is explained by greater difficulties in scheduling data collection with children and their primary caregivers in the control group, compared to the intervention group who are attending the EYEP centre.

Method of data collection

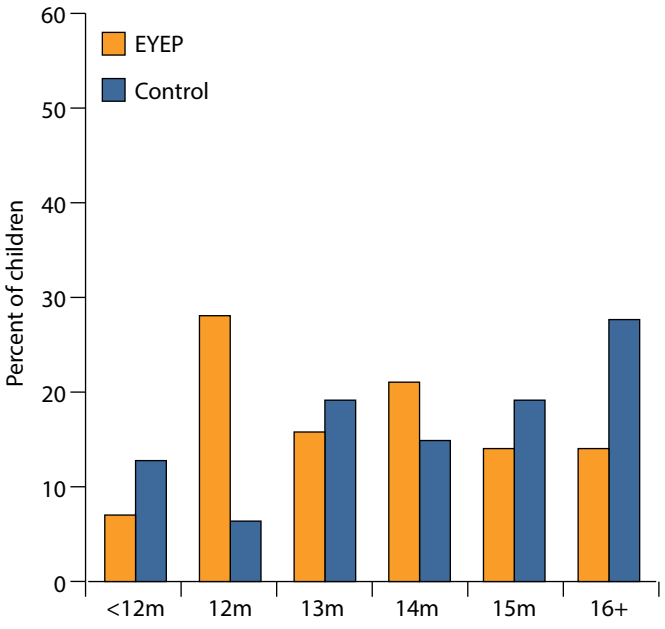
An important consideration is that data collection and data analysis in this project have been non-blind. With it being easiest to collect data for the intervention group at

Figure 1a: Length of time between date of consent to participate in the EYEP trial and date of IQ assessments at time of entry to the trial



Note: 0m means 0 -29 days. 1m means 30-59 days.

Figure 1b: Length of time between date of consent to participate in the EYEP trial and 12-month IQ assessments

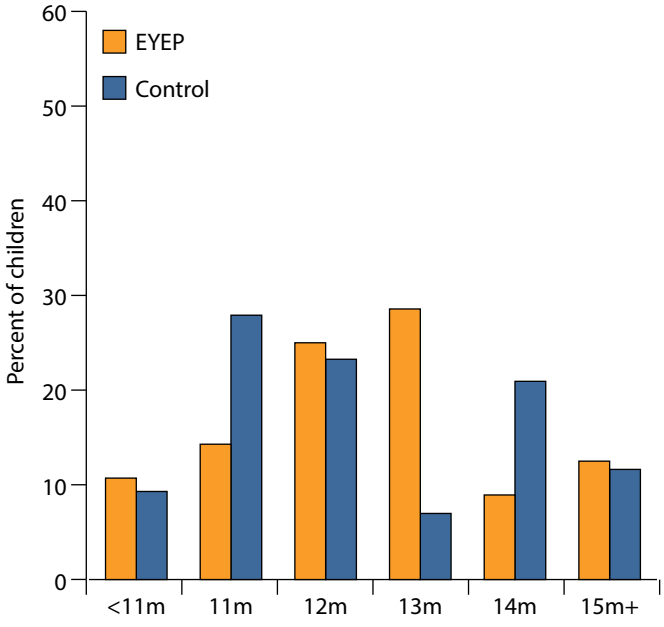


Note: 12m means 365 -394 days.

the EYEP centre, and with some data items being related to assignment status, it would have been impossible to undertake blind data collection for this trial. Similarly, continuous monitoring of the numbers of children in the intervention and control groups remaining in the trial meant it was not going to be possible to undertake the empirical analysis in a genuinely blind manner.

In order to evaluate the impact of non-blind data collection, for the main outcomes in this report which are assessor-derived, the Bayley Scales and WPPSI measures of IQ and language development, all scores have been blind checked by an independent assessor, and adjusted IQ and language scores have been constructed. The raw scores for IQ and language development are very close to the adjusted scores from the blind coder. Over 90 percent of children have the same IQ score, and about 85 percent of children have the same language score. The range of differences between the raw and adjusted scores is from plus 6 points to minus 6 points. Since these differences are relatively evenly distributed around zero, there is hardly any difference between the raw and adjusted mean IQ and language scores.

Figure 1c: Length of time between IQ assessments at time of entry to the EYEP trial and at 12 months



Note: 12m means 365 -394 days.

Did randomisation achieve balance between characteristics of the intervention and control groups at the time of entry to the trial?

A major advantage of a RCT is that it allows the impact of a program to be estimated by calculating the difference in the average values of an outcome between the intervention group and control group. Randomisation implies that trial participants assigned to the intervention and control groups have the same characteristics, with the only difference being that the intervention group has participated in the program being studied. Hence, any significant difference in the average values of an outcome between the intervention and control groups can be attributed to a causal effect of the program. This property only holds, however, where randomisation has resulted in balance between the characteristics of the intervention group and control group.

We examine whether the process of randomisation in the EYEP trial achieved balance between characteristics of the intervention and control groups. This is done using a balancing test on data collected at the time of entry to the trial. The balancing test investigates whether there are statistically significant differences in the characteristics of children and their primary caregivers between the intervention group and control group at entry to the trial.

Even where randomisation is implemented correctly, there are two possible reasons why the characteristics of children and primary caregivers in the intervention and control groups might not be balanced. First, the relatively small number of participants in the EYEP trial may have prevented balance being achieved. Second, since randomisation of children to the intervention and control groups occurred at the time of consent to participate in the trial, but the detailed data on characteristics of children and their primary caregivers is collected later, balance could have been compromised by attrition between these two points in time. One source of attrition is that there were nine children and their primary caregivers who dropped out of the trial after consenting to participate and prior to the main stage of data collection. Another source of attrition is non-responses on some variables by children and primary caregivers who remained in the trial. Where drop out or non-response has a different impact on the intervention and control groups, it has the potential to cause a significant difference between the groups in

some characteristics, even where that difference did not exist amongst the whole sample of 145 children who were recruited to the EYEP trial.

Summary statistics on the average characteristics of the intervention and control groups are reported in Appendix Table 4.1. Whether differences in characteristics between the intervention and control groups are statistically significant is assessed with p-values that are calculated using a permutation test (see Appendix 6.1 for details). Lower p-values make it more likely that there is a significant difference between intervention and control groups; and most often, a p-value less than 0.05 would be interpreted to signify a significant difference.

The summary statistics show that, generally, characteristics of children and primary caregivers in the intervention group and control group are balanced. There are, however, two caveats to this conclusion. First, there are some differences in characteristics between the intervention and control groups which are not statistically significant, but which are relatively large. An example is several of the risk factors, where there are differences of five to fifteen percentage points in the proportions of the intervention group and the control group who have the risk factor. In this case, both the size of the difference in average characteristics and the absence of statistical significance seem most likely explained by the small sample size. Second, significant differences between the intervention and control groups are found for the Bayley Scales measures for children's IQ (5% level), motor skills (5% level) and social-emotional development (10% level). A reasonably large difference also exists for language development, although it is not statistically significant.

We have investigated the differences in the Bayley Scales measures in some detail. Doing this analysis is especially important because the Bayley Scales measures for IQ and language are the two main outcome measures of children's development that are used in this report. The main reason why the Bayley Scales measures are significantly higher for the control group than the intervention group appears to be due to the sample attrition that occurred between the time of consent to participate in the trial and the main stage of data collection when IQ and language assessments were undertaken. The pattern of sample attrition was such that it lowered the Bayley Scales scores for the intervention group relative to the control group.

A formal way to test this explanation is to examine whether reweighting the data to take account of differential attrition between the intervention group and control group can achieve balance in the Bayley Scales scores (Campbell et al., 2014). The method of reweighting that was applied is described in Appendix 4.2. The weights that are used are derived from the estimated relation between attrition and the risk factors for eligibility in the EYEP trial. Results from the reweighting are reported in Appendix Table 4.3. The findings presented confirm that reweighting of the data can achieve balance in the Bayley Scales scores.

We also investigated several other potential sources of the differences in Bayley Scales scores between the intervention and control groups: first, systematic effects of individual assessors on the Bayley Scales scores; second, differences between the intervention and control groups in the (small) proportions of children who exhibited extreme delay in development; and third, whether the need to impute age for some children at the time of the twelve month data collection might have biased the Bayley Scales scores. None of these potential explanations was found to be able to explain the differences in Bayley Scales scores between children in the intervention and control groups at the time of entry to the trial.

Overall, our assessment is that randomisation at the time of consent to participate in the EYEP trial was implemented successfully. Balance was achieved for almost all variables on which data were collected at the time of entry to the trial. Where balance has not been achieved between the intervention and control groups – mainly for the Bayley Scales measures – this seems primarily to be due to attrition in the sample of children that occurred following consent to participate in the trial being given.

7. Details of participation in EYEP by the intervention group and in early years care and education services by the control group in the first 12 months

The estimated impact of EYEP is based on comparisons of outcomes between the intervention and control groups after twelve months in the trial. In that time, the main difference between the groups is the early years care and education services they receive. First, the groups may differ in their extent of engagement with early years care and education services. Second, there are likely to be differences between the quality and attributes of EYEP, which is attended by the intervention group, and the services received by children in the control group.

The implication is that the estimated impact of EYEP will depend on the quality of EYEP and engagement of intervention group children with the program, compared to the amount and quality of early years care and education received by the control group. For example, the estimated impact of EYEP will be lower, the higher are the amount and quality of services received by children in the control group; and vice-versa.

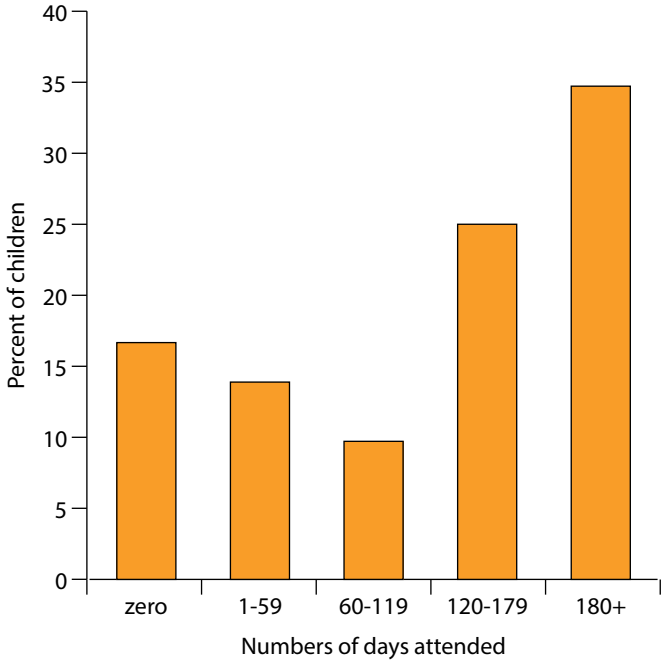
Recent analysis of the impact of demonstration early years programs has reinforced the point that it is necessary to interpret estimates of the impact of early years programs in the context of the amount and quality of services received by the control group (see Elango et al., 2015, p.8). For example, in the United States, it has been suggested that the progressive decreases since the 1960s in the size of estimated impacts of early years programs targeted at disadvantaged children is partly explained by an increase over time in the amount and quality of early years care and education available to children from disadvantaged backgrounds.

Attendance patterns at EYEP

Attendance patterns by the intervention group at EYEP in the first twelve months of the trial are described in Figures 2a to 2c. Figure 2a shows days attended for the whole intervention group. About one-third of children had low levels of attendance, either having not commenced in the program or attending for less than 60 days out of the available days. Available days are defined as the total days between the date of first regular attendance up to the date of the 12 months IQ assessment. Regular attendance is regarded as commencing once a child attends EYEP for at least three consecutive days. About one-third of children attended for between 60 to 180 days, and the remaining one-third for more than 180 days. Extra descriptive information on attendance for the whole intervention group is presented in Appendix 5.

Figures 2b and 2c relate to those children who attended for at least 60 days in the first twelve months, whose outcomes are the focus of this report. Most children in this group attended for relatively large fractions of the available days. For example, more than 90 percent attended for at least 70 percent of available days in the first twelve

Figure 2a: Days attended at EYEP by intervention group in the first 12 months



months. For those children who had smaller fractions of days attended, it seems that the main explanation was a delay commencing full-time attendance following consent to participate in the program. That gap was four months or more for about 35 percent of children. After assignment to the intervention group, an individualised plan for orientation and increasing attendance at EYEP was designed for each child. For some children, however, family circumstances prevented implementation of the scheduled plan.

Figure 2b: Percentage of available days at EYEP attended in first 12 months – Intervention group who attended EYEP for more than 60 days

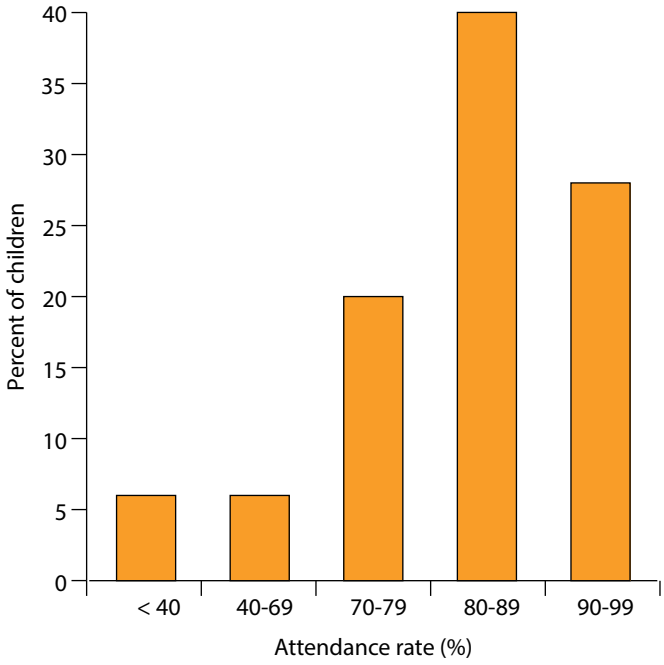
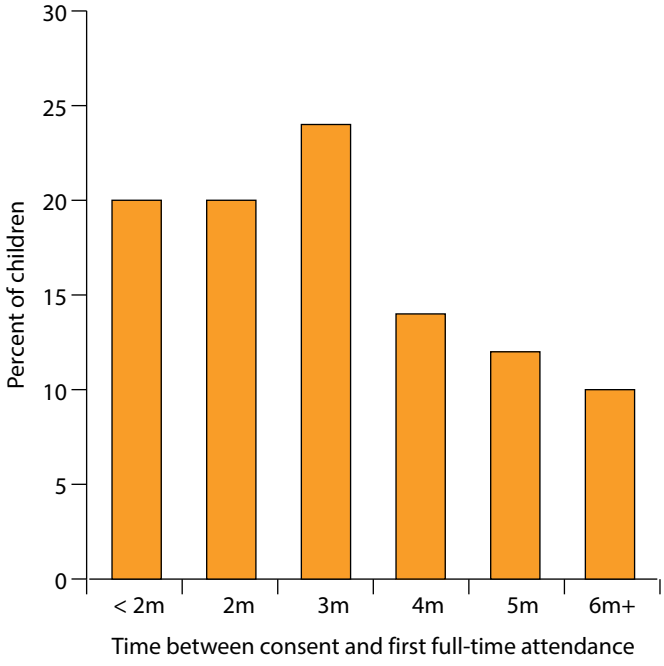


Figure 2c: Time between consent to participate in the EYEP trial and date of commencement of full-time attendance - Intervention group who attended EYEP for more than 60 days



Comparison of early years care and education services received by the intervention and control groups

Descriptive information on the type of formal early years care and education used by children in the control group in the first twelve months of the EYEP trial is shown in Figure 3a. Figures 3b and 3c compare the extent of use of early years care and education services between children in the control group and children enrolled in EYEP.

For Figures 3a to 3c, the samples of children enrolled in EYEP and in the control group are restricted to those who provided data on IQ at the time of entry to the trial and at twelve months. The sample of children enrolled in EYEP therefore differs from those for whom attendance data were reported in the previous sub-section.

In interpreting information on usage of early years care and education services by the control group, it is important to keep in mind that eligibility for the EYEP trial required a child to 'be currently engaged with family services or child protection services and have early education as part of their care plan'. It should also be noted that the measure of usage of early years care and education services for the intervention group in Figures 3b and 3c includes other services in addition to EYEP.

Figure 3a shows that 35 percent of the control group spent no time in formal care, about 40 percent at a day care centre, and the remaining 25 percent in family day care, with a nanny or using multiple types of formal care.

Figure 3b compares annual weeks of formal care and education services received by children enrolled in EYEP and in the control group. A much larger proportion of children in EYEP received what might be regarded as full-year services (40 plus weeks) than in the control group – about 80 percent compared to 40 percent. Figure 3c compares usual weekly hours of care between children in EYEP and the control group – for those children who spent time in formal care. All children in EYEP received

Figure 3a: Type of early years care and education attended by children in the control group in the first 12 months

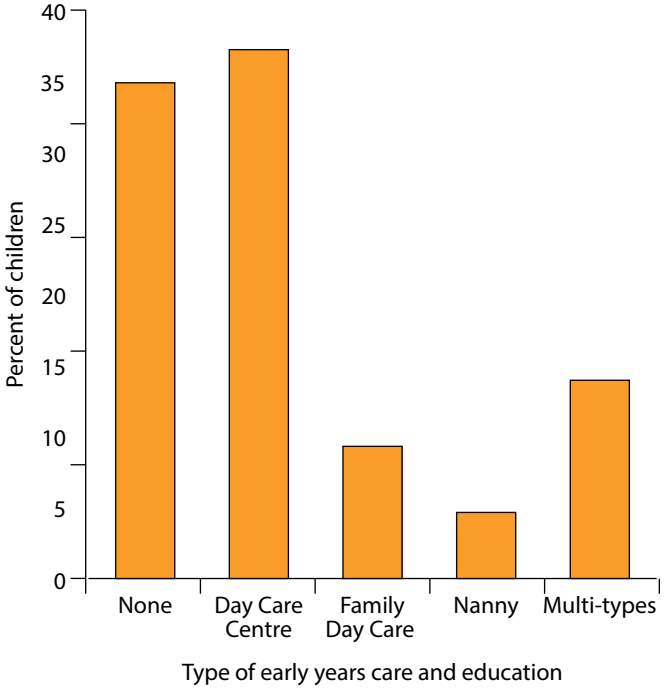
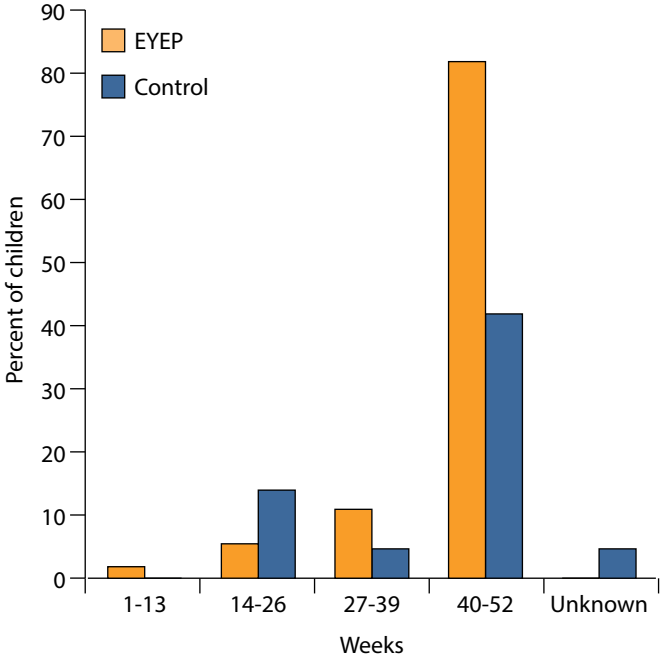


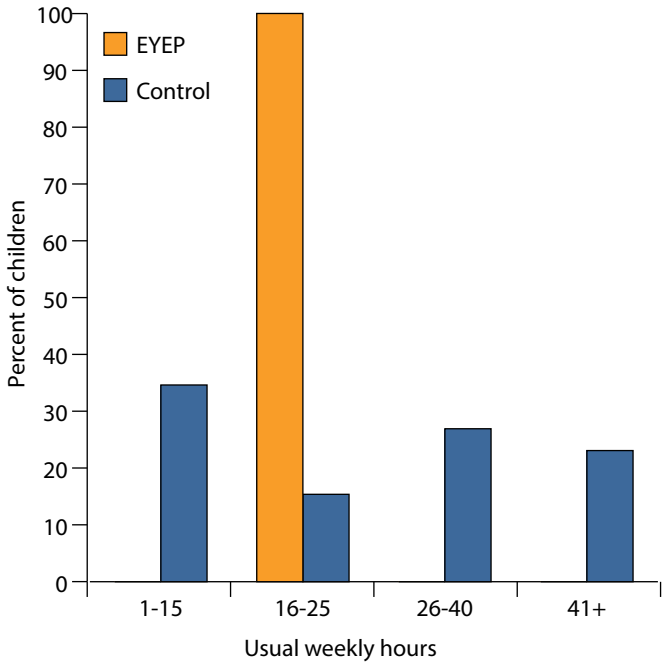
Figure 3b: Annual weeks of formal early years care and education in first 12 months of EYEP trial



services for 16 to 25 hours. For children in the control group there was a much wider spread of hours, with about 50 percent spending 25 hours or less and 50 percent receiving more than 25 hours in formal early years care and education services.

Overall, children enrolled in EYEP received a larger number of hours of early years care and education than children in the control group in the first twelve months of the EYEP trial. For the sample of children on whom IQ data were collected at the time of entry to the trial and at twelve months, average annual hours of services received were 1,126 hours for children enrolled in EYEP and 781 hours for children in the control group. Annual hours of services received by children enrolled in EYEP were relatively concentrated, with almost all children receiving services for more than 40 weeks and for 16 to 25 hours per week. By contrast, the distribution of annual hours of services for children in the control group was much more dispersed. While 35 percent had zero hours, there were some children who had much longer hours. For example, a child at the 90th percentile of the distribution of annual hours of services for the control group received 2,340 hours of care and education, compared to 1,300 hours for the child enrolled in EYEP in the same position in the distribution of annual hours.

Figure 3c: Usual weekly hours in formal care and education in first 12 months of EYEP trial – Children who spent time in formal care and education



Summary

Children enrolled in EYEP received a larger amount of early years care and education services than children in the control group in the first twelve months of the trial – on average 21.6 hours per week compared to 15.0 hours per week. It is important to recognise, however, that some of the extra hours for children enrolled in EYEP occurred outside EYEP; and that attendance at EYEP for some children was relatively limited in the first twelve months. Nevertheless, it seems reasonable to regard any impact of EYEP on outcomes for children in the first twelve months of the trial as potentially deriving both from children enrolled in EYEP receiving a larger number of hours of early years care and education services and from differences between the quality and attributes of EYEP and the services received by the control group.

8. Empirical methodology

Empirical method

The main goal of the empirical analysis is to estimate the impact of EYEP on outcomes for children or their primary caregivers after twelve months in the program. This estimated impact shows the difference in an outcome between children (or caregivers) in EYEP and those in the control group, that can be attributed to enrolment in EYEP.

As an example, suppose the impact of EYEP on HOME is estimated to be plus five (noting as background that the HOME measure has a scale from 0 to 100). This means that after twelve months, the intervention group is being assessed, on average, to score five points higher for their home environment, compared to if they had not been enrolled in EYEP and instead received usual early years care and education.

The key methodological issue in analysing the impact of EYEP is to choose an estimation method. Had there not been evidence of imbalance in some characteristics between the intervention and control groups, it would have been possible to estimate the impact of EYEP by a simple comparison of average outcomes for the groups. However, given that some variables (especially the Bayley Scales) have been found to be unbalanced, it is important to choose a method for estimating the impact of EYEP that can account for that imbalance. Otherwise, it would not be possible to know whether differences in outcomes between the intervention group and the control group at twelve months reflect the effect of EYEP or differences in the underlying characteristics of those groups.

To illustrate, suppose that in the sample from whom data have been collected at twelve months, children enrolled in EYEP are less likely than children in the control group to live with a primary caregiver who has the risk factor of alcohol or substance abuse. The finding that being enrolled in EYEP is associated with a score of plus five on the HOME measure could then reflect either the impact of EYEP or that children enrolled in EYEP are living in better home environments because their caregiver is less likely to have the risk factor of alcohol or substance abuse.

In our analysis, two estimation methods have been applied – a regression method and a propensity score matching method. Essentially, these approaches are alternative ways in which it is possible to account for differences between the characteristics of the intervention and control groups that might have caused differences at twelve months in the outcomes achieved by those groups. The objective is to control as completely as possible for those other differences in characteristics, so that the estimated impact of EYEP reflects only the effect on the intervention group having attended EYEP.

Statistical theory provides some guidance on the set of characteristics that need to be controlled for in this analysis; but ultimately it is also a matter of judgement. For each outcome that is examined, a set of covariates is selected that we regard as constituting a preferred approach to control for differences in characteristics between the intervention and control groups. The robustness of findings from the preferred approach is assessed by also using specifications with alternative sets of covariates. Full technical details of our choice of empirical methods, and how the methods are applied to estimate the impact of EYEP, are provided in Appendix 6.2.

An important reason for using two estimation methods - regression analysis and the matching method - is that both methods are known to have problems when applied to small samples such as in this study (for example, Athey and Imbens, 2016; Zhou, 2004). Adopting alternative estimation methods is intended to provide a check on the robustness of our findings.

In this report, our approach is to primarily present findings on the impact of EYEP estimated using regression analysis. Where a significant impact of EYEP is estimated using regression analysis, we also present estimates derived with the matching method. It turns out that estimates of the impact of EYEP using regression analysis and the matching method are quite similar for all outcomes.

Definition of sample

The sample of children (or primary caregivers) used to estimate the impact of EYEP is restricted to observations where data are available on children's IQ and language from the time of entry to the trial. These measures are used as covariates in the regression and matching analysis to adjust for potential imbalance between the intervention and control groups. As well, for each outcome the sample is further restricted to children (or primary caregivers) for whom data on that outcome are available at twelve months. For example, to estimate the impact of EYEP on HOME, the sample consists of children for whom data on IQ and language are available at the time of entry to the trial and for whom a HOME measure is available from the data collection at twelve months.

What impact of EYEP is estimated?

The impact of EYEP is estimated for children in the intervention group who attended the program during their first twelve months after enrolment. For the analysis undertaken for this report, to be counted as 'attending EYEP during the first twelve months', a child must have attended for at least 60 days. Hence, we exclude from the intervention group those children for whom consent was given to participate in the trial, and who were assigned to EYEP, but who never attended the program; as well as those children who attended the program in the first twelve months for less than 60 days. Outcomes for children who attended EYEP for at least 60 days in the first twelve months are compared with outcomes for children in the control group.

Deciding on the definition of 'attending EYEP during the first twelve months' involves a trade-off. On one hand, it is reasonable to assume that there is a minimum amount of time children would need to spend in EYEP for the program to have an impact on them. On the other hand, excluding children from the intervention group reduces the sample size; and restricting the intervention group to children with higher rates of attendance has the potential to bias our findings, as this group of children and their caregivers are unlikely to be representative of the whole group assigned to be enrolled in EYEP. We have tried to balance this trade-off by choosing a threshold of 60 days of attendance. This threshold causes six children who attended EYEP to be excluded, reducing the size of the intervention group from 56 to 50. In order to be transparent about how excluding these children affects the estimated impact of EYEP, we also report results including those six children as part of the intervention group. Further details of estimation of the impact of EYEP are presented in Appendix 6.3.

Statistical significance

The main findings reported are the estimated impact of EYEP on each outcome and the statistical significance level associated with that estimated impact. The statistical significance level is a measure of how confident it is possible to be that, in a larger population with similar characteristics to children in the EYEP trial, the impact of the program would be the same as found in the EYEP trial. Alternatively, statistical significance can be interpreted as a measure of the precision of the estimated impact of EYEP.

A common way to express statistical significance is as the degree of confidence that can be assigned to the impact size being different from zero if the impact was to be calculated for a larger population. We report what are known as one-tailed and two-tailed tests of significance. A one-tailed test is appropriate if it is considered that the only possible effect of participation in EYEP could have been to cause a zero or positive impact on outcomes. A two-tailed test is relevant if it is also believed that participation in EYEP could have caused a negative impact on outcomes. It is standard in analysis of early years programs to put most weight on one-tail tests. This is supported by the fact that where significant impacts have been estimated for early years programs in previous research, those impacts overwhelmingly have been positive (see Cannon et al., 2017, p.64). Generally, a 5 percent level of significance is used as the threshold for concluding that it is possible to reject that the impact size is equal to zero; that is, the program impact is statistically significant where the p-value is 0.05 or lower (see for example, Elango et al., 2015, p.27).

Statistical significance of estimates of the impact of EYEP from the regression method is assessed with standard errors calculated using the permutation method; and standard errors for the matching method are calculated using a bootstrap method.

9. Results

Children – Main results

Estimates of the impact of EYEP on children's development in the first twelve months of attending the program are shown in Table 5 (rows (1)-(4)). These estimates are from regression modelling with the preferred set of explanatory variables. The main evidence of a significant impact of EYEP is on children's IQ. For other outcomes for children it was not possible to detect any significant effect from participation in EYEP.

The impact size of EYEP on children's IQ is 3.8 points, and the impact is almost significant (at the 5% level) using a one-tailed test. The upper value of the impact on IQ that could not be rejected at the 5% level of significance is 4.1 points.

On the scale used for measuring IQ, this is a relatively large increase, representing about one-quarter of a standard deviation in a norm population. With a normal distribution of IQ scores, it implies that the mean IQ for children enrolled in EYEP would be better than the outcomes of about 60 percent of the control group; whereas if EYEP had no effect the mean IQ for children in EYEP would be better than only 50 percent of the control group. Note that this way of describing the impact of EYEP in children's IQ should be taken as an approximation, since the distribution of their IQ scores at the time of entry to the trial did not exactly follow a normal distribution.

The estimated impact size of EYEP on IQ seems comparable to estimates from early years demonstration programs in the United States. Karoly et al. (2005, p.67) review estimates of impacts on IQ for children near to or soon after commencing in primary school taken from twenty studies of early years programs in the United States. The average estimated impact on children's IQ in these studies was 0.28 of a standard deviation, very close to the estimated impact of EYEP.

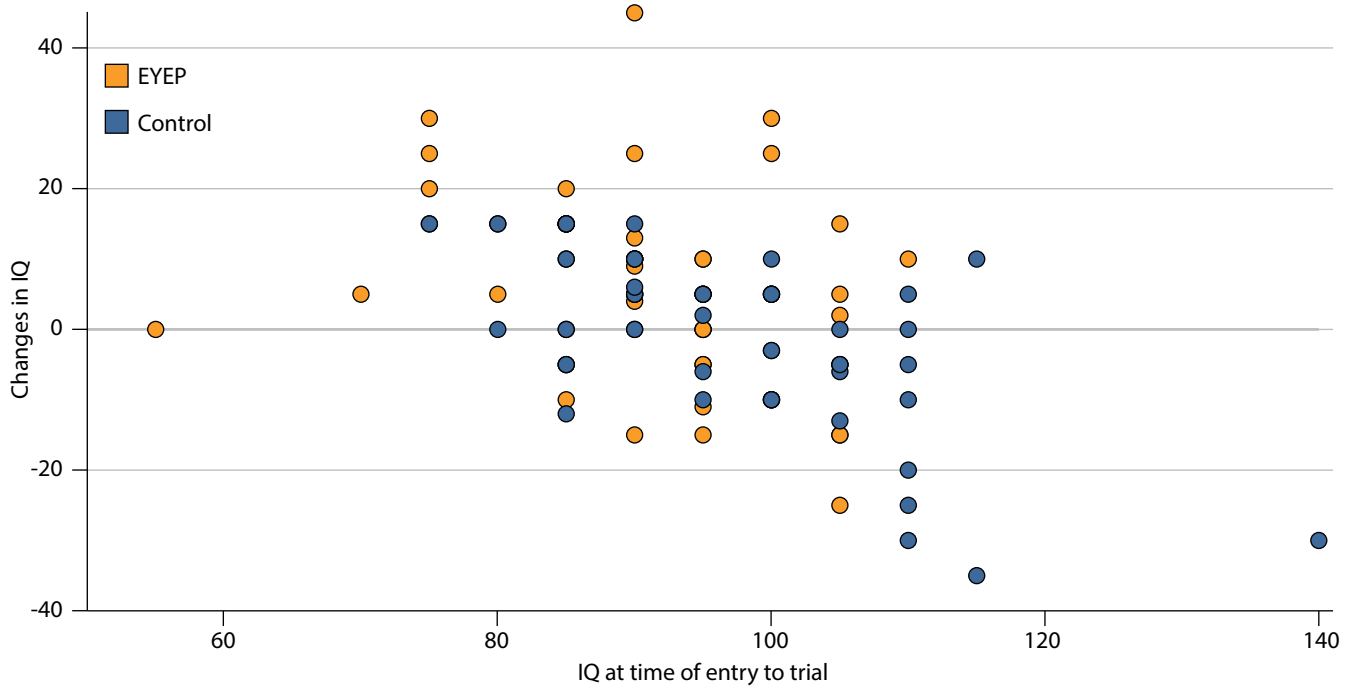
An explanation for the marginal significance of IQ is likely to be the large changes in the IQ measures for children between the time of entry to the trial and at twelve months. The variance in the change in IQ could potentially

Table 5: Impact of enrolment in EYEP for 12 months (Estimates from regression model) – Children who attended for at least 60 days

	Outcome	EYEP mean (12 months)	EYEP impact	1-tail p-value	2-tail p-value	Number of observations (EYEP/ Control)
	Children's development					
(1)	IQ	99.48	3.809	0.065	0.122	50/43
(2)	Language	92.70	0.559	0.425	0.839	50/43
(3)	Protective factors	46.90	1.953	0.163	0.327	49/45
(4)	Social and emotional (Percent in clinical range; Below norm for bottom 10% of population)	21.28	-12.20	0.107	0.219	47/44
	Primary caregiver					
(5)	Psychological distress	15.34	-1.012	0.187	0.370	47/44
(6)	Parenting daily hassles - Frequency	45.60	0.809	0.372	0.753	46/42
(7)	Parenting daily hassles - Intensity	47.88	4.558	0.113	0.225	46/42
(8)	Home environment	73.79	1.715	0.270	0.533	36/37

Note: The EYEP impacts in rows (1)-(3) and (5)-(8) are the estimated impacts of attending EYEP from an OLS regression. The EYEP impact in row (4) is the marginal impact on the percent of children below the 10% threshold on the social-emotional measure estimated from a probit model.

Figure 4: Change in IQ from time of entry to EYEP trial to 12 months after time of entry



affect the significance of the estimated EYEP impact on IQ at twelve months because IQ at the time of entry to the trial is included as an explanatory variable in the regression analysis (and also as a covariate in the matching analysis). Figure 4 presents the change in IQ for each child classified by their IQ at time of entry to the trial. The changes in IQ are centred on zero, but there are also children with large changes. For children participating in EYEP, the 25th and 75th percentile changes in IQ are -1.5 and 15 respectively; and for children in the control group the same measures are -10 and 5. It seems likely that these large changes are explained by the difficulty in doing IQ assessments on very young children, and the error component that is thereby introduced into the measure of IQ. Large changes in IQ have also occurred with very young children in other trials of early years programs (see for example, Campbell and Ramey, 1994, p.690).

Language skills and protective factors do not show large or significant impacts for children from attending EYEP. The impact of EYEP on the fraction of children who are in the clinical range for problems with social and emotional development is an improvement of 12 percentage points. This is quite a large effect when it is recognised that only 21 percent of the intervention group are in this range after twelve months in the trial, but the effect is not statistically significant.

Children – Robustness analysis

The robustness of the estimated impact of EYEP on children's IQ can be investigated by applying alternative estimation methods and specifications. The findings are presented in Table 6. Using alternative estimation methods and controlling for different sets of characteristics does produce some variation in estimates of the impact of EYEP on children's IQ. The impact sizes range from a little below to well above the estimated impact from the preferred regression model - from about one-fifth to one-half of a standard deviation. Significance levels also vary, but are generally slightly higher in the alternative specifications than in the preferred regression specification. Therefore, the estimated impact of EYEP on children's IQ of 3.8 points does seem robust.

A more detailed perspective can be obtained by looking at the specific robustness checks. Estimates from specifications of the regression model which control for different sets of characteristics (rows (1)-(2)) confirm the finding of a considerable impact size and borderline statistical significance. Controlling for whether the IQ score was derived from the Bayley Scales or WPPSI (row (3)) has only a minimal effect on the estimated impact. The results are also fairly similar when the matching method is used (rows (4)-(6)). A hybrid of the regression and matching methods is to undertake regression analysis on a matched

Table 6: Impact on IQ of enrolment in EYEP for 12 months – Sensitivity analysis

	Outcome	Method	Sample	EYEP mean (12 months)	EYEP impact	1-tail p-value	2-tail p-value	Observations (Intervention/ Control)
(1)	12-month IQ	Regression – Alternative specification (1)	Attend ≥ 60 days	99.48	5.325	0.022	0.040	50/43
(2)	12-month IQ	Regression – Alternative specification (2)	Attend ≥ 60 days	97.75	3.508	0.089	0.165	50/43
(3)	12-month IQ	Regression – Preferred specification plus control for whether Bayleys or WPPSI test used	Attend ≥ 60 days	99.48	3.469	0.087	0.155	50/43
(4)	12-month IQ	Matching – Preferred specification	Attend ≥ 60 days	99.48	5.247	0.047	0.094	50/43
(5)	12-month IQ	Matching – Alternative specification (1)	Attend ≥ 60 days	99.48	7.012	0.084	0.168	50/43
(6)	12-month IQ	Matching - Alternative specification (2)	Attend ≥ 60 days	99.34	2.933	0.224	0.450	47/42
(7)	12-month IQ	Regression on matched sample – Preferred specification	Attend ≥ 60 days	99.48	4.780	0.041	0.082	50/43
(8)	Change in IQ from entry to 12 months	Matching – Preferred specification	Attend ≥ 60 days	6.68	7.348	0.017	0.034	50/43
(9)	Change in IQ from entry to 12 months	Matching – Alternative specification (1)	Attend ≥ 60 days	6.68	10.239	0.019	0.038	50/43
(10)	Change in IQ from entry to 12 months	Matching - Alternative specification (2)	Attend ≥ 60 days	6.04	4.330	0.072	0.145	47/43
(11)	12-month IQ	Regression – Preferred specification	All	97.75	3.332	0.089	0.165	56/43
(12)	12-month IQ	Matching – Preferred specification	All	97.75	4.539	0.058	0.116	56/43

Note: For the preferred specification and alternative specification (1) it was possible to match all children in the intervention group to children in the control group. For alternative specification (2) it was not possible to match three children in the intervention group due to the large number of matching covariates.

sample. This hybrid method applies the preferred set of covariates to match EYEP participants to children in the control group, and then uses the same set of covariates in a weighted regression (with weights derived from the matching). The impact of EYEP on children's IQ using this method (row (7)) is between the results obtained from regression analysis and the matching method. The estimated impact on IQ does alter more when the outcome variable is the change in IQ from time of entry to trial to twelve months and matching is used as the estimation method (rows (8)-(10)). The estimated impact of EYEP on IQ is larger, and the estimate is more significant.

Results from analysis of all children for whom data on IQ are available at 12 months suggest that the estimated impact of EYEP is relatively robust to whether children who attended for less than 60 days are included in the intervention group (rows (11)-(12)). The estimated impacts from the regression and matching methods are respectively 3.3 points and 4.5 points. However, only the estimated impact using the matching method is close to significant at the 5% level (one-tail test).

Robustness analysis of the estimated impact of EYEP on other outcomes for children is presented in Appendix Table 7.1. The alternative specifications do not alter the

Table 7: Impact on IQ and language development of enrolment in EYEP for 12 months – By gender

Outcome	Sample	EYEP mean (12 months)	EYEP impact	1-tail p-value	2-tail p-value	Observations (EYEP/ Control)
IQ	**Attend ≥ 60 days**					
	Males	97.52	5.811	0.038	0.074	25/24
	Females	101.24	1.615	0.347	0.678	25/19
	All					
	Males	95.62	6.338	0.021	0.040	29/24
	Females	100.04	0.435	0.459	0.912	27/19
Language	**Attend ≥ 60 days**					
	Males	88.56	0.933	0.405	0.783	25/24
	Females	96.84	-2.463	0.257	0.521	25/19
	All					
	Males	86.31	0.959	0.395	0.763	29/24
	Females	95.15	-3.442	0.173	0.351	27/19

conclusion that these outcomes have not been significantly affected by the first twelve months of attendance at EYEP.

Robustness analysis has also been undertaken to consider the effect of using measures of children's IQ and language that have been adjusted by the blind coder. The findings from this analysis are reported in Appendix Table 7.2. Using the adjusted IQ and language scores also has little effect on the estimated impact of EYEP.

Children – Impacts on development by gender

Previous research has found some differences by gender in the impact of early years programs (see for example, Duncan and Magnuson, 2013, p.125). Hence, we investigate whether the estimated impact of EYEP on IQ and language development varies by gender. This is done using the regression method both for the sample of children who attended for at least 60 days in the first twelve months and for the sample of all children enrolled in EYEP who provided data on their outcomes at twelve months. Note that this analysis involves relatively small sample sizes – about 45 observations for girls and 50 for boys. Results are reported in Table 7.

A strong finding comes from the analysis by gender. For boys, there is a large and significant impact of 6 to 7 points on IQ (about one half of a standard deviation), but a much smaller and insignificant effect for girls. There is no significant impact of EYEP on language skills for either boys or girls. The most likely explanation for the different impact on IQ by gender is that the IQ scores for boys at the time of entry to the trial were lower than for girls, implying greater scope for improvement. At the time of entry to the EYEP trial, the average IQ score for girls was 95.4 and for boys was 90.0. A review of trials of early years programs in the United States by Elango et al. (2015, p.33) similarly concluded that: 'Girls develop earlier. Uniform curricula across genders appears to benefit the laggard boys on many dimensions…'. As well, the gap in achievement at the time of entry to the trial is consistent with recent evidence from the United States, which suggests that at preschool ages boys' development suffers more from an adverse home environment than girls (Autor et al., 2015). However, it is also important to note that a similar difference between girls and boys in language development existed at the time of entry to the EYEP trial, but was not associated with the same improvement for boys during the first twelve months of the trial.

Primary caregivers

Estimates of the impact of EYEP on the well-being of primary caregivers and on home environment in the first twelve months are shown in Table 5 (rows (5)-(8)). The estimated impacts are small, and not close to being statistically significant. Robustness analysis of the estimated impacts for these outcomes are reported in

Appendix 6. The alternative specifications do not alter the conclusion that these outcomes for primary caregivers have not been affected by their children's first twelve months of enrolment in EYEP.

Summary

The findings on outcomes after the first twelve months of the EYEP trial are that the program had a quite large (but only marginally significant) impact on children's IQ, with most of this impact being on boys. No significant impact on other outcomes for children or their primary caregivers was detected.

The size of impact on IQ and absence of impact on other outcomes does appear consistent with evidence from previous trials of early years demonstration programs. For example, Hojman (2015) examines six early childhood programs in the United States, and finds that gains in IQ experienced by the intervention group occur rapidly in the first few months of the program and are followed by small or zero gains in subsequent years.

It is, however, also necessary to consider the possibility that the estimated impact of EYEP on IQ is simply an artefact of sampling variability. Sampling variability implies that, with a 5% significance level, a program would be found to have a significant effect on one out of 20 outcomes. This study has found a significant effect for one out of the six outcomes that have been examined. This might suggest that the impact on IQ could be interpreted as a true effect. But it is also the case that a multiple hypothesis test, which tests the joint impact of EYEP on the set of child development measures, would show no significant impact (Anderson, 2008; Young, 2017).

As well, in interpreting the findings in this report, it should be kept in mind that evaluations of outcomes from early years program outcomes that are undertaken by the program designers generally find larger estimated impacts (for example, Duncan and Magnuson, 2013, p.114).

Overall, it seems appropriate to regard the estimates of the impact of EYEP on child development after twelve months of enrolment as encouraging, but not as yet conclusive.

References

Achenbach, T. and L. Rescorla (2000), *Manual for the ASEBA Preschool Forms and Profiles* (Burlington: VT, University of Vermont, Research Center for Children, Youth and Families).

Anderson, M. (2008), 'Multiple inference and gender differences in the effects of early interventions: A reevaluation of the Abecedarian, Perry Preschool and Early Training projects', *Journal of the American Statistical Association*, 103(484), 1481-95.

Athey, S. and G. Imbens (2016), 'The econometrics of randomized experiments', Stanford University, mimeo.

Australian Institute of Health and Welfare (2017), *Child Protection Australia 2014-15*, Child Welfare series no.66, Catalogue no. CWS 60 (Canberra: AIHW).

Autor, D., D. Figlio, K. Karbownik, J. Roth and M. Wasserman (2015), 'Family disadvantage and the gender gap in behavioural and educational outcomes', Institute for Policy Research, Northwestern University Working Paper Series WP15-16.

Bayley, N. (2006), *Bayley Scales of Infant and Toddler Development* (Third edition, San Antonio, TX: Harcourt Assessment, Inc.).

Biddle, N., R. Seth-Purdie and H. Crawford (2017), 'Risk burden, participation in early childhood education and care, and child outcomes', *Australasian Journal of Early Childhood*, 42(1), 49-59.

Bierhaus, A., J. Wolf, M. Andrassy, N. Rohleder, P.M. Humpert, D. Petrov, D. Ferstl, M. von Eynatten, T. Wendt, G. Rudofsky, M. Joswig, M. Morcos, M. Schwaninger, B. McEwen, C. Kirschbaum and P.P. Naworth (2003), 'A mechanism converting psychosocial stress into mono-nuclear cell activation', *PNAS*, 100(4), 1920-25.

Borland, J., Y. Tseng and R. Wilkins (2005), 'Experimental and quasi-experimental methods of microeconomic program and policy evaluation', pp.83-118 in *Quantitative Tools for Microeconomic Policy Analysis* (Canberra, Ausinfo).

Briggs-Gowan, M. and A. Carter (2006), BITSEA *Brief Infant-Toddler Social and Emotional Assessment* (Bloomington, MN: Pearson).

Caldwell, B., and R. Bradley (2003), *Home Inventory Administration Manual – Comprehensive Edition* (Eau Claire: Wisconsin, Minuteman Press).

Campbell, F.A. and C.T. Ramey (1994), 'Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low income families', *Child Development*, 65(2), 684-98.

Campbell, F.A., G. Conti, J.J. Heckman, S.H. Moon, R. Pinto, E.P. Pungello and Y. Pan (2014), 'Early childhood investments substantially boost adult health', *Science*, 343, 1478-85.

Cannon, J., M. Kilburn, L. Karoly, T. Mattox, A. Muchow and M. Buenaventura (2017), *Investing Early: Taking Stock of Outcomes and Economic Returns from Early Childhood Programs* (Rand Corporation).

Caspi, A., R. Houts, D. Belsky, H. Harrington, S. Hogan, S. Ramrakha, R. Poulton, R. and T. Moffitt (2016), 'Childhood forecasting of a small segment of the population with large economic burden', *Nature Human Behaviour* 01(1), Article number: 0005 doi:10.1038/s41562-016-0005.

Centre for Community Child Health (2011), *Review of the evidence base in relation to early childhood approaches to support children in highly disadvantaged communities for the Children's Ground Project*. Report prepared for the Commonwealth Department of Education, Employment and Workplace Relations (Melbourne: Murdoch Children's Research Institute).

Center on the Developing Child at Harvard University (2016a), *Applying the Science of Child Development in Child Welfare Systems*, accessed at: http://www.developingchild.harvard.edu

Center on the Developing Child at Harvard University (2016b), *From Best Practices to Breakthrough Impacts: A Science-Based Approach to Building a More Promising Future for Young Children and Families*, accessed at: http://www.developingchild.harvard.edu

Crnic, K.A. and M.T. Greenberg (1990), 'Minor parenting stresses with young children', *Child Development*, 61: 1628-37.

Currie, J. and C. Spatz Widom (2010), 'Long-term consequences of child abuse and neglect on adult on well-being', *Child Maltreatment*, 15(2), 111-20.

Dehejia, R. and S. Wahba (1999), 'Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs', *Journal of the American Statistical Association*, 94(448), 1053-1062.

Dehejia, R. and S. Wahba (2002), 'Propensity score matching for nonexperimental causal studies', *Review of Economics and Statistics*, 84(1), 151-161.

Duncan, G.J. and K. Magnuson (2013), 'Investing in preschool programs', *Journal of Economic Perspectives*, 27(2), 109-32.

Elango, S., J.L. Garcia, J.J. Heckman and A. Hojman (2015), 'Early childhood education', Human Capital and Economic Opportunity Global Working Group, Working Paper No.2015-017.

Fergusson, D.M. and L.J. Horwood (2003), 'Resilience to childhood adversity: Results of a 21 year study', pages 131-55 in S.S. Luthar (ed) *Resilience and Vulnerability: Adaption in the Context of Childhood Adversities* (Cambridge: Cambridge University Press).

Ford, E., R. Anda, V. Edwards, G. Perry, G. Zhao, C. Li and J. Croft (2011), 'Adverse childhood experiences and smoking status in five states', *Preventive Medicine*, 53(3), 188-93.

Furukawa T.A., R.C. Kessler and G. Andrews (2003), 'The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being', *Psychological Medicine*, 33(2), 357–362.

Good, P. (2005), *Permutation, Parametric, and Bootstrap Tests of Hypotheses* (Third edition) (Huntington Beach, California: Springer Science+ Business Media).

Heckman, J.J. and S. Mosso (2014), 'The economics of human development and social mobility', *Annual Review of Economics*, 6, 689-733.

Hojman, A. (2015), 'Evidence on the fade-out of IQ gains from early childhood interventions: A skill formation perspective', Mimeo, Center for the Economics of Human Development, University of Chicago.

Jordan, B., Y. Tseng, N. Coombs, A. Kennedy and J. Borland (2014), 'Improving lifetime trajectories for vulnerable young children and families living with significant stress and social disadvantage: The Early Years Education Program', *BMC Public Health*, 14, 965 (10 pages).

Karoly, L., M.R. Kilburn and J.S. Cannon (2005), *Early Childhood Interventions: Proven Results, Future Promise* (Santa Monica, California: Rand Corporation)

Kessler R.C., G. Andrews, L.J. Colpe, E. Hiripi, D.K. Mroczek, S.L. Normand, E.E. Walters and A.M. Zaslavsky (2002), 'Short screening scales to monitor population prevalences and trends in non-specific psychological distress', *Psychological Medicine*, 32, 959–976.

LeBuffe, P. and J. Naglieri (2012), *Devereux Early Childhood Assessment for Preschoolers* (Lewisville, NC: Kaplan Early Learning Company).

Mackrain, M. P. LeBuffe and G. Powell (2007), *Devereux Early Childhood Assessment for Infants and Toddlers: User's Guide* (Lewisville, NC: Kaplan Early Learning Company).

Nicholson, J.M., N. Lucas, D. Berthelsen and M. Wake (2010), 'Socioeconomic inequality profiles in physical and developmental health from 0-7 years: Australian national study', *Journal of Epidemiology and Community Health*, 66(1), 81-87.

Perry, B. (2002), 'Childhood Experience and the Expression of Genetic Potential: What Childhood Neglect Tells Us About Nature and Nurture', *Brain and Mind*, 3, 79-100.

Productivity Commission (2014), Childcare and Early Childhood Learning, Inquiry Report No.73 (Canberra).

Raver, C.C., S.M. Jones, C. Li-Grining, F. Zhai, K. Bub and E. Pressler (2011), 'CSRP's impact on low-income pre-schoolers pre academic skills: Self-regulation as a mediating mechanism', *Child Development*, 82, 362-78.

Rothman, E.F., E.M. Edwards, T. Heeren and R.W. Hingson (2008), 'Adverse childhood experiences predict earlier age of drinking onset: Results from a representative US sample of current or former drinkers', *Pediatrics*, 122, e298-e304.

Rubin, D.B. (1979), 'Using multivariate matched sampling and regression adjustment to control Bias in observational studies', *Journal of the American Statistical Association*, 7(366), 34-58.

Rubin, D.B. and N. Thomas (1996), 'Matching using estimated propensity scores, relating theory to practice', *Biometrics*, 52, 249-64.

Schweinhart, L.J., H. Barnes and D.P. Weikart (2005), *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40* (Ypsilanti, MI: High Scope Press).

Shonkoff, J.P. (2011), 'Protecting brains, not simply stimulating minds', *Science*, 333, 982-83.

Shonkoff, J.P. (2012), 'Leveraging the biology of adversity to address the roots of disparities in health and development', *PNAS*, 109 (Supplement 2), 17302-07.

Smith, J. and P. Todd (2005), 'Does matching overcome Lalonde's critique of nonexperimental estimators?', *Journal of Econometrics*, 125(1-2), 305-53.

Stuart, E.A. (2010), 'Matching methods for causal inference: A review and a look forward', *Statistical Science*, 25(1), 1-21.

St. Pierre, R.G., A.E. Riciutti and T.A. Rimdzius (2005), 'Effects of a family literacy program on low-literate children and their parents: Findings from an evaluation of the Even Start Literacy Program', *Developmental Psychology*, 41(6), 953-70.

Tapper, A. and J. Phillimore (2012), 'Prevention-based approaches to social policy: The case of early childhood development', *Evidence Base*, Issue 2 (August), accessed at: http://exelely.com/archive/22/2012/2

Tough, P. (2016), Helping Children Succeed, (Boston, MA: Houghton Mifflin Harcourt).

Tseng, Y, B. Jordan, J. Borland, T. Clancy, N. Coombs, K. Cotter, A. Hill and A. Kennedy (2017), *Changing the Life Trajectories of Australia's Most Vulnerable Children – Report no.1: Participants in the Trial of the Early Years Education Program*; accessed at: http://fbe.unimelb.edu.au/data/assets/pdf_file/0006/2403438/background-report-final-web.pdf

Turnbull, A., V. Turbville and H.R. Turnbull (2000), 'Evolution of family-professional relationships: Collective empowerment for the early 21st century'. In J.P. Shonkoff and S.J. Meisels (eds.) *Handbook of Early Childhood Interventions* (2nd edition) (Cambridge UK: Cambridge University Press).

United States Department of Health and Human Services (2010), *Head Start Impact Study Final Report* (Washington DC).

Wechsler, D. (2002), *Weschler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI III)* (San Antonio Texas: NCS Pearson).

Young, A. (2017), 'Channelling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results', mimeo. London School of Economics; accessed at: http://personal.lse.ac.uk/YoungA/ChannellingFisher.pdf

Zhou, Z. (2004), 'Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence', *Review of Economics and Statistics*, 86(1), 91-107.

Appendix 1

Staff from the Children's Protection Society involved in delivery of the Early Years Education Program and the research trial

Presidents

Alice Hill	2005-2008
Tim Mulvany	2008-2011
Alice Hill	2011-2013
Jane Munro	2013-2016
Bernard Murphy	2016-present

CEOs

Bernadette Burchell	2006-2011
Dave Glazebrook	2011-2012
Aileen Ashford	2013-present

Assistant to Research/Governance Committee

Margaret Farquharson

Management/team leaders/administration/infant mental health

Janet Williams Smith
Natalie Bou-Ghosn
Aisha Bal
Liza Farquhar
Joanne Kitto
Diana Pellegrino
Nichola Coombs
Shannan Mudie
Monica Robertson
Madeleine Saffigna
Liz Dullard
Dianne Camilleri
Debra Parker

Educators

Sonia Shard
Val Farmer
Jenny Voogt
Barbara Lacey
Natalie Boardman/O'Dath
Sandra Athanasoupolos
Nerissa Linklater
Erin Maree Sharp
Catherine Quirk
Jacquelyn Clark
Jaymi-Lee Warren
Lisa Barbaro
Marylin Ellis
Pnita Holthouse
Jennifer Lovrek
Donna Kavanagh
Helen Brand
Chiara Perri
Lisa McKibbin
Jane Cecelia William
Sarah Meldrum
Tina Howard
Robyn Ball
Sohayla Asari

Cooks

Edwina Pleming
Lea Bautista
Anne Flack
Marcela Ramos
Patrick Carmody

Researchers

Nichola Coombs
Lauren McCabe
Megan Clark
Tamera Clancy
Jane Sheehan

Appendix 2

List of risk factors to healthy child development

Child and family risk factors

- family violence, current or past
- mental health issue or disorder, current or past (including self-harm or suicide attempts)
- alcohol/substance abuse, current or past, addictive behaviours
- disability or complex medical needs eg. intellectual or physical disability, acquired brain injury
- newborn, prematurity, low birth weight, chemically dependent, foetal alcohol syndrome, feeding/sleeping/settling difficulties, prolonged and frequent crying
- unsafe sleeping practices for infants eg. Side or tummy sleeping, ill-fitting mattress, cot cluttered with pillows, bedding or soft toys which can cover an infant's face, co-sleeping with sibling or parent who is on medication, drugs/alcohol or smokes, using other unsafe sleeping place such as a couch or exposure to cigarette smoke
- disorganised or insecure attachment relationship (child does not seek comfort or affection from caregivers when in need)
- developmental delay
- history of neglect or abuse, state care, child death or placement of child or siblings
- separations from parents or caregivers
- parent, partner, close relative or sibling with a history of assault, prostitution or sexual offences
- experience of intergenerational abuse/trauma
- compounded or unresolved experiences of loss and grief
- chaotic household/lifestyle/problem gambling
- poverty, financial hardship, unemployment
- social isolation (family, extended family, community and cultural isolation)
- inadequate housing/transience/homelessness
- lack of stimulation and learning opportunities, disengagement from school, truancing
- inattention to developmental health needs/ poor diet
- disadvantaged community
- racism
- recent refugee experience

Parent risk factors

- parent/carer under 20 years or under 20 years at birth of first child
- lack of willingness or ability to prioritise child's needs above own
- rejection or scapegoating of child
- harsh, inconsistent discipline, neglect or abuse
- inadequate supervision of child or emotional enmeshment
- single parenting/multiple partners
- inadequate antenatal care or alcohol/substance abuse during pregnancy

Wider factors that influence positive outcomes

- sense of belonging to home, family, community and a strong cultural identity
- pro-social peer group

Appendix 3

Outcome measures – Details

- **Child development – IQ and language skills:** These aspects of child development are measured using the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley 2006); and the Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI) (Wechsler, 2002).

The Bayley Scales is a standardized test of development, conducted under standard conditions by a trained assessor, designed for ages from one month up to 42 months. WPPSI is an instrument for assessing the intelligence of children aged from 2 years and 6 months up to 7 years and 3 months, also administered by a trained assessor. These scales are the most widely used measures of the development of infants and toddlers in clinical and research settings. In our analysis we have used the Bayley Scales for children aged up to 42 months, and WPPSI for children aged 43 months and above.

Data on two domains from the Bayley Scales and WPPSI are presented in this report: cognitive development and language development. The measure of cognitive development from the Bayley Scales is the cognitive scale. It assesses abilities such as sensorimotor development, concept formation, memory, and simple problem solving. The corresponding measure from WPPSI is the Full scale composite score. For children aged up to 47 months this is based on four subtests: receptive vocabulary and information plus block design and object assembly; and for children aged 48 months and over on seven subtests: information, vocabulary and word reasoning, plus block design, matrix reasoning, picture concepts and processing speed. The measure of language development from the Bayley Scales is the language scale. It encompasses receptive communication (verbal comprehension, vocabulary) and expressive communication (babbling, gesturing). The corresponding measure from WPPSI is the verbal composite score. For children up to 47 months this is derived from two subtests: receptive vocabulary and information; and for children aged 48 months and over from three subtests: information, vocabulary and word reasoning.

Age-adjusted composite scores can be calculated for the Bayley Scales and the WPPSI instrument for each domain of development. Both measures scale the scores with a range of 40 to 160, a mean of 100 and standard deviation (SD) of 15. A score of 100 defines the average performance of a given age group, and scores of 85 and 115 are one standard deviation below and above the mean respectively. About 68 percent of the norm population of children achieve a score between 85 and 115. A score between 70 and 85 is defined to identify a delay in child development, and a score below 70 a significant delay in development. Since the Bayley Scales and the WPPSI are scaled equivalently against population norms, in our analysis we simply integrate the composite scores from the measures.

- **Child development – Within-child protective factors related to resilience:** This aspect of development is measured by the Devereux Early Childhood Assessment (DECA) (Mackrain et al., 2007; LeBuffe and Naglieri, 2012). It is a parent response measure.

DECA-I is used to assess infants aged from one month up to 17 months. It includes items on Initiative that assess an infant's ability to use independent thought and actions; and items on Attachment/Relationships to assess the quality of relationship between the infant and significant adults. DECA-T is used for toddlers aged from 18 months up to 35 months. It includes items on Attachment/ Relationships; items on Initiative; and items on Self-Regulation that assess a toddler's ability to gain control of and manage emotions, and sustain focus and attention. DECA-P2 is used for children aged 3 to 5 years. It includes items on Attachment/Relationships; items on Initiative; and items on Self-Regulation. Reponses to all items in each of the instruments are on a 5-point scale (0=Never to 4=Very

Frequently). The raw scores on Initiative and Attachment/Relationship can be converted into age normalised T scores and percentile rankings against a norm population. A T score and percentile ranking can also be calculated for a Total Protective Factors Scale which integrates the scores for Initiative, Attachment/Relationship and Self-Regulation. The T score for each aspect has mean of 50 and SD of 10, and ranges from 28 to 72. A score of 40 or below is defined to signify an area of need.

- **Child social and emotional development:**
  These aspects of child development are measured using the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) (Briggs-Gowan and Carter, 2006); and the Child Behaviour Checklist (CBCL) (Achenbach and Rescorla, 2000). Both are parent response measures.

The BITSEA Parent Response Form is a tool for identifying children aged up to 35 months who may have social-emotional and behavioural problems and/or delays, or deficits in social-emotional competence. In this report, we focus on the instrument for identifying socio-emotional and behavioural problems. This instrument includes 31 questions. Each question has a three-response scale (0=Not true/Rarely; 1=Somewhat True/Sometimes; 2=Very true/Often). Hence, the problem score from BITSEA ranges from 0 to 62. A percentile ranking based on age-based population norms can be assigned to each problem score.

The CBCL is a parent response index of behavioural, social and emotional functioning intended for children from 18 months up to 5 years. It consists of 99 potential problem items (for example, 'can't concentrate, pay attention for long'; 'cries a lot'; 'feelings are easily hurt') plus one open-response question. Each of the potential problem items has a three-response scale based on behaviour in the preceding two months (0=Not true; 2=Very true or often true). An extra open response question is coded in the same way. Hence, the total score on the CBCL can range from 0 to 200. A percentile ranking based on age-based population norms can be assigned to each score (although scores below the 50th percentile are aggregated).

The BITSEA and CBCL instruments are integrated to obtain a consistent measure of problems with child social and emotional development by using as the outcome measure from each instrument the proportion of children classified as having development problems in the clinical range; that is, with a score below the population norm age-based 10th percentile cut-off.

- **Parent psychological distress:**
  Parent stress is measured using the Kessler K6 screening scale (Kessler et al., 2002); and the Parenting Daily Hassles Scale (Crnic and Greenberg, 1990).

The K6 scale is a widely used measure of psychological distress, including in the 1997 Australian National Survey of Mental Health and Wellbeing (Furukawa et al., 2003). The scale has six questions about feelings over the last four weeks, with a 5-point item response scale for each question ranging from 'all of the time' (equal to 5 points) to 'none of the time' (equal to 1 point). A K6 score is derived from summing the responses of the primary caregiver to each of the six questions. The score can range from 6 to 30, with individuals scoring 6 to 13 classified as exhibiting 'low' psychological distress, 14 to 18 classified as 'medium' psychological distress, and 19 to 30 classified as 'severe' psychological distress.

The Parenting Daily Hassles scale aims to assess the frequency and intensity/impact of 20 experiences that can be a 'hassle' to parents. A child's primary caregiver is asked to score each of the 20 hassles on a 4-point scale for frequency (1=Rarely to 4=Constantly), and a 5-point scale for intensity (1=Low to 5=High). Hence, the total frequency score can range from 0 to 80 and the total intensity score from 0 to 100. Scores above (respectively) 50 and 70 are considered to show high frequency and significant pressure on parents. Sub-categories of the intensity score for challenging behaviour and parenting total tasks are arrived at by considering subsets of the intensity questions. These scores can range respectively from 0 to 35, and 0 to 40.

- **Home environment:**
  Home environment is assessed using the Home observation and measurement of environment (HOME) (Caldwell and Bradley, 2003).

HOME is a home-based rating of the home environment by an assessor/observer. It is designed to achieve systematic measurement of the environment based on observation of interaction between the primary caregiver and their child, and interview data on significant aspects of the family's interpersonal and physical environment. The Infant-Toddler instrument (up to three years) includes 45 total items (for

example, 'Parent spontaneously praises child at least twice; 'Push or pull toys'; 'Child has 3 or more books of his/her own'). The Early Childhood instrument (three years and over) includes 55 items (for example, 'Child is encouraged to learn the alphabet'; 'Parent introduces visitor to child'; 'Neighbourhood is aesthetically pleasing'). All items on both instruments are scored as plus/minus. Hence the raw scores can range from (respectively) 0 to 45 and 0 to 55. The scores from the Infant-Toddler and Early Childhood instruments are integrated by rescaling both scores to range from 0 to 100.

Appendix 4

Tables relating to balancing tests on data at time of entry to the EYEP trial

Appendix 4.1 - Table: Variable balance between intervention and control groups at time of entry to the EYEP trial

	EYEP group – Mean (I)	Control group – Mean (C)	Difference: I - C	p-value of difference	Number of observations
Family risk factor: Attachment/relationship issues (%)	0.542	0.548	-0.006	0.970	145
Family risk factor: Alcohol or substance abuse (%)	0.361	0.411	-0.050	0.608	145
Family risk factor: Disability/Complex medical issues (%)	0.264	0.329	-0.065	0.478	145
Family risk factor: Mental health issues (%)	0.722	0.781	-0.059	0.591	145
Family risk factor: Family violence, current or past (%)	0.528	0.562	-0.034	0.748	145
Family risk factor: Social isolation (family, community, cultural) (%)	0.333	0.178	0.155	0.117	145
Family risk factor: Inadequate housing/ Transience/Homelessness (%)	0.208	0.137	0.071	0.442	145
Parent risk factor: Parent/carer under 20 years old (%)	0.167	0.123	0.043	0.545	145
Parent risk factor: Lack of ability or willingness to prioritise children's needs (%)	0.250	0.384	-0.134	0.186	145
Parent risk factor: Rejection of child (%)	0.083	0.164	-0.081	0.263	145
Parent risk factor: Harsh, inconsistent discipline, Neglect or Abuse (%)	0.333	0.493	-0.160	0.124	145
Parent risk factor: Inadequate supervision (%)	0.278	0.425	-0.147	0.161	145
Sum of risk factors	4.653	4.890	-0.238	0.564	145
Male (%)	0.569	0.548	0.021	0.774	145
Age at time of referral to trial (Years)	0.903	0.973	-0.070	0.563	145
Age at time of consent to participate in trial (Years)	1.464	1.475	0.011	0.935	145
Both parents attended consent interview (%)	0.458	0.425	0.033	0.767	145
Number of study children in family	1.440	1.490	-0.050	0.743	99
Member of a family with multiple children in study (%)	0.340	0.490	-0.150	0.114	99
Primary caregiver is immigrant (%)	0.242	0.063	0.179	0.056	133
Main language spoken at home non-English (%)	0.314	0.190	0.124	0.291	133
Age of primary caregiver (Years)	30.30	30.94	-0.63	0.698	132
Primary caregiver has post-school qualification (%)	0.434	0.403	0.031	0.780	129
Primary caregiver- Psychological distress (K6 score 1 to 30)	14.446	15.245	-0.819	0.499	129

	EYEP group – Mean (I)	Control group – Mean (C)	Difference: I - C	p-value of difference	Number of observations
Primary caregiver – Proportion with severe psychological distress (%)	0.235	0.278	-0.043	0.579	129
Primary caregiver – Parenting daily hassles – Frequency (0 to 80)	44.446	45.441	-0.995	0.643	120
Primary caregiver – Parenting daily hassles – Intensity (0 to 100)	43.701	41.896	1.805	0.604	118
Primary caregiver – Percent not in labour force (%)	0.824	0.733	0.090	0.337	128
Equivalent weekly family income ($2016 qtr1)	363.81	418.25	-54.41	0.275	108
Equivalent weekly family income – Proportion with less than $250	0.286	0.269	0.016	0.885	108
Financial stress (Index 0 to 4)	2.738	2.381	0.158	0.707	127
Birth weight (kgs)	3.014	2.910	0.103	0.586	116
Birthweight – Proportion less than 1.5kgs	0.085	0.034	0.051	0.309	118
Duration of pregnancy (Weeks)	37.672	37.733	-0.061	0.919	124
Child ever breastfed (%)	0.691	0.857	-0.166	0.075	131
Bayley Scale – IQ	90.221	94.911	-4.690	0.035	124
Bayley Scale – Language	86.044	89.661	-3.617	0.159	124
Bayley Scale – Motor skills	85.618	92.482	-6.864	0.003	124
Bayley Scale – Social-emotional	95.833	103.269	-7.436	0.079	112
Bayley Scale – Adaptive behaviour	88.450	89.185	-0.735	0.827	124
DECA – Total protective score (28 to 72)	45.388	47.691	-2.303	0.208	122
DECA – Total protective score – Proportion 'in need' (%)	0.269	0.218	0.050	0.504	122
Home observation (0 to 100)	64.375	68.613	-4.238	0.121	119
Child has at least one out of a set of on-going problems (eg., hearing problems, eyes or seeing properly, food or digestive allergies)	0.623	0.745	-0.122	0.234	108
Number of times child had injuries requiring medical attention	0.547	0.550	-0.003	0.972	124
Number of times hospitalised since birth	0.426	0.345	0.081	0.528	116
Interval between consent to participate in EYEP trial and main stage of data collection at the time of entry to the trial (Months)	0.449	1.358	-0.909	0.061	136

Appendix 4.2 – Reweighting of data from time of entry to the EYEP trial

Reweighting of observations is done using the risk factors reported at the time of entry to the EYEP trial. The reweighting procedure involves two stages (Campbell et al., 2014).

The first stage is regression analysis of the association between the risk factors and attrition. The definition of attrition is that a child: (i) Had data collected on their risk factors at the time of consenting to participate in the trial; and (ii) Did not have data collected on the Bayley Scale scores at the main stage of data collection at the time of entry to the trial. The regression analysis was undertaken by estimating alternative specifications which included as covariates all possible combinations of up to six risk factors. The model which minimised the Akaike Information Criterion was chosen as the preferred specification. This preferred specification includes as covariates: gender; age at time of consent to participate in the EYEP trial; family risk factor – alcohol or substance abuse; family risk factor – disability/complex medical issues; family risk factor – mental health issues; parent risk factor – harsh, inconsistent discipline, abuse or neglect.

The second stage is to correct for attrition. Coefficient estimates from the preferred regression model are used as inverse probability weights to calculate weighted average Bayley Scales scores for children in the intervention and control groups for whom data on those scores are available. Reweighting is done separately for the intervention and control groups. For the intervention group, there are four children from single child families for whom no cognitive development score is available. These four children are all boys from a single child family which did not have the risk factors of social isolation or homelessness. Therefore the sample used for reweighting is restricted to boys from single child families who did not have these risk factors at the time of consent to participate in the trial.

Appendix 4.3 reports the findings from the reweighting. For each of the four Bayley Scales measures the mean scores for children who are enrolled in EYEP and for the control group are shown – the top row shows the unadjusted means, and the bottom row shows the means after reweighting. The p-value for whether there is a significant difference in the mean scores for the children enrolled in EYEP and in the control group is also shown.

Appendix 4.3 - Table: Bayley Scales scores – At time of data collection at time of entry to the EYEP trial - Raw data and reweighted

Variable	EYEP group	Control group	p-value for difference	Number of observations
Cognitive Raw	90.22	94.91	0.035	124
Cognitive Reweighted	90.27	92.19	0.641	124
Language Raw	86.04	89.66	0.159	124
Language Reweighted	86.04	87.40	0.679	124
Motor Skills Raw	85.61	92.48	0.003	124
Motor Skills Reweighted	86.00	89.70	0.204	124
Socio-Emotional Raw	95.83	103.26	0.079	112
Socio-Emotional Reweighted	96.00	99.30	0.605	112

Appendix 5

Extra descriptive information on attendance at EYEP in the first 12 months by the intervention group

Figure 5.1: Percentage of available days at EYEP attended in first 12 months – All children in intervention group

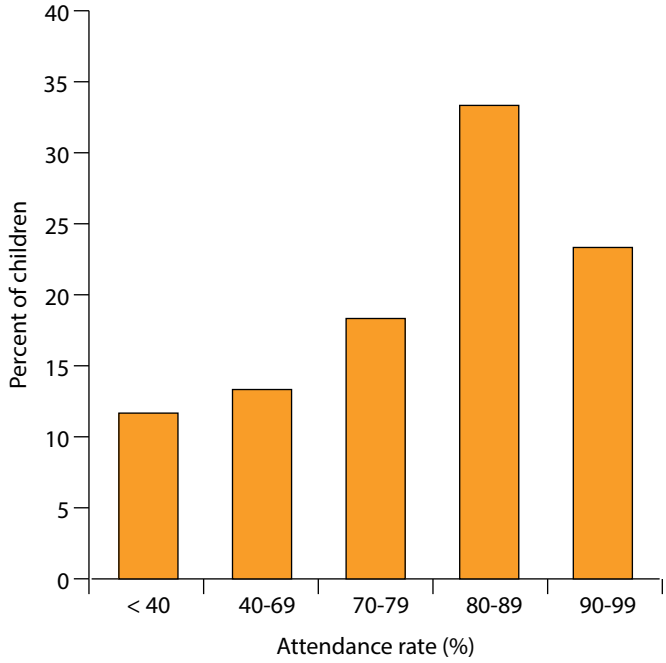
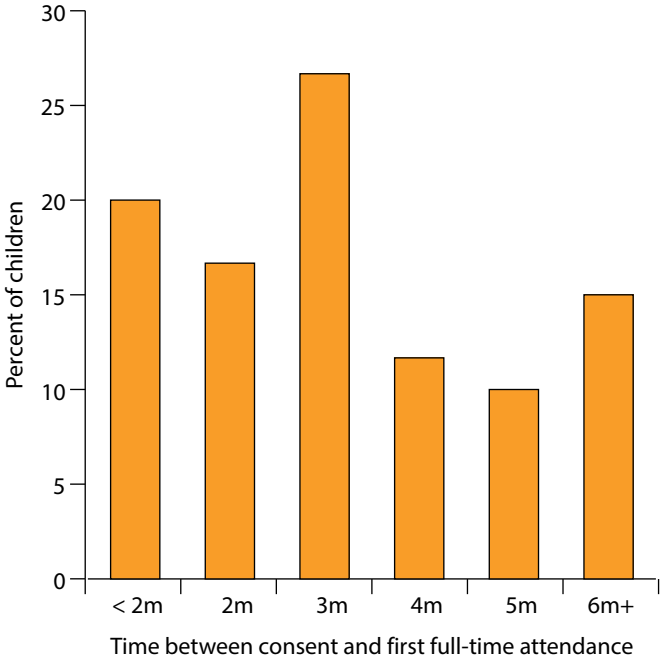


Figure 5.2: Time between consent to participate in EYEP trial and date of commencement of full-time attendance – All children in intervention group



Appendix 6

Details of statistical methods

6.1 Permutation test

Applying the same block randomisation approach as for the original assignment of trial participants to the intervention group and the control group, the sample of participants in the EYEP trial is reassigned 10,000 times between those groups. For each reassignment, the difference in the mean of each variable between the intervention and control groups can be calculated. From the process of repeated reassignment, a distribution of differences in means for each variable can be generated. That distribution is then applied to undertake a two-tailed test for whether there is a significant difference in the mean values of the variable between the intervention group and the control group. For a general reference, see Good (2005).

6.2 Details of the empirical methods used to estimate the impact of EYEP

Regression

The impacts of EYEP on outcomes using the regression method are estimated from the model:

$$Y_i = \alpha + \beta EYEP_i + \gamma X_i + \varepsilon_i \qquad (1)$$

Where: Y_i is an outcome measure for a child (or primary caregiver); $EYEP_i$ is an indicator for whether a child (or primary caregiver) was in the intervention group who attended EYEP for at least 60 days in the first twelve months; and X_i is the set of other covariates that are controlled for in the regression model. The impact of EYEP for children who attended at least 60 days in the first twelve months is equal to β. Standard errors are calculated using the permutation method.

The regression method is known to have several shortcomings (Borland et al., 2005). First, it imposes a linear functional form. Where the linear assumption does not hold, then regression analysis will not provide valid estimates of program impact. Second, where the program impact is heterogeneous between participants, the regression method produces an estimate of program impact that is a weighted average across participants, where weights are determined by observable characteristics of participants. These weights will not necessarily correspond to the weights required for the estimated program impact to be the average effect of treatment on the treated. Third, the regression method does not impose any common support condition. Hence, it is possible that program impact estimates are derived from comparisons of outcomes for program participants and non-participants who differ significantly in their observable characteristics. Finally, concerns have been raised about interpretation of program impacts that are derived from regression analysis on small samples of data from randomised experiments (Athey and Imbens, 2016).

Matching

The matching method involves comparing an outcome for each child (or primary caregiver) who attended EYEP for at least 60 days in the first twelve months with a matched control group of children (or primary caregivers). The aggregate impact of EYEP is calculated as the average of the differences in the outcome between each child (or primary caregiver) in the intervention group and their respective matched control group.

A formal description of the estimated EYEP effect using the matching method is:

$$\beta = \left(\frac{1}{n}\right) \sum_{i \in EYEP_i=1} [Y_i - \sum_{j \in EYEP_i=0} w(i,j) Y_j] \qquad (2)$$

Where n is the number of intervention group observations, $EYEP_i$ is an indicator for whether a child (primary caregiver) was assigned to the intervention group, w(i,j) is the weight placed on the jth potential control group observation in constructing a comparison for the ith intervention group observation, and Y_i and Y_j

are outcomes for children (primary caregivers) in the intervention and control groups respectively. Standard errors are calculated using a bootstrap method with 400 replications.

Matching of children in the intervention group to children in the control group is implemented using a Propensity Score Model (PSM) approach. Essentially this involves matching intervention and control group observations on the basis of their predicted probability of assignment to enrolment in EYEP.

Stage one of the PSM approach is to estimate a probit model for whether a child (or primary caregiver) is assigned to be enrolled in EYEP. Covariates included in the model are described below. A balancing test is used to find an appropriate functional form of the probit model for participation in EYEP (see Dehejia and Wahba, 1999, 2002, and Smith and Todd, 2005). The 'balancing test' is a test of whether, after conditioning on the predicted probability of program participation, there is a significant difference between the value of any matching variable for children in EYEP and the control group.

Stage two of the PSM is to match intervention and control group observations. This is done using results from the PSM probit models. Each intervention group observation is matched to a weighted average of control group observations in a 5% confidence interval (caliper method), and a difference in outcome is calculated. Matching is undertaken using the linear predicted score from PSM, which is preferred to the predicted probability as this allows symmetry in selection of control observations using the caliper method. Re-sampling of control group observations across different intervention group observations is allowed. Weights for control observations are derived using kernel weighting:

$$w(i,j) = G^{ij} / [\sum_{j \in EYEP_i=0} G^{ij}] \qquad \text{(3a); and}$$

$$G^{ij} = G[\frac{X_i\hat{\delta} - X_j\hat{\delta}}{a_{5\%}}] \qquad \text{(3b)}$$

Where G^{ij} is the kernel for the ith intervention group observation and jth control group observation, $X_i\hat{\delta}$ and $X_j\hat{\delta}$ are linear predicted scores for the respective intervention group and control group observations, and $a_{5\%}$ represents the use of a 5% confidence interval bandwidth around the linear predicted score for the intervention group observation. The biweight kernel is used.

For the matching method to be a valid estimator of the impact of EYEP, it is sufficient that (Rubin, 1979):

(a) Conditional Independence Assumption (CIA) - Conditional on a set of observable variables, participation in treatment is unrelated to outcomes in the absence of treatment; and

(b) Common support assumption - For each possible combination of observable variables there is a non-zero probability of non-participation.

The CIA effectively requires that matching between intervention and control group observations should be conditional on all variables that affect both enrolment in EYEP and outcomes in the absence of EYEP (Rubin and Thomas, 1996; Stuart, 2010). Our justification for why the CIA is likely to hold is twofold: first, that there has been random assignment to the intervention and control groups; and second, that the matching covariates included are key variables likely to have affected assignment or the outcome being analysed (see discussion below).

The common support assumption is implemented through our approach of matching each EYEP participant to a weighted average of control group observations within a window around the propensity score of the EYEP participant. For the main specifications of the matching method that we estimate, all available EYEP participants can be matched.

The matching method has several advantages over regression methods. It does not impose the restriction of linear function form, and estimates an impact equal to the average effect of treatment on treated. As well, while not solving the common support problem, the matching method makes explicit the common support from which the program impact is identified, thereby facilitating appropriate interpretation of estimates of program impacts. However, the propensity score matching method has been found to perform poorly with small sample sizes, with a major problem being that in small samples the effect of increased variance dominates the effect of the method in reducing bias (Zhou, 2004).

What covariates to control for?

The statistical theory underlying the matching method establishes that in order to satisfy the CIA it is necessary to include as covariates all variables known to be related to assignment to the intervention/control groups or to the outcome (Rubin and Thomas, 1996; Stuart, 2010). Uncertainty about which variables determine assignment and outcomes, however, means that choice of covariates for matching involves judgement. In the case of analysis of EYEP, the choice of covariates also needs to be guided by the relatively small sample size.

To enable comparisons between findings from the alternative estimation methods, the same sets of covariates have been used with both methods. Our general approach has been to: (a) Choose a preferred specification with a set of characteristics for which there is the strongest 'a priori' evidence of a relation with outcomes or where the analysis of balance at the time of entry to the EYEP trial found strongest evidence of a significant difference between the intervention and control groups; and (b) Choose other specifications to allow sensitivity analysis.

For analysis of the impact of EYEP on IQ and language, the alternative specifications are chosen deliberately to include smaller and larger sets of covariates. A caveat regarding the specification with a smaller set of covariates (referred to in the table below as alternative specification (1)) is that it does not achieve balance between the intervention and control groups. For analysis of other outcomes, the alternative specification is chosen to test sensitivity to including extra covariates such as IQ and language scores at time of entry to the trial.

The full details of the sets of covariates included for each outcome and each specificationin are included in Appendix 6.2: Table.

6.3 Definition of impact of EYEP that is estimated

The impact of EYEP that is estimated is the effect of enrolment in EYEP for children who attended for at least 60 days in first twelve months. This is formally referred to as the impact of 'treatment on the treated' (Borland et al., 2005, pp.89-90).

Other impacts of potential interest are: (a) The average treatment effect – The average impact of EYEP if all children assigned to the intervention group had attended the program over the first twelve months; and (b) The average effect of intention to treat – the average impact of EYEP across all children assigned to the intervention group regardless of whether they attended the program in the first twelve months.

Estimating the average treatment effect for EYEP is highly problematic. To estimate this effect would require knowledge of the impact of participation in EYEP for children (or primary caregivers) who never commenced in the program, or who did attend during the first twelve months but did not provide data. It might be thought that this problem could be overcome by reweighting outcome data from children (or primary caregivers) in the intervention and control groups who did provide that data at twelve months in order to adjust for the children for whom the data are not available. Suppose, for example, that at the time of entry to the trial the intervention group has four children – two males and two females; but that by the data collection at twelve months one of the males is no longer in the trial. One solution might seem to be to reweight the remaining male in the intervention group at double the weight of the females. The problem is that, for this approach to provide a valid estimate of the average treatment effect, it is necessary that the outcome for the male who left the trial would have been the same as the male who remains in the trial. But this is highly unlikely. Instead, the fact of having exited the trial reveals that the male who exited is likely to differ from the male who did not exit. Hence, reweighting would lead to a biased (probably upward biased) estimate of the average impact of EYEP.

Estimating the average effect of intention to treat for EYEP does seem feasible, and will be the subject of future research. Note that, provided greater exposure to EYEP is associated with larger impacts from the program, the average effect of intention to treat will be lower than the average effect of treatment on the treated that has been analysed for this report.

Appendix 6.2 Table: Matching covariates by outcome and specification (DV= dummy variable):

Outcome	Preferred specification	Alternative specification (1)	Alternative specification (2)
IQ; Language	Gender; Age at twelve months IQ test; Difference in age between IQ tests at time of entry to trial and twelve months; DV for carer age 25-34; DV for carer age 35+; DV for whether carer has post-school qualification; DV for K6 category Medium; DV for K6 category High; DV for whether carer immigrant; DV for whether English is main language at home; DV for whether both parents present at consent meeting; IQ score at time of entry to trial; Language score at time of entry to trial.	DV for whether carer immigrant; DV for whether both parents present at consent meeting DV for whether child ever breastfed; IQ score at time of entry to trial; Language score at time of entry to trial; Motor skills score at time of entry to trial.	Gender; Age at twelve months IQ test; Difference in age between IQ tests at time of entry to trial and twelve months; DV for carer age 25-34; DV for carer age 35+; DV for whether carer has post-school qualification; DV for K6 category Medium; DV for K6 category High; DV for whether carer immigrant; DV for whether English is main language at home; DV for whether both parents present at consent meeting; IQ score at time of entry to trial; Language score at time of entry to trial; DV for whether child ever breastfed; DV for risk factor – alcohol or substance abuse; DV for risk factor – disability/complex medical issues; DV for risk factor – harsh, inconsistent discipline, neglect or abuse.
DECA; BITSEA/ CBCL	Gender; Age at twelve months IQ test; DV for carer age 25-34; DV for carer age 35+; DV for whether carer has post-school qualification; DV for K6 category Medium; DV for K6 category High; DV for whether both parents present at consent meeting.	Gender; Age at twelve months IQ test; DV for carer age 25-34; DV for carer age 35+; DV for whether carer has post-school qualification; DV for K6 category Medium; DV for K6 category High; DV for whether carer immigrant; DV for whether English is main language at home; DV for whether both parents present at consent meeting; DV for whether child ever breastfed; IQ score at time of entry to trial; Language score at time of entry to trial; DV for risk factor – harsh, inconsistent discipline, neglect or abuse.	
Parenting Daily Hassles	Same as DECA; plus DV for multiple children at consent meeting.	Same as for DECA; plus DV for multiple children at consent meeting.	
K6	Same as DECA; plus DV for multiple children at consent meeting and K6 score at time of entry to trial.	Same as DECA; plus DV for multiple children at consent meeting and K6 score at time of entry to trial.	
HOME	Gender; Age at twelve months IQ test; DV for carer age 25-34; DV for carer age 35+; DV for whether carer has post-school qualification; DV for K6 category Medium; DV for K6 category High; DV for whether both parents present at consent meeting.	Gender; Age at twelve months IQ test; DV for carer age 25-34; DV for carer age 35+; DV for whether carer has post-school qualification; DV for K6 category Medium; DV for K6 category High; DV for whether both parents present at consent meeting; DV for whether child ever breastfed; DV for whether carer is immigrant; DV for risk factor – rejection of child.	

Appendix 7

Extra results

Appendix 7.1: Table - Impact of enrolment in EYEP for 12 months – Sensitivity analysis

	Outome	Method/Sample	EYEP mean (12 months)	EYEP impact	1-tail p-value	2-tail p-value	Observations (Intervention/ Control)
(1)	12-month Language	Regression – Alternative specification (1)	92.70	1.145	0.344	0.688	50/43
(2)	12-month Language	Regression – Alternative specification (2)	92.70	0.397	0.453	0.894	50/43
(3)	12-month Language	Matching – Preferred specification	90.57	-0.701	0.487	0.914	50/43
(4)	Change in Language from entry to trial to 12 months	Matching – Preferred specification	2.339	0.489	0.429	0.858	50/43
(5)	12-month Language	Regression – Preferred specification - All intervention group	90.57	-0.437	0.423	0.867	56/43
(6)	12-month Language	Regression on matched sample – Preferred specification	92.70	0.803	0.428	0.856	50/43
(7)	Protective factors	Regression – Alternative specification (1)	46.90	1.967	0.171	0.346	49/45
(8)	Protective factors	Regression – Preferred specification - All intervention group	46.63	1.952	0.163	0.328	54/45
(9)	Social and emotional (10% threshold)	Regression – Alternative specification (1)	21.28	-6.75	0.250	0.503	47/44
(10)	Social and emotional (25% threshold)	Regression – Preferred specification - All intervention group	20.75	-13.16	0.085	0.175	53/44
(11)	Psychological distress	Regression – Alternative specification (1)	15.34	0.096	0.463	0.930	47/44
(12)	Psychological distress	Regression – Preferred specification - All intervention group	15.18	-1.234	0.126	0.252	51/44
(13)	Parenting daily hassles - Frequency	Regression – Alternative specification (1)	45.60	1.486	0.293	0.594	46/42
(14)	Parenting daily hassles - Frequency	Regression – Preferred specification - All intervention group	45.81	0.771	0.371	0.754	50/42
(15)	Parenting daily hassles - Intensity	Regression – Alternative specification (1)	47.88	3.911	0.166	0.332	46/42
(16)	Parenting daily hassles - Intensity	Regression – Preferred specification - All intervention group	47.87	4.077	0.136	0.271	50/42
(17)	HOME	Regression – Alternative specification (1)	73.79	1.751	0.279	0.556	36/37
(18)	HOME	Regression – Preferred specification - All intervention group	72.19	0.237	0.491	0.979	40/37

Note: Unless otherwise stated the intervention group for all estimates was children who attended for at least 60 days in the first twelve months.

Appendix 7.2: Table - Impact of enrolment in EYEP for 12 months – Children who attended for at least 60 days – With adjustments to Bayley Scales Cognitive and Language measures by blind coder

	Outcome	EYEP mean (12 months)	EYEP impact	1-tail p-value	2-tail p-value	Number of observations (Intervention/ Control)
	Children's development					
(1)	IQ	97.55	3.773	0.067	0.125	50/43
(2)	Language	90.75	0.617	0.417	0.820	50/43
(3)	Protective factors	46.90	2.040	0.153	0.309	49/45
(4)	Social and emotional (Percent in clinical range; Below norm for bottom 10% of population)	21.28	-12.92	0.093	0.192	47/44
	Primary caregiver					
(5)	Psychological distress	15.34	-1.025	0.184	0.364	47/44
(6)	Parenting daily hassles - Frequency	45.60	0.840	0.368	0.742	46/42
(7)	Parenting daily hassles - Intensity	47.88	4.639	0.109	0.218	46/42
(8)	Home environment	73.79	1.670	0.276	0.548	36/37

Note: The EYEP impacts in rows (1)-(3) and (5)-(8) are the estimated impacts of attending EYEP from an OLS regression. The EYEP impact in row (4) is the marginal impact on the percent of children below the 10% threshold on the social-emotional measure estimated from a probit model.

Changing the Life Trajectories of Australia's Most Vulnerable Children

Report No. 2
The first twelve months in the Early Years Education Program: An initial assessment of the impact on children and their primary caregivers