MELBOURNE INSTITUTE
of Applied Economic
and Social Research

Changing the Life Trajectories of Australia's Most Vulnerable Children

EYEP replication trial report No. 2

Can the Early Years Education Program (EYEP) be Successfully Replicated? Early Evidence from the First 12-month Outcomes of Participants

April 2026

THE UNIVERSITY OF MELBOURNE
MELBOURNE INSTITUTE

Parkville INSTITUTE

EYEP Replication Trial Report 2

Can the Early Years Education Program (EYEP) Be Successfully Replicated? Early Evidence from the First 12-month Outcomes of Participants

Yi-Ping Tseng[1], Nichola Coombs, Daniel Fischer, Karina Tetkowski
Melbourne Institute of Applied Economic and Social Research, University of Melbourne

Corresponding author: Yi-Ping Tseng (y.tseng@unimelb.edu.au)

Acknowledgement

The Early Years Education Program (EYEP) was initiated by Kids First, previously the Children's Protection Society (CPS), an independent not-for-profit child and family services organisation based in the north- east of Melbourne which was founded in 1896. The program was designed and implemented by CPS in collaboration with Associate Professor Brigid Jordan AM and Dr. Anne Kennedy AM. The EYEP RCT was conducted by a multidisciplinary team led by Professor Jeff Borland AO from the University of Melbourne. A full list of contributors is included in the RCT reports which can be downloaded from https://melbourneinstitute.unimelb.edu.au/research/education/early-years-education.

The EYEP replication trial is led by Parkville Institute, a registered charity established to support infants and young children living with significant family stress and social disadvantage to enter school as confident and successful learners who are developmentally and educationally equal to their peers. The outcome evaluation of this trial (including this report) is conducted by the Melbourne Institute of Applied Economic and Social Research (MI) under the service agreement with Parkville Institute.

This study is funded by the Australian Government Department of Education; Victorian Department of Education; Paul Ramsay Foundation; The Bryan Foundation; The Berg Family Foundation; and The TDM Foundation; as well as in-kind contributions from the Parkville Institute and Melbourne Institute. The findings and views in this report, however, are those of the authors and should not be attributed to the funders, Parkville Institute, or the Melbourne Institute.

This report utilised data collected from the EYEP replication trial and the EYEP RCT. We are indebted to the children and families who have been willing to participate in the EYEP RCT and the EYEP replication trial. We express our gratitude to staff members at the replication centres for their expertise, care, professionalism, and assistance in data collection. We are grateful for Professor Jeff Borland's insightful discussions. We are thankful to Taylor Ey, Dr. Sarah Fraser and Dr Jane Sheehan, and Tanya Gupta for their contributions in collecting and/or processing data for the replication trial.

This study also uses part of questionnaires from Growing Up in Australia: Longitudinal Survey of Australian Children (LSAC). The LSAC questionnaires are the property of the Commonwealth Department of Social Services (DSS). LSAC is being undertaken in partnership between DSS, Australian Bureau of Statistics (ABS), and Australian Institute of Family Studies (AIFS). The findings and views reported in this study should not be attributed to DSS, ABS, or AIFS.

[1] Yi-Ping Tseng and Nichola Coombs are the authors of this report. Daniel Fischer is the data base manager and Karina Tetkowski is a research coordinator.

Table of Contents

Executive Summary

Background and purpose

The Early Years Education Program (EYEP) is an intensive, centre-based early childhood education and care (ECEC) program specifically designed to support children experiencing significant family stress and social disadvantage. A randomised controlled trial (RCT) conducted between 2011 and 2018 demonstrated that EYEP produces large and statistically significant impacts on children's development, with improvements of 7.6 IQ points, 6.8 language points, and 6.2 points reduction in behaviour problem scores at 36 months (Tseng et al., 2022).

While the RCT provided proof of concept, translating this evidence into widespread impact requires demonstrating that the program can achieve comparable outcomes when implemented in diverse real-world contexts. The EYEP replication trial tests whether the model can be successfully delivered across three children's centres—in metropolitan and regional Victoria, and metropolitan Queensland—while maintaining effectiveness.

Report 1 (the baseline report, released June 2025) describes participants who were referred to the EYEP replication trial and completed the consent interview before 30 September 2024. This report has two purposes: (1) to update trial progress and the profile of all children referred up to 30 June 2025; and (2) to examine whether developmental trajectories of children in replication trial from baseline to 12-month follow-up are comparable with those observed in the original RCT. These analyses provide early evidence on whether the EYEP can be successfully replicated.

The replication trial

The three replication centres opened in staggered sequence: Centre 1 (metropolitan Victoria) in January 2023, Centre 2 (metropolitan Queensland) in August 2023, and Centre 3 (regional Victoria) in January 2024. Parkville Institute provides comprehensive implementation support to maintain program quality and fidelity, including regular provision of evidence-based advice, professional development, and mentoring. The University of Melbourne conducts the independent outcome evaluation.

Eligibility criteria mirror the EYEP RCT: children under three years at referral, experiencing two or more family or parenting risk factors, engaged with family or child protection services, and requiring early education as part of their care plan. An additional requirement is eligibility for the Australian Government's Additional Child Care Subsidy (child wellbeing).

Recruitment and population reached

By 30 June 2025, 231 eligible children had been referred across the three centres. Of these, 168 children (73%) consented to participate, comparable to the RCT consent rate of 77%. However, conversion from consent to commencement of regular attendance varied substantially across replication centres: 49% at Centre 1, 80% at Centre 2, and 48% at Centre 3, compared to 81% in the RCT.

The replication trial successfully engaged highly vulnerable families. Children were referred at an average age of 17 months, with families presenting with an average of 5.9 risk factors. The most prevalent risk factors were parental mental health concerns (79.2%), family violence (67.3%), and attachment/relationship concerns (50.6%).

Outcome measures

This report focuses on a set of child and carer outcomes aligned with the RCT reporting framework, including:

- Child cognitive development (IQ) and language development assessed using Bayley Scales of infant and toddler development (BSID-IV) and Wechsler preschool and primary scale of intelligence (WPPS-IV).
- Child total protective factors related to resilience (social and emotional health) measured using Devereux Early Childhood Assessment Program (DECA)
- Child social–emotional problems in clinical range and in problem range (i.e. below the bottom 10^{th} and 25^{th} percentiles of population norm, respectively) derived from Brief infant toddler social emotional assessment (BITSEA) and Child behaviour checklist (CBCL).
- Primary carer outcomes: psychological distress (K6), Parenting Daily Hassles (frequency total score and intensity total score)

Analytic sample and empirical method

Analysis focuses on children who attended for at least 60 days in the first 12 months and had complete assessment data at baseline and 12 months. This restriction serves two purposes: ensuring comparability with the RCT treatment group, which used the same attendance threshold, and focusing on children with meaningful program exposure.

To benchmark replication outcomes against the RCT treatment group, the report presents:

- Average within-person change in outcomes between baseline and 12-month follow-up.
- Difference-in-differences estimates comparing replication-trial participants with the RCT treatment group, estimated using a regression-adjusted matching method to control for family characteristics and risk factors.
- Sensitivity analyses that vary the estimation method and the set of covariates.

Baseline characteristics: A more vulnerable population

Children in the replication trial entered with substantially greater developmental challenges than the RCT treatment group across all assessed domains. Average baseline IQ was 80.1 in the replication trial compared to 90.0 in the RCT, representing a difference of approximately two-thirds of a standard deviation. Language development showed a similar pattern, with mean baseline scores of 77.4 compared to 86.3 in the RCT. Social-emotional development (measured at baseline using Bayley Social Emotional Scale questionnaires) was also markedly lower, with average baseline scores of 83.5 compared to 95.8 in the RCT.

The prevalence of developmental delay was substantially higher in the replication trial. Nearly three-quarters of replication trial children (73.2%) scored below 85—indicating developmental delay—in at least one domain (cognitive or language), compared to one-third (33.8%) in the RCT. Using standard classifications, 24.7% of replication trial children had significant delay (IQ <70) at baseline, compared to only 4.4% in the RCT.

Families in the replication trial also presented with significantly greater cumulative risk burden, with an average of 5.92 referrer reported risk factors compared to 4.77 in the RCT. However, the risk

profiles differed in character. The replication trial showed higher rates of social isolation (38.7% versus 25.5%) and child medical complexity (40.5% versus 26.4%), while the RCT showed higher rates of parenting capacity concerns such as harsh discipline or neglect (41.3% versus 8.9%). These differences may reflect variation in referral pathways: the RCT received more referrals from statutory child protection services, while the replication trial receives more referrals from health professionals, particularly maternal and child health nurses (55.4% of referrals).

After 12 months: Comparable cognitive and language gains, superior total protective factor gains

Children in the replication trial showed average changes of +4.6 points in IQ, +7.2 points in language, and +11.6 points in total protective factor standard scores. The observed change in IQ was not statistically significant owing to substantial variation in individual change scores across children.

Compared with children in the RCT who had similar characteristics and risk factors using a regression adjusted matching method, average changes in IQ and language were comparable. Outcomes slightly favoured the replication trial — with an average IQ gain of 2.9 points and language gain of 0.5 points higher than the RCT — but these differences were not statistically significant. The replication trial showed a significant average increase in total protective factor scores that was 14 points higher than the RCT average.

At 12-month follow-up, a smaller proportion of children in the replication trial were in "possible problem range" for social–emotional development (scores below the 25th percentile of population norms): 31 per cent compared with 53 per cent in the RCT. After accounting for characteristics and risk factors, the difference between the two trials increased to 51 percentage points. The replication trial and the RCT had a similar proportion of children in "clinical range" of social emotional problems (below the 10th percentile of population norms).

Sensitivity analyses indicated the findings were robust. However, some families in which caregivers had severe psychological distress were less likely to complete the 12-month protective factor and social–emotional assessments even though their children are still participating in the program; consequently, the apparently superior outcomes on these measures may not fully reflect results for the most distressed families.

Primary carer outcomes: similar between the replication trial and the RCT

There was no significant difference between the replication trial and RCT in how caregiver psychological distress changed over 12 months, indicating comparable outcomes. Similarly, the frequency of daily parenting hassles did not differ between groups.

For the intensity of parenting hassles—how stressful these daily challenges felt—one analytical approach showed a significant difference favouring the replication trial (9.4 points lower), though this was not significant when using a different analytical approach. This finding should therefore be interpreted cautiously.

Interpretation and implications

The 12-month findings provide encouraging early evidence about the replication trial's trajectory. Despite entering the program with substantially greater developmental vulnerability, children in the replication trial achieved cognitive and language gains at 12 months that were comparable to those seen in the RCT treatment group. The replication trial's larger gains in total protective factors and the reduced proportion of children with social–emotional difficulties are particularly noteworthy.

The substantial early increases in protective factors are especially encouraging because these foundational capacities — self regulation, initiative and attachment/relationships — are expected, under EYEP's theory of change, to support subsequent cognitive and language development and social emotional health. Strengthening these protective capacities early therefore offers a plausible pathway to sustaining developmental gains over time.

Several implications for implementation follow from these results. The replication trial is reaching children with greater developmental vulnerability and medical complexity than the RCT, which calls for higher levels of individualisation, closer coordination with allied health services, and potentially additional inclusion supports beyond EYEP's enhanced staff–child ratios. Different referral patterns observed in the replication trial suggest the need to strengthen connections with statutory child protection services to ensure the program reaches families across the full spectrum of intended need.

There is also substantial variation in recruitment and retention across centres. Retention rates at the two Victorian centres are much lower than at the Queensland centre even after controlling for families' risk factors and referral pathways. This retention challenge underscores the importance of flexible, persistent and trauma informed approaches to supporting families through barriers to sustained participation.

Finally, the comparable 12month outcomes indicate that Parkville Institute's implementation support model is functioning effectively in early delivery for those children who have participated in the program for 12 months or longer. Further development in support needed to overcome the retention challenges will be beneficial for future implementation.

Next steps

The replication trial continues to follow participants through their second and third years of program participation. The final report (Report 3) will present complete evaluation findings, including updated 12-month outcomes incorporating late recruitments and 24-month outcomes for the cohort available.

The 24-month assessment will be particularly important for determining whether the pattern of strengthening impacts observed in the RCT is replicated across the three sites. Integration of fidelity assessment findings with outcome data will provide comprehensive understanding of factors supporting successful replication.

This evidence will inform recommendations for targeted scale-up of the EYEP model as part of Australia's early childhood education and care service system and contribute to broader knowledge about effective approaches to replicating evidence-based early childhood interventions.

1. Introduction

Early childhood represents a critical period when adversity and trauma can profoundly impact development and learning, with effects lasting throughout the life course. In Australia, nearly 50,000 preschool-aged children receive child protection services annually, and recent data from the Australian Early Development Census (AEDC) reveal that developmental vulnerability at school entry has reached its highest level since data collection began in 2009 (AIHW, 2024; AEDC, 2024). In 2024, 23.5% of Australian children were developmentally vulnerable on at least one domain, and 12.5% were vulnerable on two or more domains—increases that reverse the positive trends observed in earlier census cycles.[2]

Research has demonstrated that early exposure to adverse experiences can disrupt brain development, particularly affecting emotional regulation and learning capacity (Perry, 2002; Shonkoff, 2012). The impact of adversity is cumulative: Australian longitudinal studies using data from the Longitudinal Study of Australian Children (LSAC) show that children exposed to four or more family and environmental risk factors—such as poverty, parental mental health issues, housing instability, and family violence—are substantially more likely to experience developmental delays and be vulnerable across multiple AEDC domains by school entry (Edwards & Baxter, 2017; Guy et al., 2016). Data linkage studies combining AEDC records with health and child protection administrative data confirm that cumulative risk operates through both direct effects on child development and indirect effects through reduced access to early childhood services (Pilkington et al., 2019). Without early intervention, these developmental disruptions often become entrenched, leading to persistent difficulties in education, employment, and health (Heckman, 2008).

More than one in four Australian children live in 'childcare deserts'—areas with limited or no access to ECEC places—and access shortages are most acute in regional areas, outer suburbs, and lower-income communities where vulnerability is highest (Mitchell Institute, 2023). In recognition of the importance of Early Childhood Education and Care (ECEC) on child development, in February 2025, Parliament amended the Child Care Subsidy legislation to introduce a 'Three Day Guarantee,' ensuring all families can access a minimum of 72 hours per fortnight of subsidised early childhood education and care (ECEC) regardless of parental work or study requirements. Additional investments include the $1 billion Building Early Education Fund, explicitly targeting areas of undersupply with a focus on communities experiencing higher levels of vulnerability and disadvantage. The economic imperative for early intervention is substantial: the Cost of Late Intervention report reveals that failing to identify and address health and developmental issues in young Australians cost taxpayers $22.3 billion in 2024—a 47% increase since 2019 (Minderoo Foundation, 2024).

However, while these policies increase access to ECEC, critical questions remain on accessing quality ECEC services for disadvantaged families. Where services are available, general ECEC services may not have sufficient resources to meet the complex needs of children who have experienced trauma or significant family instability (Productivity Commission 2024).

The Early Years Education Program (EYEP) was developed as an evidence-based response to this challenge, providing intensive early education and care specifically designed to address the complex needs of children experiencing significant family stress and social disadvantage. The EYEP randomised

[2] The disruption of ECEC and school foundation year caused by COVID-19 may have contributed to the reversal of the trend.

controlled trial (EYEP RCT) conducted between 2011 and 2018 demonstrated that the program produces large and statistically significant impacts on children's IQ (7.6 points), language (6.8 points), and social-emotional outcomes at 36 months despite the fact that substantial number of total ECEC hours were received by control group children (approximately two-third of treatment group). (Tseng et al., 2022)

While the RCT provided 'proof of concept' that the EYEP model can significantly improve vulnerable children's developmental trajectories, translating this evidence into widespread impact requires demonstrating that the program can achieve comparable outcomes when implemented in diverse real-world contexts. It is well established that innovative social programs often experience 'voltage drop'—a reduction in effectiveness during scale-up—due to factors including challenges in maintaining implementation fidelity, differences in population characteristics, variation in contextual and systemic factors, and organizational capacity constraints (Al-Ubaydli et al., 2019; Durlak & DuPre, 2008). Members of the EYEP RCT research team further developed the replication framework with careful consideration of factors that might affect program effect when scaling a program and conducting the EYEP replication trial. Understanding the conditions under which effective interventions maintain their impact when replicated is essential not only for EYEP but for informing broader efforts to scale evidence-based early childhood programs.

The replication trial is being conducted across three children's centres—in metropolitan and regional Victoria, and metropolitan Queensland—to test whether the program can be successfully implemented in different geographical, jurisdictional, and organizational contexts while maintaining its effectiveness. The study design includes several features that distinguish it from most replication studies in the literature. First, a rich set of data are collected covering six domains of child and primary carer outcomes, family background, and risk factors. Second, unit record data from both the EYEP RCT and the replication trial enable direct econometric comparisons of developmental trajectories, taking participant characteristics into account. Third, comprehensive documentation of the implementation process and fidelity challenges and enablers allow cross-referencing of outcome findings with implementation quality. The findings from this replication trial will inform recommendations for targeted scale-up of the model as part of the suite of early childhood education and care services in Australia and contribute to policy development regarding interventions for vulnerable children. The research also contributes to the broader implementation science literature on service delivery of targeted early childhood interventions.

This report is the second of the three outcome evaluation reports planned for this project. The replication trial is still in progress, and the recruitment of participants has not been completed at the time of producing this report. The report serves dual purposes: (1) providing an updated profile of participants at replication trial entry including late recruitments and report any significant difference between earlier cohort (those reported at baseline) and later cohort. (2) Examining developmental progress between baseline and 12-month assessment for children whose participation has passed one year mark. Understanding whether children in replication trials are achieving developmental trajectories comparable to the RCT treatment group provides early evidence of whether the program is on track to replicate the outcomes of RCT.

This report is organised as follows. Section 2 describes the EYEP model, summarises evidence from the original RCT with particular attention to the temporal pattern of impacts, and provides an overview of the trial progress. Section 3 describes the scope of this report and data sources. Section 4 presents updated baseline characteristics for all children recruited before 30 June 2025 and describes the 12-month follow-up sample. Section 5 examines developmental outcomes after 12 months,

including child development measures, primary carer outcomes, and detailed comparisons with the RCT treatment group's progress at 12 months. Section 6 discusses potential implications for program implementation, and Section 7 outlines next steps for the ongoing evaluation.

2. The Early Years Education Program and evidence from the randomised controlled trial

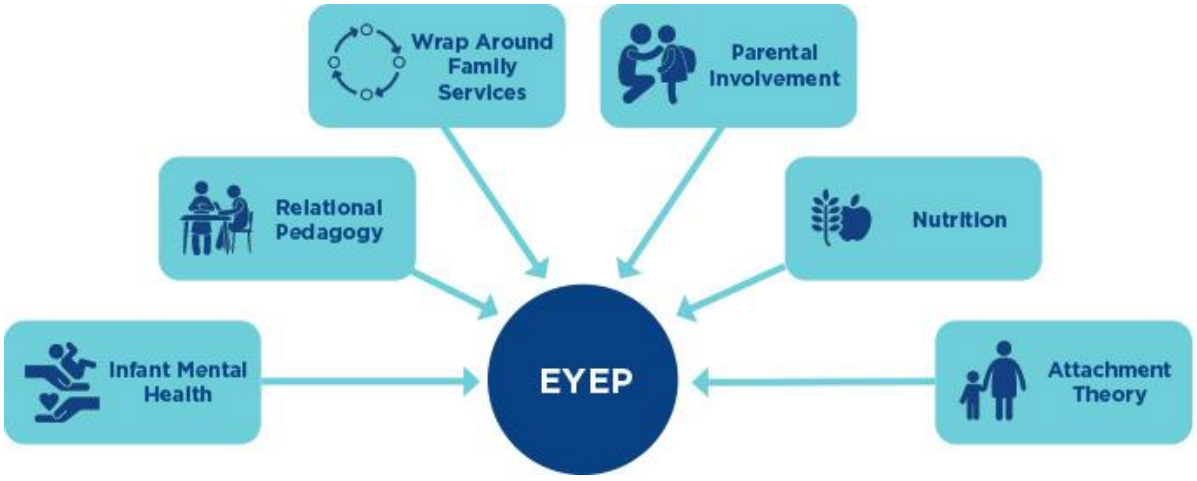
2.1 The Early Years Education Program model

Early Years Education Program (EYEP) was initiated by Kids First, previously the Children's Protection Society (CPS), an independent not-for-profit child and family services organisation based in the northeast of Melbourne which was founded in 1896. The program was designed and implemented by CPS in collaboration with Associate Professor Brigid Jordan AM and Dr. Anne Kennedy AM. The structural and process quality elements of the model are described in Jordan and Kennedy (2019).

This innovative centre-based early years education and care program specifically targets children experiencing significant family stress and social disadvantage (Jordan and Kennedy, 2019). With a dual focus, the program addresses both the consequences of significant family stress on brain development and learning deficiencies. The objective is ensuring vulnerable children enter formal schooling developmentally and educationally on par with their peers, possessing the knowledge, skills, and attributes necessary for successful ongoing learning.

In a standalone childcare setting, the program delivers holistic approach to care and education grounded in relational pedagogy, attachment theory, and trauma-informed care. It involves direct child intervention to address identified developmental and learning needs, mitigate adverse experience impacts, and reduce risk factors that can produce poor developmental and learning outcomes. The care approach emphasises attachment and trauma-informed practice, highlighting the importance of establishing strong, responsive relationships with children. Through the primary educator model, the program aims to foster significant caregiver attachments, particularly critical for children whose home environments may feature disrupted or compromised attachment relationships.

Figure 2.1 The EYEP model



Source: Tseng Y. and Borland J (2020)

Australia's National Early Years Learning Framework, Belonging, Being and Becoming (AGDE V2.0, 2022) and the National Quality Standard (ACECQA, 2018), establish the foundational principles and practices for the EYEP framework. Children's learning within EYEP is supported through individualised

goals developed collaboratively with families and implemented through intentionally planned play-based experiences spanning all developmental domains and the five National learning outcomes.

EYEP provides children with a program for five hours a day, five days per week, 50 weeks annually, over three years. The program maintains high staff-to-child ratios (1:3 for children under three years and 1:6 for older children) and employs qualified, experienced early childhood teachers and educators. An innovative program feature is its multidisciplinary leadership team, comprising a full-time centre director, full-time pedagogy leader, part-time on-site infant mental health consultant, and part-time on-site family practice consultant. These professionals collaborate within an adaptive leadership framework (Heifetz et al., 2009) to support and guide educators' work using a multidisciplinary approach and children's best interest framework.

Family engagement constitutes a critical EYEP component. The program encourages active parental participation in children's education through meetings with their child's primary educator every 12 weeks to develop shared learning goals. Families are welcome to include support people in these meetings. The centre family practice consultant may attend these meetings while also supporting families' access to local health and community services and addressing service usage barriers such as affordability or interpersonal dynamics.

EYEP's comprehensive care and education model is supported through monthly meetings attended by Parkville Institute's senior advisors and the centre leadership team to ensure program fidelity. This structure ensures continuous fidelity monitoring and ethical, effective implementation to meet at-risk children's needs. Through this rigorous, multidisciplinary approach, EYEP aims to provide vulnerable children with educational foundations necessary to start school developmentally and educationally equal to their peers.

2.2 Evidence from the randomised controlled trial

A multidisciplinary team led by Professor Jeff Borland AO from the University of Melbourne conducted the EYEP randomised controlled trial (EYEP RCT). This team brought expertise spanning economics, early childhood education, infant mental health, social work, evaluation, strategy, policy development, and management. Complete contributor lists are available in EYEP reports at https://melbourneinstitute.unimelb.edu.au/research/education/early-years-education.

Between 2011 and 2016, the EYEP RCT recruited 145 children from 99 families. Eligibility required children to be under three years at referral, experience two or more family or parenting risk factors, and have early childhood education included in their care plan. Families received random assignment to either the treatment group (offered EYEP for three years) or the control group (receiving usual care—a combination of care by parents/guardians or other ECEC services).

The trial successfully engaged a highly vulnerable population, with 70% of participating children experiencing four or more risk factors. The most prevalent risk factors included parental mental illness, alcohol or substance use, family violence, and harsh or inconsistent parenting. Participating families demonstrated substantially greater disadvantage than families with low socioeconomic backgrounds in the Longitudinal Study of Australian Children (LSAC). For instance, RCT-participating families experienced higher rates of financial crisis (32% versus 18.8%) and children's primary carers showed elevated rates of psychological distress (25.8% versus 4.45%) (Tseng et al., 2017).

EYEP's impact on children's IQ emerged within the program's first 12 months and continued strengthening over subsequent two years. At 36-month follow-up, the estimated average IQ impact

approximates one-half of the standard deviation, twice as large as the average impact estimates from early years demonstration programs in the United States (Duncan and Magnuson, 2013). The program effects on language grew throughout the trial, achieving statistical significance at 24 months and reaching its strongest point at 36 months. Social-emotional development impacts became large and significant at 24 months and were sustained at 36 months (Tseng et al., 2022). For some outcomes, program impacts derived from both improvement of treatment group as well as deterioration of control group standardised outcomes.

At the 36-month assessment, children in the EYEP group demonstrated greater average improvement of 7.6 IQ points and 6.8 language points compared to children in the usual care group, alongside reduced behaviour problems (6.1 points, equivalent to 0.6 standard deviation). IQ and language impacts were substantially larger for children whose initial scores fell below 90, achieving average impacts of 13.6 IQ points and 12.7 language development points.

By program completion (36 months), average IQ and language scores for intervention group children reached 99.6 and 99.5, respectively, aligning closely with the general population average of 100. These results provide strong evidence of EYEP's capacity to significantly alter developmental and learning trajectories for highly vulnerable children.

The temporal pattern of impacts observed in the EYEP RCT provides important context for interpreting findings from the replication trial's 12-month assessment. The progressive emergence of impacts over time suggests that the 12-month assessment represents an early indication of whether the replication is on track rather than the conclusion of replication trial outcomes.

3. The replication trial study design

Members of the original EYEP RCT research team are conducting this replication trial. With Australian Government funding, Parkville Institute (PI) was established as a registered charity dedicated to supporting young children and infants experiencing significant family stress and social disadvantage. The founding directors, Associate Professor Brigid Jordan AM and Dr. Anne Kennedy AM, were both members of the original RCT team. PI holds overall responsibility for the replication project and program practice. Melbourne Institute (MI) team members Associate Professor Yi-Ping Tseng and Dr. Nichola Coombs, who were also part of the original RCT team, are responsible for conducting outcome evaluations to determine whether the replication trial produces outcomes comparable to those achieved in the EYEP RCT. This organisational structure leverages both teams' expertise while preserving the independence of the outcome evaluation.

The replication trial study design is outlined briefly below. More details were documented in the baseline report, Tseng et al. (2025).

3.1 Research questions

The replication study will examine three key research questions:

1. Can referral and recruitment processes successfully identify and engage the target population across diverse communities?

2. Will the intervention produce improvements in children's developmental outcomes comparable to those observed in the original RCT?

3. Is it possible to replicate the model with fidelity? What factors enable model fidelity and what challenges arise?

The Melbourne Institute outcome evaluation team addresses Questions 1 and 2 while contributing insights into Question 3 through data collection processes and outcome data analyses. Parkville Institute formally conducts the fidelity assessment to address Question 3. The outcome evaluation delivers independent assessment to produce evidence of program effects, while the fidelity assessment provides essential complementary information regarding implementation quality. These two evaluation streams together yield comprehensive understanding of program effects and the implementation factors influencing them.

The University of Melbourne Human Research Ethics Committee has approved this study (ID 23608 and 23994).

3.2 Recruitment of participants

Eligibility criteria for the replication trial mirror those used in the original EYEP RCT, targeting children who are:

- Between 0 and 3 years of age at referral
- Assessed as experiencing two or more risk factors according to the Department of Human Services 2007 Best Interest Case Practice Model (risk factors listed in Appendix A)
- Currently engaged with family services or child protection services
- Identified as requiring early education as part of their care plan

Recruitment follows a community driven approach, with each centre holding primary responsibility for community engagement and referral sourcing, while PI delivers ongoing guidance to centres regarding outreach and recruitment strategies, ensuring alignment with program objectives and effective reach to children eligible for this service.

3.3 Data collection

Data collection encompasses multiple domains:

- Standardised developmental assessments conducted by qualified clinical researchers, evaluating cognitive development, language skills, parent-child relationships, and educator-child relationships.
- Primary carer interviews conducted by research coordinators utilizing parent questionnaires and standardised assessment tools. These instruments collect information on family characteristics, service utilization, child behaviour and emotional functioning, child resilience, parental psychological distress, and daily parenting experiences.
- Educator assessments of children's behaviour, emotional functioning, and resilience using standardised assessment tools.
- Centre-collected attendance records for engagement pattern analyses.
- Administrative records supporting PI's program fidelity assessment.

Child and family outcome data are collected at three critical junctures in each child's program participation journey: at baseline (within three months of enrolment), at 12 months post-enrolment, and at twenty-four months post-enrolment. All data collection personnel are qualified clinical researchers experienced in engaging with young children experiencing adversity and their parents, ensuring data collection quality.

3.5 Analytical framework

This research project aims to assess whether the intervention model can be successfully replicated. Rather than employing the conventional approach of quantifying program impact through comparison of program participants with a non-intervention comparison group, we directly examine whether participating children's trajectories are similar to those of the original RCT's intervention group. Comparing an intervention group with a usual-care (non-intervention) group produces program impact estimates that vary according to the amount and types of usual care received, which can be influenced by numerous institutional factors, including the accessibility of mainstream early childhood services across different jurisdictions. Consequently, impact differences cannot be attributed exclusively to replication quality. Comparison with the original RCT treat group enables more precise evaluation of whether comparable outcomes were achieved by the replicated intervention. More detailed discussions of estimation strategies are discussed in section 5.

4. Recruitment progress and the trial participants

The recruitment of trial participants was still on going at the time of preparation for this report. Characteristics of families who completed the recruitment process by September 2025 were presented in baseline report (Tseng et al 2025). In this report, we provide updates of recruitment for all families referred prior to June 2025.[3] As the characteristics of early and late recruitment cohorts are very similar with no statistically significant differences between the two groups, only combined results are presented. Comparisons of the characteristics of the two groups are presented in Appendix C.

4.1 Recruitment progress

The three centres opened in a staggered sequence, beginning with Centre 1 in January 2023, followed by Centre 2 in August 2023, and Centre 3 in January 2024. This phased implementation allowed careful attention by PI to program fidelity and operational capacity at each centre.

Table 4.1: Recruitment Progress

	Centre 1	**Centre 2**	**Centre 3**
Centre open	January 2023	August 2023	January 2024
Maximum num. of children (Licensed places)	42	48	39
Num. of children attending the program (Num of children awaiting commencement)	25(1)	23(0)	18 (0)
Numbers of children attending – including children referred after 30 June 2025	25	24	23

Note: Recruitment was on going at the time of report preparation, this table reports only families who were referred prior 30th June 2025 unless otherwise mentioned. Participating status is recorded as of 15 December 2025.

Recruitment has continued across all three centres. As of 30 June 2025, the centres had been open for between 18 months and 30 months. Table 4.2 presents recruitment progress and retention rates for each centre, alongside comparable data from the EYEP RCT.

In total 231 eligible referrals had been recorded across the three centres, with 168 children consenting to participate (overall consent rate: 73%). Consent rates were similar between Centres 1 and 3 (70% each) and higher in Centre 2 (83%). These consent rates are comparable to the EYEP RCT consent rate of 77% (excluding children referred after a sibling had commenced participation).

Recruitment pace varied across centres, as shown in Table 4.2. Centre 1 enrolled the majority of children in its first year of operation, with enrolment tapering in subsequent years. Centre 2 showed steady enrolment across both years. Centre 3 maintained consistent enrolment across its two years of

[3] This is to allow time for completion of recruitment process and data collection, processing, and analyses.

operation. For comparison, the EYEP RCT enrolled children at a relatively steady pace across its first three years.

Conversion from consent to commencement of regular attendance also varied across centres. Of children who consented, 49% commenced regular attendance in Centre 1, 80% in Centre 2, and 48% in Centre 3. This compares to 81% in the EYEP RCT treatment group. These differences likely reflect a combination of factors including family circumstances (such as housing instability, transport difficulties, and safety concerns), local service system contexts, and the challenges of supporting highly vulnerable families to transition from referral to sustained program participation.

Table 4.2 Recruitment pace and retention rates across centres

	Replication trial			EYEP RCT	
	Centre 1	Centre 2	Centre 3	Treatment	All
Geographical area of centre	Metro. VIC	Metro. QLD	Regional VIC	Metro. VIC	
Duration since opened (up to June 2025)	2 yr 6 m	1 yr 11 m	1 year 6 m		
Num. consented in year 1	51	24	27	18	35
Num. consented in year 2	24	16	21	16	35
Num. consented in year 3	5	n.a	n.a	11	29
Num. consented in year 4-5	n.a	n.a	n.a	27	46
Total num. consented	80	40	48	72	145
Total num. referred (eligible)	114	48	69	n.a.	177
Consent rate	70%	83%	70%	n.a.	80% (82%)
% of consented commenced regular attendance	49%	80%	48%	81% (83%)	n.a.

Note: For comparison purposes, consent rate and percentage commenced regular attendance for EYEP RCT, excluding children referred to the trial after their sibling's commenced participation in the trial[4]. The figures for full sample are in parenthesis.

When interpreting these comparisons, it is important to note differences in geographical context and service system dynamics between the replication centres and the original RCT setting. Differences in local referral pathways, early childhood education and care market conditions, and availability of wraparound services are likely to influence both recruitment pace and retention patterns. Parkville Institute is monitoring how local implementation conditions shape these patterns.

4.2 Referral organisations

Table 4.3 presents the distribution of referral sources across the three centres. Maternal and child health nurses were the dominant referral source, accounting for 55% of all referrals and particularly prominent in the Victorian centres. Family services represented the second major pathway, more

[4] The inconsistency in consent rate for RCT reported between baseline report and this report is due to correction of a typing error.

prominent in Centres 2 and 3 than Centre 1. Child protection, community health, and other health services each contributed smaller proportions.

These patterns likely reflect differences in local service systems, professional networks, and pathways through which vulnerable families access services. The mix of referral sources may shape the profile of families reached and their subsequent program engagement.

Table 4.3: Referral organisations

Referral organisation (%)	Centre 1	Centre 2	Centre 3	Total
Family service	**11.3**	**35.0**	**29.2**	**22.0**
Community health	**20.0**	**0.0**	**0.0**	**9.5**
Child protection	**3.8**	**10.0**	**12.5**	**7.7**
Maternal and child health nurse	**62.5**	**37.5**	**58.3**	**55.4**
Child development or mental health clinic	**2.5**	**17.5**	**0.0**	**5.4**

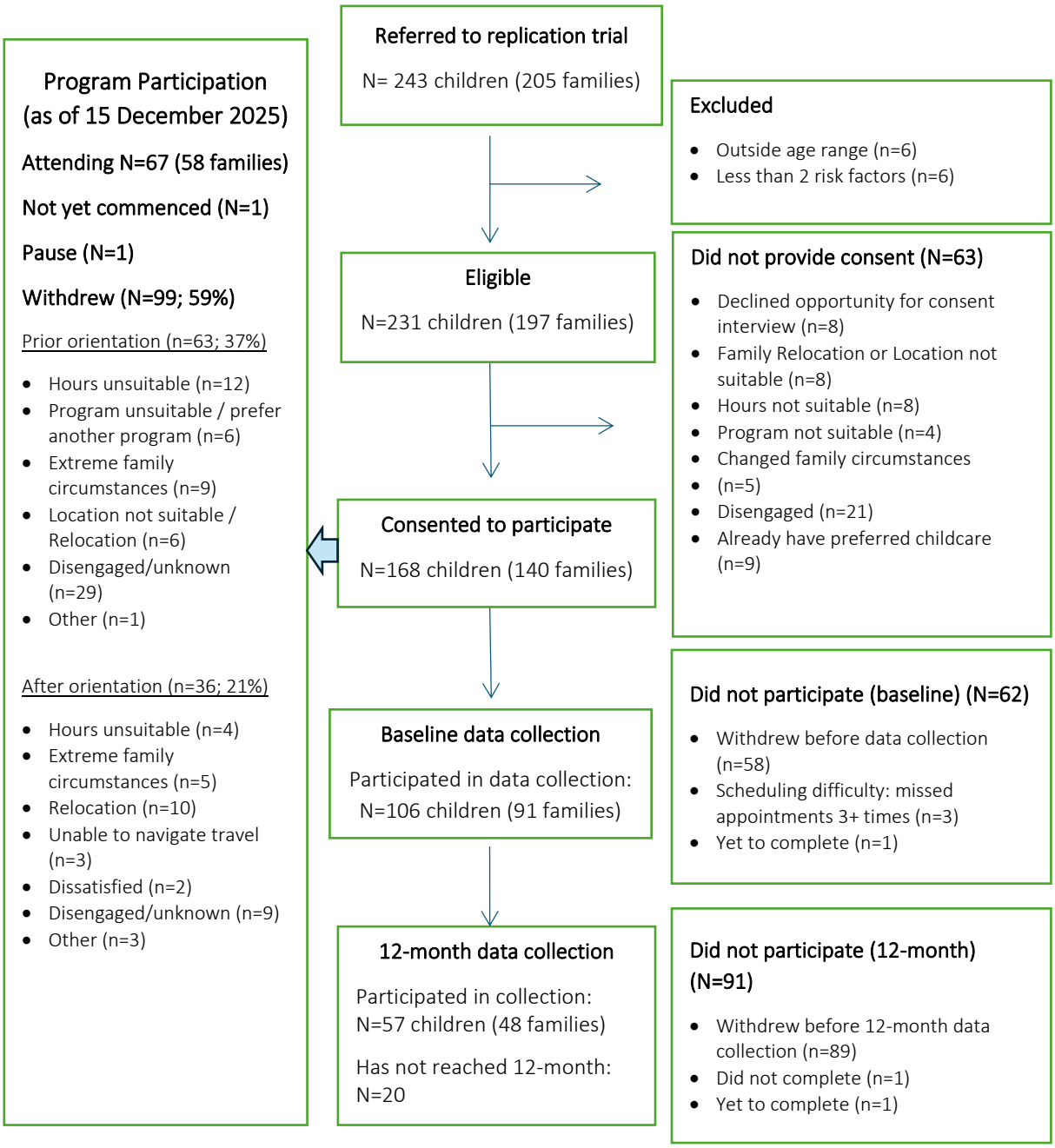
4.3 Overview of trial progress and engagement

This section summarizes participant flow from referral through to completion of baseline and 12-month data collection. Figure 4.1 provides an overview of the number of children retained at each stage and the primary reasons for non-consent and withdrawal.

As shown in Figure 4.1, a total of 243 children from 205 families had been referred to the program by June 2025. Twelve children were excluded because they did not meet eligibility criteria: six were outside the eligible age range and six had fewer than two risk factors leaving 231 eligible children from 197 families. Of the 231 eligible children, 168 children from 140 families consented to participate in the replication trial. Of those who did not provide consent, many families did not provide specific reasons (n=29), either families declined consent interviews or disengaged without reason (some were unable to be contacted) despite the high number of follow-up attempts. Common reasons cited are relocation or location barriers (n=8), or hours not being suitable (n=8). Smaller numbers were due to changed family circumstances (n=5) or the program not being suitable (n=4).

As of 15 December 2025, program participation status for the 168 children who consented was as follows: 67 children from 58 families were attending, one child had not yet commenced, one child's participation was on pause, and 99 children (59% of those who consented) had withdrawn from the program. More than half of them withdrew prior to orientation (n=63, 37% of all children consented to participate), with the frequently cited primary reasons being hours not suitable (n=12), extreme family circumstances (n=9), and relocation or location not suitable (n=6). Many children (n=29) were disengaged without explicitly mentioning the reason. A further 36 children (21%) withdrew after commencing orientation, with relocation being the most cited reason (n=10). Other common reasons include extreme family circumstances, hours not suitable and practical access barriers such as transport difficulties. These withdrawal patterns are consistent with the complex and dynamic life circumstances of families targeted by EYEP, including housing instability, competing crises, and multiple service system demands.

Figure 4.1: Replication trial recruitment and program participation flow chart



Note: Recruitment is in-progress at the time of data processing. The flowchart includes only those families that were referred prior to 30 June 2025 and the program participation report these children participation status as of 15 December 2025. A small number of children (<10) disengaged or withdrew more than 3 months after initial referral were referred again subsequently into the trial. These are counted as separate referrals as the recruitment outcomes were different for the two referrals. Withdrawal reasons that have only 1 observation are grouped into others. Unless participants explicitly indicate that they no longer wish to participate in data collection, they are eligible for follow-up data collection regardless of their participation in baseline data collection.

At the time of analysis, 57 children from 48 families had participated in 12-month follow up data collection. Among children who had reached the 12-month timepoint and remained enrolled in the program, data collection participation rates were high: only one child did not participated in 12-

month data collection and one child's assessment was pending at the time of reporting.[5] Many children had not yet provided 12-month data because they withdrew before reaching the 12-month timepoint (n=92) or had not yet reached the 12-month follow-up window (n=20). This resulted in a smaller 12-month outcome sample than the number originally consented.

The retention patterns observed reflect the challenges of conducting research with highly vulnerable families experiencing multiple stressors. These patterns underscore the importance of flexible scheduling, persistent follow-up, and trust-building approaches in both program implementation and research evaluation.

4.5 Family and parenting risk factors

The following provides updates to the information presented in baseline Replication Report 1 on the comparison of the prevalence of risk factors among participants in replication trial with those in the original RCT using the whole sample which include all children referred to the replication trial before June 2025. The updated comparison indicates that the replication trial continues to reach families with a broadly similar child profile to the EYEP RCT treatment group. Children are similar between the two trials in age at referral (average 1.42 years) and gender composition (51.8% boys in replication versus 56.9% in RCT; this difference is not statistically significant).

Consistent with the baseline report, families in the replication trial present with high overall cumulative burden of family stressors at referral. The average number of recorded risk factors is significantly higher in the replication trial than the RCT (5.9 versus 4.7).[6] At the level of individual risk factors, the replication trial participants show higher prevalence across several family adversity indicators. Social isolation is significantly more common in the replication trial (38.7% versus 25.5%). The proportion exposed to family violence is also higher (67.3% versus 52.8%) although this difference is only approaching statistical significance. More children in the replication trial are also living with a disability or complex medical issues (40.5% versus 26.4%).

In contrast, the RCT treatment group shows higher prevalence on parenting capacity risk indicators. "Harsh, inconsistent discipline, neglect or abuse" is substantially more common in the RCT than the replication trial (33.3% versus 8.9%). The RCT families also had a higher proportion of teenage parents (16.7% versus 7.1%).

These findings reinforce the pattern noted in the replication trial baseline report: while the replication trial is reaching children and families with high levels of cumulative adversity—including complexity related to child disability and medical issues, and social isolation—the RCT treatment group is more heavily characterised by parenting risk markers that are more proximal to statutory child protection concerns.

---

[5] Participation is defined as they provided data in any of data collection. Some participants may not provide all data items.

[6] Because participants' referral pathways differ substantially between the RCT and the replication trial, we analysed the total number of risk factors among replication trial participants by referral organisation. The average number of risk factors for children referred by child protection or family support practitioners was slightly higher than for those referred by health professionals. Given that the majority of RCT participants were referred by family support or child protection practitioners, it is unlikely that the higher number of risk factors in the replication trial is driven by differences in referrers.

These differences are likely to reflect variation in referral pathways and service system contexts across the trials. The original RCT received a large share of referrals via Child FIRST and statutory child protection services, whereas referrals in the replication trial include a higher proportion from health professionals (particularly maternal and child health nurses) and comparatively few from statutory child protection. This referral composition may shape the risk profile of families who enter the program. Parkville Institute will continue to examine referral processes and barriers to ensure that children most aligned with EYEP's intended target group are able to access the program.

Participants recruited after September 2024 are very similar to those recruited prior (i.e. those included in the baseline report). Comparisons of the two groups are reported in Appendix Table C1.

Figure 4.2 Family risk factors of consented children: Replication trial vs. RCT treatment group

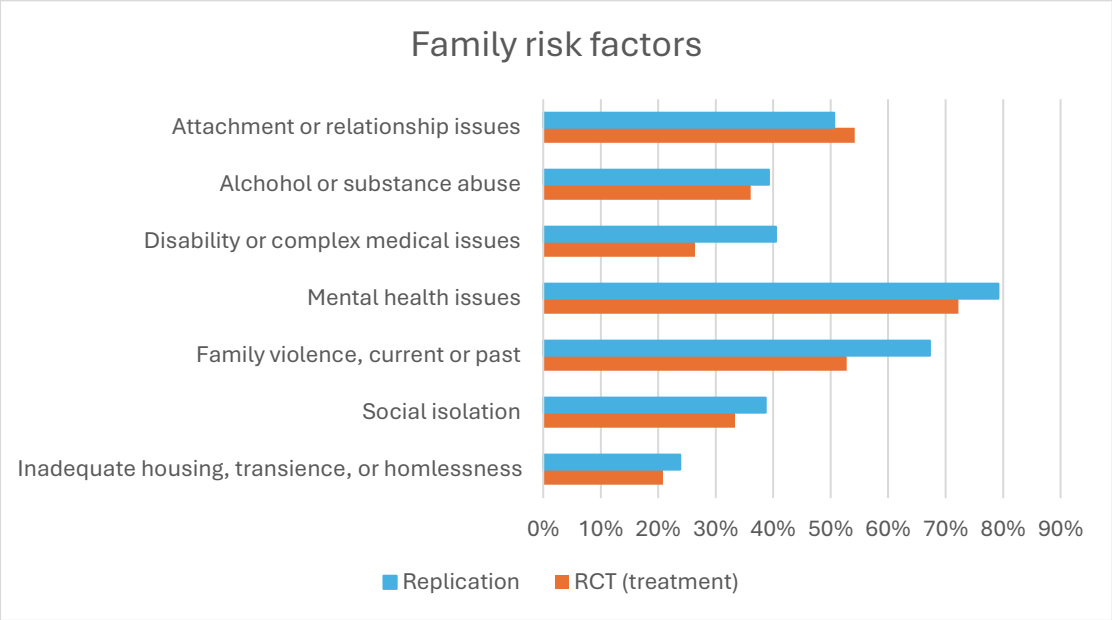
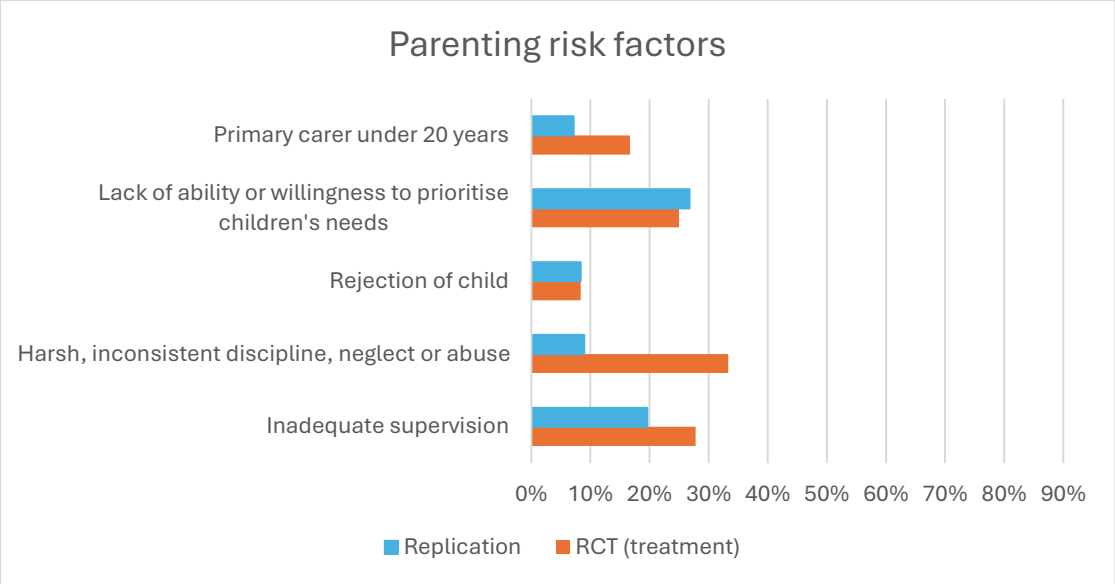


Figure 4.3 Parenting risk factors of consented children: Replication trial vs. RCT treatment



.

4.6 Child health and development at trial entry

This section presents updated baseline characteristics for all children referred before the end of June 2025. As documented in the baseline report (Tseng et al., 2025), children who entered the replication trial by September 2024 showed substantial developmental vulnerability across all assessed domains, and were more vulnerable than children in the original RCT. The updated results with all the sample continue to show this pattern.

Baseline child development was assessed using the Bayley Scales of Infant and Toddler Development, Fourth Edition. The cognitive and language measures are direct child assessments conducted by qualified clinicians, while the social-emotional measure is parent-reported information from interviews conducted by qualified technicians using Bayley Social Emotional Scale questionnaires. As all scores are standardised, the general population is expected to have a mean score of 100 and a standard deviation of 15. Following Johnson et al. (2014), Standardised scores are classified into the following categories[7]:

- Superior range: Score 130 and above
- Above Average: Score 115-129
- Average Range: Score 85-114
- Mild Delay: Score 70-84 (1-2 standard deviations below mean)
- Moderate to Severe Delay: Score below 70 (lower than 2 standard deviations below mean)

Figures 4.4–4.6 show the distributions of baseline scores for replication trial participants, alongside the RCT and expected distribution in general population benchmarks. The corresponding mean scores and statistical tests of the differences between the participants of the two trials are presented in Table 4.3.

The average baseline cognitive and language scores of children in the replication trial are about two-thirds of a standard deviation lower than those of children in the RCT (80.1 versus 90.0 for cognitive, and 77.4 versus 86.3 for language), and more than one standard deviation lower than those of the general population.

[7] Johnson et al. (2014) provided a classification for the Bayley-III, which was adopted in the EYEP RCT report No. 1. However, the replication trial had to update to the Bayley-IV due to the phasing out of the Bayley-III. To our knowledge there are no research articles that compare the scores of Bayley III and IV. There is no recommendation on classification of developmental delay using Bayley IV. Thus, in baseline and this report, we follow the classification from the RCT report 1 for consistency and comparison purposes.

Figure 4.4 Baseline Bayley scales of infant and toddler development: Cognitive

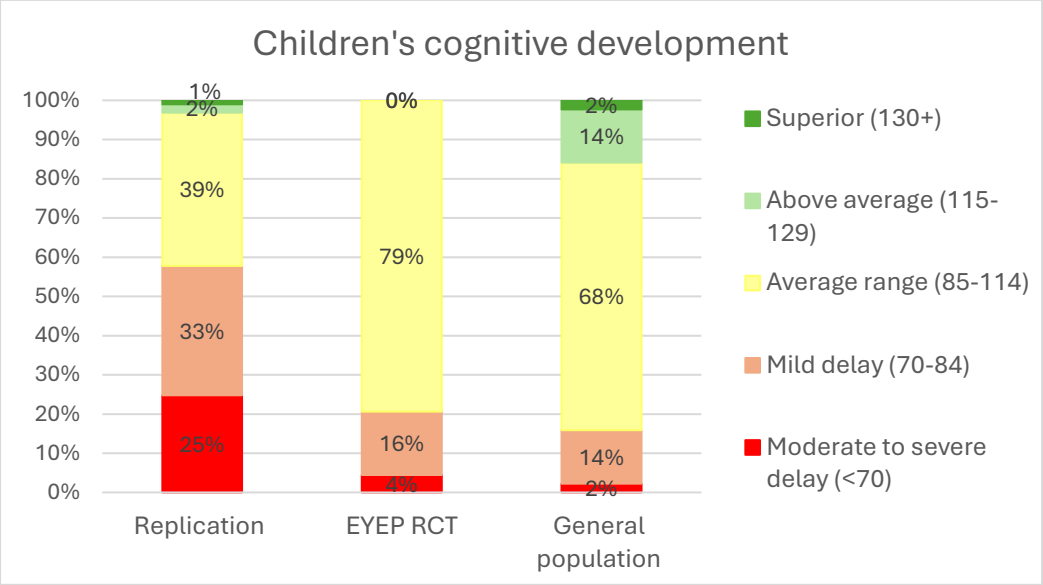
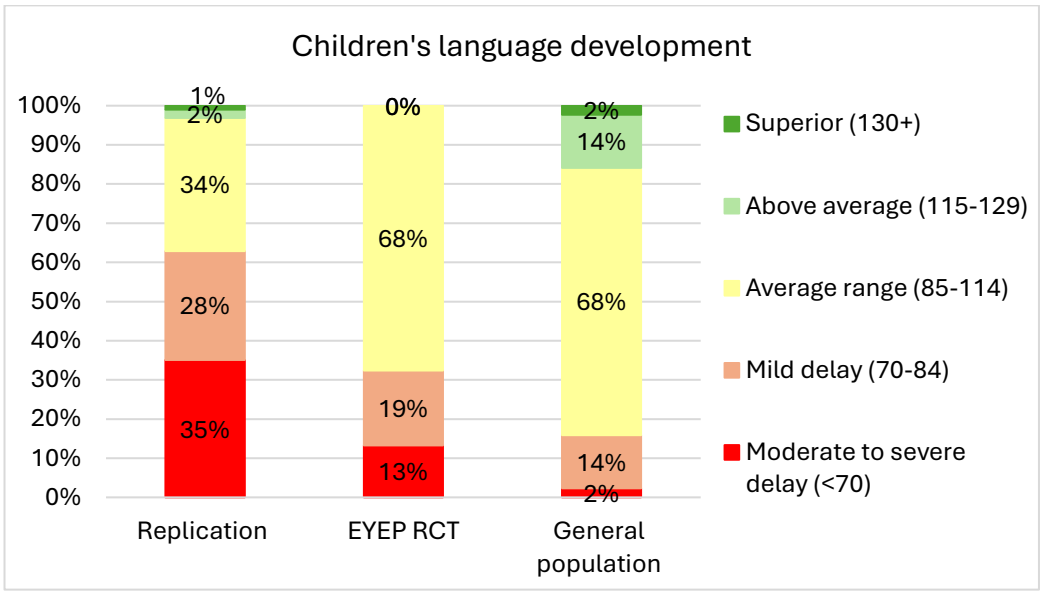
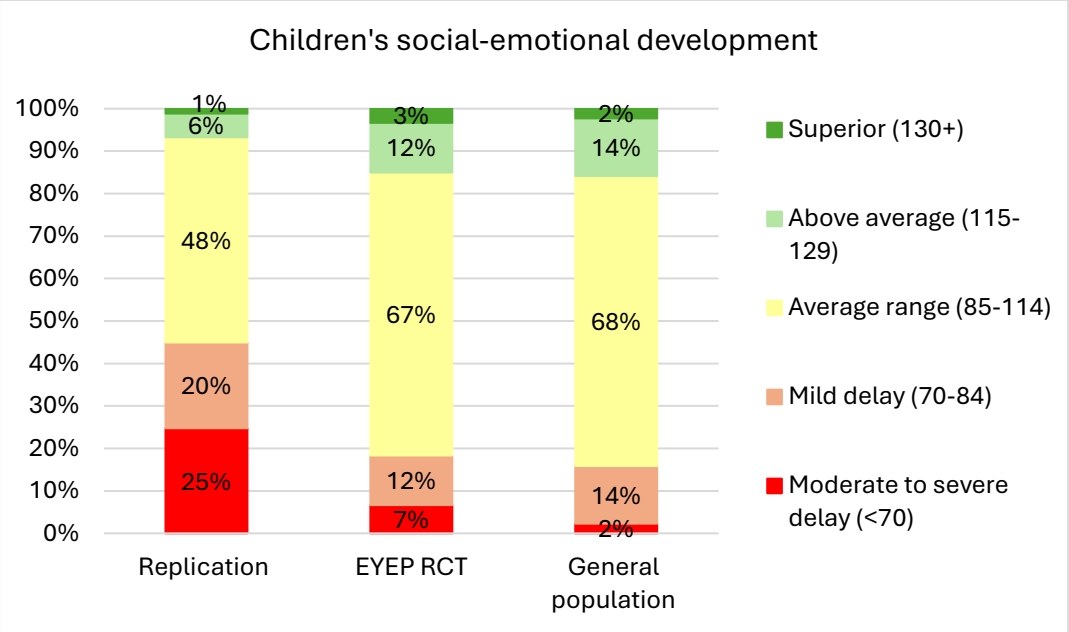


Figure 4.5 Baseline Bayley scales of infant and toddler development: Language



Developmental delay is also much more common in the replication trial sample. Over half of the children met criteria for cognitive delay (57.7% versus 20.6% in the RCT), and nearly two-thirds met criteria for language delay (62.9% versus 32.4%). Considering cognitive and language together, 73.2% of children in the replication trial scored below 85 in at least one of these domains, compared with 33.8% in the RCT treatment group.

Figure 4.6 Baseline Bayley scales of infant and toddler development: Social-Emotional



Social-emotional delay at baseline was likewise more prevalent among replication trial children than RCT children (44.9% versus 18.3%), alongside a 12.3-point gap in average scores. While children in the RCT did not show significant social-emotional developmental differences compared with the general population, children in the replication trial demonstrated greater social emotional developmental differences, with an average score of 16.9 points (more than one standard deviation) lower than the general population.

As shown in table 4.4, parent-reported child health indicators were broadly similar across the two cohorts for hospitalisation history and birth weight. However, parents in the replication trial were more likely to report that their child had a developmental delay or neurodevelopmental disorder (22.6% vs 9.7%), and this difference was statistically significant. The systematic pattern of lower clinician-assessed developmental scores across domains suggests these differences reflect a more developmentally vulnerable population being reached by the replication trial, rather than reporting anomalies. These baseline differences provide important context for interpreting 12-month impacts and underscore the need to account for differences in initial developmental status and risk profiles in comparative analyses.

Table 4.4: Child health and development at entry to the study

	Replication (T)	RCT tm (C)	Differences	
			Dif. (T-C)	p-value
Child BAYLEY scale baseline assessment standard scores, assed by qualified clinicians (population average=100, std=15 for all standard scores)				
Cognitive average score	80.1	90.0	-9.9	<0.001
Language average score	77.4	86.3	-8.9	0.004
Social-Emotional average score	83.5	95.8	-12.3	<0.001
% with any delays of cognitive or language	73.2%	33.8%	39%	<0.001
Parent reported child health and development				
Average child birth weight (kg)	2.7	3.0	-0.3	0.151
Hospitalised once since birth (excluding birth)	23.3%	21.9%	1.5%	0.837
Hospitalised multiple times since birth (excluding birth)	7.8%	9.4%	-1.6%	0.769

Developmental delay or Neurodevelopmental disorder	22.6%	9.7%	12.9%	0.030
Child ever had formal childcare prior to trial	34.4%	22.7%	11.7%	0.122
Number of children	97	68		

5. Outcomes after 12 months

This section examines changes in outcomes from baseline to 12 months for children in the replication trial and compares these changes with children with similar family characteristics in EYEP RCT treatment group. The primary question is whether children in the replication trial are achieving developmental and learning progress comparable to children who attended EYEP in the RCT.

5.1 Overview of outcome variables

Six main outcomes were selected to align with the EYEP RCT research reports (Tseng et al., 2018), enabling direct comparison of developmental trajectories between the replication trial and RCT treatment group. The outcomes include children's cognitive and social-emotional development, as well as primary caregiver psychological wellbeing and parenting daily hassles. Table 5.1 lists all outcomes, with detailed descriptions provided below.

Table 5.1 Summary of outcome measures

	Outcomes	Measures
1	Child development - IQ	Bayley Scales of infant and toddler development (BSID-IV); Wechsler preschool and primary scale of intelligence (WPPS-IV)
2	Child development – Language skills	Bayley Scales of infant and toddler development III (BSID-IV); Weschler preschool and primary scale of intelligence (WPPSI-IV) – Verbal IQ score
3	Child development – Protective factors related to resilience (initiative, self-regulation, attachment/relationships, behavioural concerns)	Devereux Early Childhood Assessment Program (DECA)
4	Child emotional and social development	Brief infant toddler social emotional assessment (BITSEA); Child behaviour checklist (CBCL)
5	Parenting stress	The parenting daily hassles scale
6	Parent psychological distress	K6

Cognitive and language development are evaluated using standardised tests: the Bayley Scales of Infant and Toddler Development, Fourth Edition (Bayley-IV, 2019); and the Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition (WPPSI-IV) (Wechsler, 2012). These are the most widely used measures of development of infants and toddlers in clinical and research settings.

The replication trial uses Bayley-IV for children aged up to 42 months, and WPPSI-IV for children aged 43 months and above. Age-adjusted composite scores can be calculated for IQ and language domains for both measures. Both are scaled with mean of 100, and standard deviation of 15. Since the Bayley Scales and WPPSI are scaled equivalently against population norms, scores from these measures are integrated in our analysis.[8] Whether combining different measures affects findings is tested as part of empirical analysis.

[8] Although Bayley and WPPSI standard score has the same mean and standard deviation, they have slightly different range where the lower limit of WPPSI standard score is lower than the limit of Bayley standard scores.

Within-child protective factors related to resilience 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 for infants aged 1 month to less than 18 months, DECA-T for toddlers 18 to less than 36 months, and DECA-P2 for children aged 3 to 5 years. Responses 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-normalized T scores and percentile rankings. The T score has mean of 50 and SD of 10, ranging from 28 to 72. A score of 40 or below signifies an area of need.

Child social and emotional development is 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. BITSEA is used for children from 1 year up to three years, and CBCL for children older than three years. The BITSEA identifies children aged less than 36 months who may have social-emotional and behavioural problems and/or delays or deficits in social-emotional competence. The problem score 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 for children from 18 months to 5 years. Total score ranges from 0 to 200. A percentile ranking based on age-based population norms can be assigned to each score. The BITSEA and CBCL instruments are integrated to obtain a consistent measure using the proportion of children classified as having development problems in the clinical range. The clinical range is defined as scores below the population norm age-based 10th percentile cut-off. We also construct another variable defined as a score below the population norm age-based 25th percentile cut-off.

Parent psychological distress 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. The scale has six questions about feelings over the last four weeks. A K6 score is derived from summing responses, ranging from 6 to 30, with individuals scoring 6 to 13 classified as 'low' psychological distress, 14 to 18 as 'medium', and 19 to 30 as 'severe' psychological distress.

The Parenting Daily Hassles scale assesses frequency and intensity/impact of 20 experiences that can be a 'hassle' to parents. Frequency score ranges from 0 to 80 and 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.

5.2 Data and attrition analyses

We focus on children’s development trajectories rather than outcome levels, for the 12-month outcome analyses presented in this report. The sample for outcome analyses includes all participants who completed both baseline (t1) and 12-month (t2) developmental assessments, regardless of date of referral in order to maximise statistical power. Most of the 12 months assessments are undertaken 11-14 months after children commenced orientation. All children assessed at 12-month follow up are still attending the centre at the time of assessment which is the same as the RCT sample included in the comparison. The RCT treatment sample includes all children who participated in EYEP for at least 60 days prior to the assessments which is the same as the main sample used in the RCT 12-month report (report 2, Tseng et al 2018).

Table 5.2 presents the sample size and baseline information for participants with outcome change measures. There is no comparable social-emotional outcome measure for children under 12 months of age at baseline, therefore the outcome measure analysed is the score at 12 months rather than progress between baseline and subsequent outcomes. Although there are more participants, the sample size is smaller for the replication trial analytic sample than in the RCT due to lower retention rates in addition to many children not yet reaching the 12-month data collection reference window.

The children in the outcome analytic sample have lower average baseline developmental scores than the RCT which is consistent with the profile for the whole sample recruited before June 2025. Nonetheless, the differences in baseline language scores between the RCT and replication trials are larger for the 12-month outcome analytical sample than the whole sample recruited.

Interestingly, there is no difference in K6 scores between the replication trial and RCT in the baseline report (Tseng et al 2025) but there is a significance difference between the two trials in 12-month outcome analytical sample where the average K6 score of replication trail sample is lower than the score of the RCT sample. There are subtle differences in retention pattens between two trials where parents with lower psychological distress in the replication trial are more likely to respond to the surveys, but the RCT goes the other direction. A further analysis of sample attrition suggests that parents with severe psychological distress at trial entry were less likely to respond to interviews for both child development assessment and parent questionnaire interviews at 12-month follow up even though children were still participating in the program[9]. It is unclear whether there is a bias in any directions as the reliability of parents' child development assessment for those with mental health issues is an under-researched area, especially when the focus is on changes in outcomes over time. Unfortunately, the current sample size is insufficient to undertake further analyses in this report. We will re-investigate this in the final report.

Table 5.2 overview of analytic sample

	Replication (REP)		RCT		Difference	
	N	Baseline measure	N	Baseline	REP-RCT	P-value
IQ	43	82.9	50	92.6	-9.7	0.139
Language	43	76.0	50	89.7	-13.7	0.025
Total protective factor– resilience	35	40.1	48	46.5	-6.5	0.000
Social-emotional development - problem	42	NA	47	NA	NA	NA
Parenting daily hassles – Frequency	33	42.1	40	45.1	-3.0	0.543
Parenting daily hassles – Intensity	33	37.9	39	44.7	-6.8	0.377
Psychological distress	39	13.5	47	15.1	-1.6	0.019
Note: As we focus on outcome changes between baseline and 12-month follow up (except social-emotional development), analytic sample includes participants with both baseline and 12-month outcomes. Comparable social-emotional problem measures are not available for children under 12 months of age; therefore, results are not reported here.						

To assess the generalisability of the outcome findings, we also examined attrition patterns in relation to risk factors and baseline scores on outcome measures. We report attrition patterns for IQ (and language), where outcomes are assessed by qualified professionals, and for the total protective factor score related to resilience (assessed using DECA), which is derived from parent interview responses.

[9] Statistical analyses results are available upon request.

IQ and language outcomes at 12 months can be assessed in the classroom and are therefore less dependent on parent presence. Attrition in the IQ sample is largely due to children leaving the program; only a small number (n <5) were lost because they were unable to complete the WPPSI assessment due to their existing developmental or neurological conditions.

Tables 5.3 and 5.4 present differential attrition for those items that were ever found to differ significantly (or nearly significant) between participants with and without outcome measures. Full results are available on request. As the replication trial is still on going, to assess attrition fairly, we only include participants who had/would have completed 12-month data collection window.

Beyond the marked difference in overall retention rates, the patterns of who remains in the program differ substantially between the RCT and the replication trials. In the replication trial, children in the IQ analytic sample (i.e. with both baseline and 12-month IQ) and those excluded (missing the 12-month IQ) do not differ significantly in family risk factors or baseline scores. The mean baseline language score is 7.4 points lower in the analytic sample than in those excluded, but this difference is not statistically significant. By contrast, the RCT shows more pronounced selection: families experiencing mental-health difficulties are more likely to remain in the program, whether mental health is measured using referrer-identified risk factors or the parent-responded K6. The RCT analytic sample also has substantially higher mean IQ and language scores at baseline.

Table 5.3 comparison of data attrition patterns— IQ change measure

	IQ outcome analytic sample (I)	Not in IQ outcome analytic sample (E)		Differences (I-E)	
	mean	mean		% pt	p-value
	Replication trial consented participants				
Risk factors: %					
Alcohol or Substance use	26.5	36.2		-9.8	0.377
Mental health issues	85.3	75.4		9.9	0.209
Baseline score:					
IQ	79.0	77.0		1.9	0.666
Language	71.2	78.6		-7.4	0.170
K-6 Psychological distress scale	13.9	13.9		0.0	0.981
Sample size (maximum)	34	69			
	RCT consented participants				
Risk factors:					
Alcohol or Substance use	34.0	40.9		-6.9	0.665
Mental health issues	82.0	50.0		32.0	0.037
Baseline score:					
IQ	92.6	82.8		9.8	0.029
Language	89.7	77.1		12.5	0.013
K-6 Psychological distress scale	15.4	12.0		3.4	0.038
Sample size (maximum)	50	22			

Note: P-value <0.05 denotes statistically significant at 5% level and P-value <0.1 denotes significant at 10% level. Sample of replication trial include those who has/would have completed 12 months outcome data collection windows (i.e. consent to participate in the trial prior to 30 June 2024). The IQ outcome analytic sample comprises participants with both baseline and 12-month IQ scores. Sample sizes in the attrition analyses vary according to which baseline measures are analysed, because not all participants completed every baseline assessment.

The attrition pattern for the DECA outcome measure is similar. However, families with drug or alcohol problems were less likely to complete interviews for child development assessments. These differences are attributable both to withdrawal from program participation and to difficulties scheduling parent interviews for families who remained enrolled.

Table 5.4 comparison of data attrition patterns—total protective factor (DECA) score change

	DECA score change analytic sample (I)	Not in the DECA score change sample (E)		Differences (I-E)	
	mean	mean		% pt	p-value
	Replication trial consented participants				
Risk factors: %					
Alcohol or Substance use	21.2	38.6		-17.4	0.101
Mental health issues	81.8	77.1		4.7	0.577
Baseline outcomes: (score)					
IQ	80.6	75.5		5.2	0.267
Language	74.3	75.3		-1.0	0.861
K-6 Psychological distress scale	13.6	14.2		-0.5	0.711
Sample size (maximum)	33	70			
	RCT consented participants				
Risk factors:					
Alcohol or Substance use	33.3	41.7		-8.3	0.599
Mental health issues	81.3	54.2		27.1	0.070
Baseline outcomes: (score)					
IQ	92.1	85.0		7.1	0.076
Language	89.5	78.7		10.9	0.017
K-6 Psychological distress scale	15.4	12.2		3.3	0.039
Sample size (maximum)	48	24			

Note: P-value <0.05 denotes statistically significant at 5% level and P-value <0.1 denotes significant at 10% level. Sample of replication trial include those who has/would have completed 12 months outcome data collection windows (i.e. consented to participate in the trial prior to 30 June 2024). The total protective factor (DECA) score analytic sample comprises participants with both baseline and 12-month total protective factor scores. Sample sizes in the attrition analyses vary according to which baseline measures are analysed, because not all participants completed every baseline assessment.

The differential attrition patterns further widen the gap in age-adjusted IQ and language scores between the two trials. On average, the replication trial is providing intervention to children whose baseline IQ is 11.5 points lower and whose language scores are 15 points lower than those of children in the RCT, with a higher proportion reported as having confirmed developmental delay or neurological conditions. Some children may yet have undiagnosed conditions, which are contributing to these baseline disparities.

The extremely low baseline language scores have significant implications: children may need to make gains in language before they can fully benefit from other components of the intervention. This issue can be linked to findings that will be discussed in later sections.

5.3 Estimation methods

The modelling strategy is guided by the results of balancing tests, that is, comparisons of participant characteristics between the two trials. Because the final sample for each outcome differs slightly, we conduct separate balancing tests to examine whether the characteristics of the samples differ significantly between the trials for each outcome. Any variables that are found to differ significantly are included as controls in the corresponding statistical models.

For the sample of IQ and language outcomes, the key differences between the two trials are consistent with the findings reported in Section 4.5: social isolation, harsh or inconsistent parenting, neglect or abuse, and the total number of risk factors. These differences are therefore adjusted for in the analysis. Children's baseline IQ and language scores also differ substantially between the trials. We address these differences by modelling progress (change in scores) rather than score levels as the outcome variable. Regarding whether children should additionally be compared at the same baseline IQ and language levels, there are important trade-offs to consider. Given the marked baseline differences, further conditioning on baseline scores would either require implausible extrapolation or generate very large weights (for example, a single child serving as a comparison for many treated children), depending on the econometric method used. For this reason, we do not include baseline outcome scores in the main specification. For other basic control variables, we largely follow the EYEP RCT report by Tseng et al. (2021), where covariates were selected through a comprehensive program of testing.

Our final main specification includes the following sets of variables:

- Child's gender, age, time between baseline and 12-month assessment
- Primary carer's age, education, psychological distress, whether they speak a language other than English at home, and presence of both parents
- Family risk factors and parenting risk factors:
    - Alcohol or substance use
    - Disability/complex medical issues
    - Family violence, current or past
    - Social isolation (family, community, and cultural)
    - Harsh, inconsistent discipline, neglect or abuse

As the K6 mental illness screening tool scores are included, we do not include referrer reported mental health risk factor. The total number of risk factors is also omitted to avoid multicollinearity as many risk factors have already been controlled for. We also test on another simplified specification, excluding 3 variables— whether parents speak language other than English at home and presence of both parents, and alcohol and substance use risk factor.

As for choice of econometric model, there is no perfect one due to the sample size and the large differences between the replication trial and RCT sample. We decided to test the sensitivity on three models—regression adjusted matching, regression adjusted inverse probability reweighting and regression methods.

Outcome change scores are calculated as the difference between 12-month (t2) and baseline (t1) values (t2-t1). For social–emotional outcomes, comparable progress measures are not available for the full age range (given different instruments for children under and over 12 months); therefore, we report the 12-month proportion in clinical concern ranges rather than baseline-to-follow-up change.

5.4 Summary of attendance in the program

Understanding attendance is important for the interpretation of program outcomes. This section summarises program attendance for children in the replication trial compared with those in the RCT. We focus on regular attendance, which begins once the leadership team and the primary educator agree that the child is able to use the primary educator as a secure base. After regular attendance commences, the child is expected to attend five hours a day, five days a week.

In the RCT, orientation completion was not emphasised, and the administrative attendance data did not distinguish orientation from regular attendance. Consequently, researchers used an ad hoc definition of the start of regular attendance: a child attending EYEP for a full day for at least three consecutive days. Because of these different definitions and recording practices, the attendance data between the trials are not strictly comparable, and comparisons should be treated with caution.

The first-year attendance observation window spans from the start of regular attendance to the 12-month follow-up IQ assessment. However, attendance records for the replication trial are available only up to September 2025. Therefore, total days attended are reported only for children with a complete observation window[10].

Table 5.5 summarises program attendance. Attendance statistics for all children and for IQ outcome analytic sample are presented. The comparison of attendance for IQ outcome analytic samples between the two trial is important for the interpretation of outcomes which will be presented in the next section.

Table 5.5 Overview of program attendance prior to 12-months follow up data collection

	All children commenced regular attendance		Children in IQ change outcome analytic sample	
	Replication	RCT	Replication	RCT
Average duration attended	353.4	290.3	409.2	315.0
Average total days attended	185.2	145.2	234.7	167.9
Average attendance rate	79.8	71.5	83.3	80.2
Note: 1. IQ analytic sample are those with both baseline and 12-month follow-up IQ scores and attended the program for more than 60 days. 2. Total days attended include those whose observation window completed prior to 30 September 2025.				

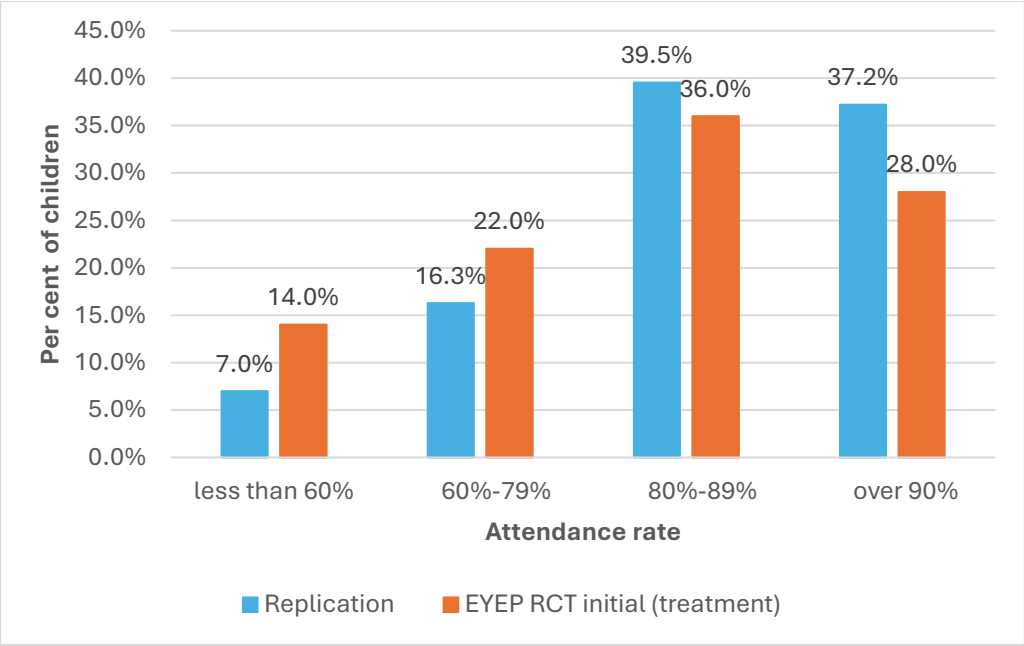
All children in the RCT and replication trial analytic samples attended the EYEP program for more than 60 days. As expected, children in the IQ analytic samples attended for slightly longer than the full samples, because most missing 12-month IQ assessments resulted from withdrawal from the program. On average, children in the replication trial attended 235 days, compared with 168 days in the RCT; this difference reflects the slightly later timing of the IQ assessment in the replication trial and marginally higher attendance rates.

Attendance rates (days attended divided by days expected to attend) are generally high in both trials. Because attendance rates are less affected by differences in data collection methods, we further compare the distributions of attendance rates across the trials.

[10] Children who did not complete 12 months IQ assessments, the end of attendance observation window is defined as the orientation start date plus 365+90.

As shown in Figure 5.1, over 75% of children in the replication trial have an attendance rate above 80%, compared with about 64% in the RCT. The EYEP program in the replication trial is funded through the Additional Childcare Subsidy (ACCS-Child Wellbeing), which permits up to 42 allowable absence days per child in each financial year. This funding rule may have contributed to the higher attendance rates observed in the replication trial. However, it may also lead families facing extreme circumstances to stop or suspend participation once allowable absences are exhausted; this occurred for a small number of children in the replication trial. This administrative feature therefore warrants further consideration in the future implementation of the program.

Figure 5.1 Comparison of attendance rate distributions for IQ analytic sample



5.5 Results: changes in outcomes between baseline and 12-month follow up

This section examines how children and families in the replication trial progressed over their first 12 months in the program and compares this progress with children in the RCT treatment group at the same timepoint.

Table 5.5 provides an overview of outcomes across all assessed domains. For each outcome, the table shows average changes from baseline to 12 months for both the replication trial and RCT treatment group and test whether the changes are (statistically) significantly different from zero. Two stars indicates significant at 5% level (or lower) and one star indicates significant at 10% level. It also shows the difference between the samples in the two trials, that is difference-in-difference (Replication trial – RCT) in two ways: first, without any adjustment in characteristics; second, after accounting for the differences in characteristics and risk factors using regression adjusted matching method. Statistical significances are denoted by stars as described above. As there is no priori hypothesis on which trial does better, two tail tests are used. Results of sensitivity analyses that vary covariates or econometric estimation methods as described above in section 5.3 are presented in Table 5.6 to examine the robustness of findings.

It is important to note that a child's improvement over time does not necessarily only reflect program impact. Children's development can progress naturally even without intervention and can also regress for those who have experienced toxic stress or have neurological conditions (Perry, 2002; Shonkoff, 2012; Alameddine 2025). The average outcome changes reported in this study also cannot be compared with the impact estimates reported in EYEP research reports where the impact estimates represent differences between children who received the EYEP intervention and those who received usual care.

Table 5.6 Comparison of changes in outcomes between baseline and 12-month follow-up

	Replication (REP)	RCT treatment	Difference-in-difference (REP-RCT)	
			Comparison without control variables	Matching-Regression adjusted (model 1)
Outcomes of children				
IQ (t2-t1)	4.6	6.7**	-2.1	2.9
Language (t2-t1)	7.2**	3.0*	4.2	0.5
Total protective factor related to resilience (t2-t1)	11.3**	0.5	10.8**	14.4**
Social-emotional development (% below bottom 10th percentile population norm) (t2)	14%	21%	-7ppt	-5ppt
Social-emotional development (% below bottom 25th percentile population norm) (t2)	31%	53%	-22ppt**	-51ppt**
Outcomes of primary carers:				
Parenting daily hassles – Frequency (t2-t1)	0.9	0.6	0.3	-0.6
Parenting daily hassles – Intensity	2.3	3.2	-1.0	-6.6*
Psychological distress (K6)	2.4**	0.2	2.1	2.7
Note: (1) All the replication trial sample commenced regular attendance in the program. (2) Social emotional development does not have comparable measures for baseline, therefore the outcome measures are 12 month outcomes rather than outcome progress between baseline and 12 month (t2-t1). (3) * denotes significant at 10% level and ** denotes significant at 5%				

The following summarise the findings by domains:

Child outcome: Cognitive Development (IQ)

As shown in Table 5.6, the average change in IQ from program entry to 12 months for children in the replication trial was an improvement of 4.6 points. Although this magnitude is not trivial (nearly one-third of a standard deviation), it is not statistically significant because of the large variation among children. Children in the RCT treatment group improved significantly by 6.7 points over the same period—2.1 points higher than the replication trial average—but that difference is not statistically significant.

After accounting for the substantial differences in baseline characteristics between the two groups using regression adjusted matching methods, there was no significant difference in IQ change between the replication trial and RCT treatment group. Point estimates actually favoured the replication trial slightly, though not statistically significantly. This is finding is robust across different estimation methods and covariates.

Several considerations are important when interpreting the IQ and language results. First, three children in the replication trial could not be assessed at 12 months using standardised tests (WPPSI) due to developmental delays that prevented engagement with standardised assessment procedures. Second, children presenting with severe developmental delay at baseline may include those with emerging or established neurodevelopmental conditions. For these children, we would not expect the same magnitude of standardised score change over 12 months as for children starting closer to average who did not have neurodevelopmental conditions.[i] Third, the replication trial includes substantially more children with severe delay (IQ <70) at baseline than the RCT, which creates challenges in creating statistically comparable matched groups.

Despite these considerations, the overall pattern indicates that IQ change for children in the replication trial is comparable to that achieved in the RCT treatment group, despite entering the trial with substantially greater cognitive developmental vulnerability.

Table 5.6 Sensitivity analyses: Comparison of changes in outcomes

	Difference-in-difference (REP-RCT)			
	Matching-Regression adjusted (model 1)	Matching-Regression adjusted (model 2)	IPW- Regression adjusted (model 1)	Regression (model 1)
Outcomes of children				
IQ (t2-t1)	2.9	5.7	4.9	-3.3
Language (t2-t1)	0.5	2.2	-1.0	2.4
Total protective factor related to resilience (t2-t1)	14.4**	10.9**	15.1**	10.7**
Social-emotional development (% below bottom 10th percentile population norm) (t2)	-5ppt	-1ppt	-1ppt	-3ppt
Social-emotional development (% below bottom 25th percentile population norm) (t2)	-51ppt**	-51ppt**	-51ppt**	-34ppt**
Outcomes of primary carers:				
Parenting daily hassles – Frequency (t2-t1)	-0.6	0.8	-1.0	0.9
Parenting daily hassles – Intensity	-6.6*	-5.3	-6.4	-5.1
Psychological distress (K6)	2.7	2.5	3.0	1.8
Note: (1) Model 1 and model 2 differ by sets of control variables included. Model1 includes family characteristics and risk factors in both weighting model and regression adjustment. Model 2 used a reduced set of matching variables (removed three variables that are not statistically significant in balancing test from model 1 but regression adjustment included the same variables as model 1). (2) All replication trial sample commenced regular attendance. (3) Social emotional development does not have comparable measures for baseline; therefore the outcome measures are 12-month outcomes rather than outcome progress between baseline and 12 month (t2-t1). Statistical tests of t2-t1 are not performed for social emotional measure. (4) * denotes significant at 10% level and ** denotes significant at 5%				

Child outcome: Language Development

Children in the replication trial demonstrated substantial language gains from program entry to 12 months, improving by an average of 7.2 points. This improvement was statistically significant. Average language change in the RCT treatment group showed smaller and non-statistically significant improvement of 3.0 points over the same period.

However, when we account for baseline differences between groups, the estimated difference between the two trials is not statistically significant. The estimates varied depending on analytical approach—ranging from the replication showing slightly better to somewhat worse outcomes—and none reached statistical significance.

This sensitivity to statistical methods likely reflects the marked baseline differences between groups, particularly the substantially higher proportion of children in the replication trial with moderate to severe language delays. The results tell us two important things: first, children in the replication trial made statistically significant language progress over 12 months; second, after accounting for their lower starting points, there is no clear evidence that their language change differs from what was observed in the RCT treatment group.

Child outcome: Total Protective Factor related to resilience (DECA)

Children in the replication trial demonstrated substantial improvement in protective factors related to resilience over 12 months, with scores increasing by an average of 11.6 points. In contrast, the RCT treatment group showed minimal change (+0.5 points). This difference of 11.1 points was highly statistically significant.

After accounting for characteristics and risk factors, this differences between the two trials remained highly significant and very large (14.4 points, nearly 1.5 of standard deviations). The estimates are all highly significant regardless of which econometric method and model specification was used, with the adjusted difference ranged from 10.7 to 15.1.

Notably, the analytic samples for IQ outcomes and for other parent-reported outcomes differ slightly. Attrition patterns indicate that families with drug or alcohol risk factors were less likely to take part in interviews for child-development assessments. In addition, primary carers with severe psychological distress at baseline were less likely to complete the 12-month protective-factors assessment, even when their children continued to attend the program. Consequently, observed gains in protective factors should be interpreted as reflecting outcomes for families who remained engaged with the assessment process, and may not generalise to the full baseline sample.

Child outcome: Social-Emotional Development

Social-emotional outcomes were measured using standardised parent-report questionnaires that identify children whose development places them at clinical risk—that is, with scores below what would be expected for their age. We examined two thresholds: the bottom 10th percentile of the general-population norm (clinical range) and the bottom 25th percentile of general population norm (possible problems). Lower proportions in these bands indicate better outcomes.

The proportion of children in the clinical range (bottom 10th percentile) is similar between the two trials. After adjustment for differences in family characteristics and risk factors, no statistically significant difference was detected between the trials.

For the broader threshold (bottom 25th percentile), the replication trial shows a substantially lower proportion of children in possible problem range at 12 months: 31% in the replication trial versus 53% in the RCT. After accounting for differences in observable characteristics, the adjusted difference between the trials is 51 percentage-points (replication trial lower) and highly significant. This result is robust across alternative analytical approaches. Regression analyses produced a somewhat smaller adjusted estimate—closer to the unadjusted difference—but it nevertheless remained highly significant. Given that children in the replication trial had substantially lower baseline social-

emotional standard scores on the Bayley social-emotional scale than those in the RCT, the 12-month outcomes are particularly encouraging.

Primary carer outcomes: Psychological Distress (K6)

Caregiver psychological distress increased slightly in the replication trial cohort over 12 months (+2.4 points) and remained essentially unchanged in the RCT treatment group (+0.2 points). After accounting for baseline differences, there was no statistically significant difference between cohorts in how psychological distress changed over 12 months.

The slight increase in distress within the replication cohort indicates that despite their children's participation in the program, primary carers continued to experience substantial life stressors. This underscores that the focus of EYEP is to address children's learning and development within the context of ongoing family adversity.

Primary carer outcomes: Parenting Daily Hassles

Parenting Daily Hassles are assessed in two dimensions: how frequently hassles occur and how intense or disruptive they feel for the parent. Higher scores indicate more problems, so decreases are more favourable.

For frequency of hassles, changes were small in both trials (replication: +0.9; RCT: +0.6), with no significant difference between groups. This indicates that the daily challenges of parenting did not substantially change over the first 12 months in either cohort.

For intensity of hassles—how stressful or disruptive these daily challenges feel—the pattern was more complex. Without adjustment, changes were similar in both cohorts (replication: +2.3; RCT: +3.2). However, after accounting for baseline differences, particularly when controlling for baseline outcome scores, a significant difference emerged favouring the replication trial. In this specification, the difference was 9.4 points, indicating that caregivers in the replication trial experienced a smaller increase (or greater reduction) in how intensely they felt daily parenting hassles.

This pattern—hassles occurring with similar frequency but being experienced as less intense—makes conceptual sense in families participating in the program but continuing to experience high levels of adversity. It may reflect improvements in child behaviour and self-regulation, better daily routines, increased connections to support services, or greater parental confidence in managing challenges because of participation in the program. Importantly, this occurred even though caregiver psychological distress overall did not differ between cohorts.

However, this finding should be interpreted cautiously because it was significant only in analyses that controlled for baseline outcome scores and not in other model specifications. The sensitivity to analytical approach means we cannot be as confident in this result as we are in findings that were consistent regardless of method.

6. Potential Implications for program implementation

The findings in this report highlight several implications for implementation of the EYEP model in replication settings. First, the replication trial continues to reach children and families experiencing substantial and often co-occurring adversity, and the baseline data indicate a cohort that is, on average, significantly more developmentally vulnerable than the RCT treatment group. This has practical implications for implementation of the model in early childhood education and care settings.

Children presenting with moderate to severe developmental delay and/or complex medical needs may require a higher level of individualisation, communication and regulation supports, and greater coordination with allied health and family services. In this context, maintaining consistency of staffing, predictable routines, and the relational "core" of EYEP becomes even more critical, as these program elements underpin children's engagement and capacity to benefit from enriched learning experiences.

Relatedly, the replication findings raise an important implementation issue about "fit" between EYEP's model and broader inclusion support systems. In mainstream kindergarten and ECEC settings, children with developmental delay or disability may be eligible for additional inclusion supports (for example, Victorian Kindergarten Inclusion Support and, in some contexts, national Inclusion Support Program mechanisms) that are designed to fund targeted adjustments, specialist input, and in some cases additional staffing to enable participation in ECEC. Where a sizeable subgroup of EYEP children present with higher developmental support needs, EYEP's favourable adult to child ratios—while essential—may not be sufficient to provide the intensity of individualised developmental support required without reducing educator availability to the broader group. In practice, educators can be required to provide frequent 1:1 support, effectively diluting relational capacity and potentially compromising the overall learning environment for other children in their care. To sustain both inclusion and program quality, a "braided" approach is recommended: retaining EYEP's trauma-informed relational pedagogy as the foundation, while systematically pursuing child-specific inclusion supports and early childhood intervention inputs (including NDIS-funded supports where eligible) that would ordinarily follow children in mainstream settings. Clear thresholds for additional resourcing, an internal "tiering" approach to intensity of need, and structured coordination with allied health and family services may be required to maintain both equitable participation for children with higher needs and stable ratios for the whole room.

Second, differences in the profile of families reached in the replication trial—relative to the RCT—underscore the importance of local referral pathways and service system dynamics. Where referrals are more strongly driven by health settings and less frequently involve statutory child protection or intensive family support pathways, the program may be reaching a somewhat different segment of the intended population. This raises two complementary implementation priorities: (i) strengthening and clarifying referral partnerships so that children and families with the highest levels of need are not inadvertently missing out, and (ii) ensuring that program messaging and eligibility processes are clear, consistent, and well understood by key referral agencies. In addition, the retention rate in the replication trial is lower than that in the original RCT after adjusting for differences in family risk factors and the rates also differ across replication trial centres. The fidelity assessments component of evaluation will be important in identifying barriers to referral and engagement, and in identifying opportunities to improve reach and equity of access.

Third, the emerging pattern of changes in outcomes provides guidance about what aspects of implementation may impact most in the first year. Where early gains are observed in protective factors and aspects of social–emotional functioning, these may represent proximal changes that support later developmental recovery, including in cognitive and language domains. This interpretation is consistent with the EYEP theory of change, in which improvements in children's regulation, relationships, and engagement are expected to precede and enable gains in learning-related outcomes. In practical terms, this reinforces the need to prioritise the quality and consistency

of educator–child relationships, emotionally available caregiving, and purposeful relational pedagogical practice, alongside active family engagement.

7. Next Steps

The replication trial continues to follow participants through their second and third years in the program. Data collection at 12- and 24-month timepoints remains ongoing for children across different stages of program participation.

The temporal pattern of impacts observed in the original EYEP RCT provides important context for interpreting the findings presented in this report and for understanding the trajectory expected in future assessments. In the RCT, modest initial effects at 12 months strengthened substantially and broadened across developmental domains over subsequent years. At 24 months, language impacts became statistically significant and social-emotional development showed marked improvements. By 36 months, all three main child development outcomes demonstrated large, statistically significant improvements: IQ (7.6 points), language (6.8 points), and behavioural problems (0.6 standard deviation reduction). For children with baseline scores below 90, impacts were even larger (Tseng et al., 2022).

This progressive emergence of impacts in the RCT suggests that modest or emerging effects at 12 months in the replication trial may strengthen substantially with continued program participation. The 12-month assessment presented in this report therefore provides critical early evidence confirming that the replication sites are achieving developmental and learning trajectories comparable to the RCT treatment group, while recognizing that the full extent of program impacts may not yet be apparent.

The next and final report (Report 3) will present the complete evaluation findings. This report will provide updated 12-month outcomes incorporating late recruitments and present 24-month outcomes for the full cohort. The 24-month data will be particularly important for assessing whether the pattern of strengthening impacts observed in the RCT is replicated across the three sites. Report 3 will compare both 12-month and 24-month outcomes with the corresponding RCT treatment group timepoints, employing regression-adjusted matching methods to account for any differences in baseline characteristics.

The cascading reporting approach employed across the evaluation ensures that each report captures both the longitudinal trajectory of earlier participants and the characteristics of newly engaged families. This approach provides increasingly robust evidence about the program's effectiveness while documenting the ongoing recruitment and engagement processes that would be essential during future program scaling.

The final report will synthesise findings across all evaluation components to provide comprehensive evidence on whether the EYEP model can be replicated across diverse settings while maintaining effectiveness comparable to the original RCT. This evidence will inform recommendations for targeted scale-up of the model as part of Australia's early childhood education and care service system and contribute to broader knowledge about effective approaches to replicating evidence-based early childhood interventions.

References

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

AEDC (Australian Early Development Census). (2024). AEDC National Report 2024. Retrieved from https://www.aedc.gov.au/

Alameddine, A. (2025). Developmental regression associated with PTSD in children: A poorly defined and understudied phenomenon. Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, 13(1), 54–56. doi:10.2478/sjcapp-2025-0006. PMID: 40486672; PMCID: PMC12145200.

Al-Ubaydli, O., List, J. A., & Suskind, D. L. (2019). The science of using science: Towards an understanding of the threats to scaling experiments (NBER Working Paper No. 25848). National Bureau of Economic Research.

Australian Government Department of Education. (2022). Belonging, Being & Becoming: The Early Years Learning Framework for Australia (Version 2.0). Canberra: Australian Government Department of Education.

Australian Government Department of Education. (2025). Building Early Education Fund. Retrieved from https://www.education.gov.au/early-childhood/about/building-early-education-fund

Australian Institute of Health and Welfare (AIHW). (2023). Australia's children. Canberra: AIHW, Australian Government.

Australian Institute of Health and Welfare (AIHW). (2024). Child Protection Australia 2022–23, Supplementary Data Table S2.3. AIHW Child Protection Collection 2022–23.

Barwick, M., Dubrowski, R., & Petrica, K. (2020). Knowledge translation: The rise of implementation. American Institutes for Research.

Bayley, N., & Glen, P. (2019). Bayley Scales of Infant and Toddler Development, 4th Edition: Australian and New Zealand Standardised Edition.

Biddle, N., Crawford, H., & Seth-Purdie, R. (2017). Risk burden, participation in early childhood education and care, and child outcomes. Australasian Journal of Early Childhood, 42(1), 42–52. doi:10.23965/AJEC.42.1.06

Blakemore, T., Strazdins, L., & Gibbings, J. (2009). Measuring family socio-economic position. In Department of Families, Housing, Community Services and Indigenous Affairs (Ed.), Australian Social Policy No. 8 (pp. 121–168).

Briggs-Gowan, M. J., & Carter, A. S. (2006). Brief Infant-Toddler Social and Emotional Assessment (BITSEA) Manual, Version 2.0. New Haven, CT: Yale University.

Cannon, J. S., Kilburn, M. R., Karoly, L. A., Mattox, T., Muchow, A. N., & Buenaventura, M. (2017). Investing early: Taking stock of outcomes and economic returns from early childhood programs. RAND Corporation.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. Teachers College Record, 112(3), 579–620.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. Implementation Science, 2, 40. https://doi.org/10.1186/1748-5908-2-40

Crnic, K. A., & Greenberg, M. T. (1990). Minor parenting stresses with young children. Child Development, 61(5), 1628–1637.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. American Journal of Community Psychology, 41(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. Journal of Economic Perspectives, 27(2), 109–132.

Guralnick, M. J. (2019). Effective early intervention: The developmental systems approach. Brookes Publishing.

Guy, S., Furber, G., Leach, M., & Segal, L. (2016). How many children in Australia are at risk of adult mental illness? Australian & New Zealand Journal of Psychiatry, 50(12), 1146–1160. https://doi.org/10.1177/0004867416640098

Halle, T., Metz, A., & Martinez-Beck, I. (Eds.). (2013). Applying implementation science in early childhood programs and systems. Brookes Publishing.

Hancock, K., Christiensen, D., & Zubrick, S. (2018). Development and assessment of cumulative risk measures of family environment and parental investments in the Longitudinal Study of Australian Children. Social Indicators Research, 133(1), 355–378. https://doi.org/10.1007/s11205-017-1607-3

Heckman, J. J. (2008). Schools, skills, and synapses. Economic Inquiry, 46(3), 289–324.

Heckman, J. J., & Karapakula, G. (2019). Intergenerational and intragenerational externalities of the Perry Preschool Project (NBER Working Paper No. 25889). National Bureau of Economic Research. Retrieved from https://www.nber.org/papers/w25889

Heifetz, R., Grashow, A., & Linsky, M. (2009). The practice of adaptive leadership: Tools and tactics for changing your organisation and the world. Boston, MA: Harvard Business Press.

Johnson, S., Tamanna, M., & Marlow, N. (2014). Using the Bayley-III to assess neurodevelopmental delay: Which cut-off should be used? Paediatric Research, 75, 670–674.

Jordan, B., Tseng, Y., Coombs, N., Kennedy, A., & Borland, J. (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.

Jordan, B., & Kennedy, A. (2019). The Early Years Education Program (EYEP) model. Changing the Trajectories of Australia's Most Vulnerable Children, Report No. 3. Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

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

Kim, H. Y., Cho, G. J., Ahn, K. H., Hong, S. C., Oh, M. J., & Kim, H. J. (2024). Short-term neonatal and long-term neurodevelopmental outcome of children born term low birth weight. Scientific Reports, 14, Article 2274. https://doi.org/10.1038/s41598-024-52154-9

LeBuffe, P. A., & Naglieri, J. A. (2012). Devereux Early Childhood Assessment for Preschoolers, Second Edition (DECA-P2). Lewisville, NC: Kaplan Early Learning Company.

MacKrain, M., LeBuffe, P. A., & Powell, G. (2007). Devereux Early Childhood Assessment for Infants and Toddlers: Technical Manual. Lewisville, NC: Kaplan Early Learning Company.

Metz, A., Halle, T., Bartley, L., & Blasberg, A. (2013). The key components of successful implementation. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), Applying implementation science in early childhood programs and systems (pp. 21–42). Brookes Publishing.

Minderoo Foundation. (2020). Thrive by Five: Early Years Initiative. Retrieved from https://www.minderoo.org/thrive-by-five/

Minderoo Foundation. (2024). Cost of Late Intervention Report [compiled by The Front Project]. Retrieved from https://www.minderoo.org/

Mitchell Institute. (2023). Childcare deserts in Australia. Victoria University.

Mitchell Institute. (2024). Research Spotlight: Australian Early Development Census 2024. Victoria University. Retrieved from https://www.vu.edu.au/mitchell-institute/

Murray, J., Menezes, A. M. B., Hickman, M., Maughan, B., Gallo, E. A. G., Matijasevich, A., Gonçalves, H., Anselmi, L., Susser, E., Wertz, J., & Barros, F. C. (2019). Childhood behaviour problems predict crime and violence in late adolescence: Brazilian and British birth cohort studies. Social Psychiatry and Psychiatric Epidemiology, 50(4), 579–589. https://doi.org/10.1007/s00127-014-0976-z

Parliament of Australia. (2025a). New Tax System (Family Assistance) Act 1999 [Amendment]. Retrieved from https://www.aph.gov.au/

Parliament of Australia. (2025b). Early Childhood Education and Care (Three Day Guarantee) Bill 2025. Retrieved from https://www.aph.gov.au/

Perry, B. D. (2002). Childhood experience and the expression of genetic potential: What childhood neglect tells us about nature and nurture. Brain and Mind, 3, 79–100.

Pilkington, R., Montgomerie, A., Grant, J., Gialamas, A., Malvaso, C., Smithers, L., et al. (2019). An innovative linked data platform to improve the wellbeing of children: The South Australian Early Childhood Data Project. In Australia's welfare 2019 data insights (Chapter 8). Canberra: Australian Institute of Health and Welfare.

Productivity Commission. (2024). A path to universal early childhood education and care (Inquiry report no. 106, Vol. 1). Canberra.

Productivity Commission. (2025). Report on Government Services 2025: Early childhood education and care. Australian Government.

Shaw, D. S., Sitnick, S. L., Reuben, J., Dishion, T. J., & Wilson, M. N. (2020). Transactional effects among maternal depression, neighbourhood deprivation, and child conduct problems from early childhood through adolescence: A tale of two low-income samples. Development and Psychopathology, 28(3), 819–836. https://doi.org/10.1017/S095457941500109X

Shonkoff, J. P. (2011). Protecting brains, not simply stimulating minds. Science, 333, 982–983.

Shonkoff, J. P. (2012). Leveraging the biology of adversity to address the roots of disparities in health and development. Proceedings of the National Academy of Sciences, 109(Supplement 2), 17302–17307.

Smart, D., Sanson, A., Baxter, J., Edwards, B., & Hayes, A. (2016). Home-to-school transitions for financially disadvantaged children: Final report (LSAC Research Report). Australian Institute of Family Studies.

Talarico, F., Liu, Y. S., Metes, D., Wang, M., Wearmouth, D., Kiyang, L., Wei, Y., Gaskin, A., Greenshaw, A., Janus, M., & Cao, B. (2023). Risk factors for developmental vulnerability: Insight from population-level surveillance using the Early Development Instrument. Journal of Medical Internet Research, 25, e210705. https://doi.org/10.2196/42116

Tseng, Y., Jordan, B., Borland, J., Clancy, T., Coombs, N., Cotter, K., Hill, A., & Kennedy, A. (2017). Changing the trajectories of Australia's most vulnerable children: Report No. 1. Participants in the trial of the Early Years Education Program. Melbourne Institute of Applied Economic and Social Research, University of Melbourne. Retrieved from https://melbourneinstitute.unimelb.edu.au/research/education/early-years-education

Tseng, Y., Jordan, B., Borland, J., Guillou, M., Hill, A., Kennedy, A., & Sheehan, J. (2019). 24 months in the Early Years Education Program: Assessment of the impact on children and their primary caregivers. Changing the Trajectories of Australia's Most Vulnerable Children, Report No. 4. Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Tseng, Y., & Borland, J. (2020). Can a targeted early years program improve outcomes for children facing extreme adversity? Melbourne Institute Research Insights, 31/20. Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Tseng, Y., Jordan, B., Borland, J., Clark, M., Coombs, N., Cotter, K., Guillou, M., Hill, A., Kennedy, A., & Sheehan, J. (2022). Changing the life trajectories of Australia's most vulnerable children: 36 months in the Early Years Education Program — Assessment of the impact on children and their primary caregiver. Report No. 5. Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Victorian Department of Human Services. (2007). Child development and trauma guide. Melbourne, Australia.

Wechsler, D. (2012). Wechsler Preschool and Primary Scale of Intelligence — Fourth Edition (WPPSI-IV). San Antonio, TX: The Psychological Corporation.

Yoshikawa, H., Rosman, E. A., & Hsueh, J. (2002). Resolving paradoxical criteria for the expansion and replication of early childhood care and education programs. Early Childhood Research Quarterly, 17(1), 3–27. https://doi.org/10.1016/S0885-2006(02)00129-1

Appendices

Appendix A. 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

Appendix B. Description of data source

The replication trial data

Information on participants in the replication trial in this report is taken from the referral information, baseline (initial) and 12-month follow-up data on children and their primary carers and centres administrative records of child's program attendance.

Baseline data collection occurred after a child's primary caregivers had provided consent to participate in the trial, generally no later than 3 months after children commenced orientation. The 12-month follow-up data collection window is between 12 and 15 months from children commenced orientation. However, in some circumstances, data were collected after the preferred window due to scheduling difficulties and resources constraints. Children centres sent children's attendance records to research team quarterly.

Attendance data

This report also incorporated the children's attendance data gathered by the centres. Information on both orientation and regular attendance was collected. In the original RCT, orientation and regular attendance records were not explicitly differentiated, and regular attendance was defined ad hoc by researchers based on patterns of initial attendance, where regular attendance is regarded as commencing once a child attends EYEP for at least three consecutive full days. For the replication trial, orientation data was explicitly collected based on the definition of orientation provided by PI. Additional clarifications from the data collection team were necessary to ensure consistent data. Consequently, the orientation attendance records and regular attendance data of the two trials were not entirely comparable.

The EYEP RCT

Baseline data on the characteristics of children and their primary caregivers are available for a maximum of 136 out of the 145 children for the EYEP RCT. In this report, we focus on the treatment group, where data are available for a maximum of 70 out of 72 children. Bayley standardised scores are available for 68 children. There was additional non-response on some individual variables and data items in the baseline data collection. See the data source section of EYEP report No. 1 (Tseng et al., 2017) for a more detailed description of the data collection.

The standardise child development measures for IQ and language were collected using the Bayley Scales of Infant and Toddler Development (Bayley), 3rd Edition for children 42 months or younger and Wechsler preschool and primary scale of intelligence (WPPSI-III) for children over 42 months of age. The replication trial was not able to employ the same scale for consistency due to the phasing out of Bayley and WPPSI III. The fourth edition of the Bayley and WPPSI are used.

Appendix C. Supplementary tables and figures

Table C1 Child characteristics and family risk factors by timing of recruitment (eligible referrals)

	Children consented to participate			
	Early cohort (T)	Late cohort (C)	Differences (T-C) difference	p-value
Demographics				
Gender (% boys)	52.0	51.7	0.3	0.971
Age at referral	1.51	1.38	0.13	0.361
Risk factors (% children with specific risk):				
Attachment/relationship issues	52.0	50.0	2.0	0.840
Alcohol or Substance use	48.0	35.6	12.4	0.201
Disability/complex medical issues (child or primary carer)	36.0	42.4	-6.4	0.521
Mental health issues	76.0	80.5	-4.5	0.630
Family violence, current or past	74.0	64.4	9.6	0.261
Social isolation (family, community, and cultural)	34.0	40.7	-6.7	0.462
Inadequate housing/ transience/ homelessness	18.0	26.3	-8.3	0.263
Parent/carer under 20 yrs old	6.0	7.6	-1.6	0.729
Lack of ability or willingness to prioritize child's needs	22.0	28.8	-6.8	0.387
Rejection of child	4.0	10.2	-6.2	0.165
Harsh, inconsistent discipline, neglect or abuse	10.0	8.5	1.5	0.771
Inadequate supervision	20.0	19.5	0.5	0.945
Total numbers of risk factors	5.98	5.89	0.09	0.884
Currently at risk of harm (%)	38.0	23.7	14.3	0.133
Number of children	50	118		

Note: Early cohort referred to those referred to the replication trial prior to August 2024 and late cohorts were those referred after.

[i] There were insufficient sample in RCT with extremely low IQ and language baseline score to further controlled for the differences. See

THE UNIVERSITY OF
MELBOURNE

MELBOURNE
INSTITUTE