JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
2021, AHEAD-OF-PRINT, 1-20
https://doi.org/10.1080/19345747.2021.1931999

# Replication and Extension of Family-Based Training Program to Improve Cognitive Abilities in Young Children

Rachel R. Romeo [a,b], Julia A. Leonard [a,c], Ethan Scherer[d], Sydney Robinson[a,e], Megumi Takada[a,f], Allyson P. Mackey[a,c], Martin R. West[d], and John D. E. Gabrieli[a,d,g]

[a] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA [b] Department of Psychology, Harvard University, Boston, Massachusetts, USA [c] Division of Psychology, College of Arts and Sciences, University of Pennsylvania, Philadelphia, Pennsylvania, USA [d] Graduate School of Education, Harvard University, Boston, Massachusetts, USA [e] Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, Pennsylvania, USA [f] Graduate School of Education, Stanford University, Stanford, California, USA [g] Integrated Learning Initiative, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

**ABSTRACT**

Childhood socioeconomic status (SES) is associated with persistent academic achievement gaps, which necessitates evidence-based, scalable interventions to improve children's outcomes. The present study reports results from a replication and extension of a family-based training program previously found to improve cognitive development in lower-SES preschoolers. One hundred and one primarily low-SES families with 107 children aged 4–7 years were randomly assigned to the intervention or passive control group. Intent-to-treat regression models revealed that children whose families were assigned to the intervention group did not exhibit significant benefit on composite measures of nonverbal IQ, executive functioning, or language skills, though post-hoc analyses suggested marginal improvement on the fluid reasoning subcomponent of nonverbal IQ. Treatment-on-treated models revealed a significant positive effect of intervention attendance on fluid reasoning and a negative effect on vocabulary. We discuss potential causes for the non-replication, including differences in the sample composition, size, and assessment choices. Results suggest the need to more broadly assess scalable interventions with varying populations and ensure appropriate cultural and geographical adaptations to achieve maximum benefits for children from diverse backgrounds.

**ARTICLE HISTORY**

Received 9 May 2020
Revised 12 February 2021
Accepted 23 April 2021

**KEYWORDS**

Family-based intervention, cognitive development, early childhood, replication, socioeconomic status

**CONTACT** Rachel R. Romeo rromeo@mit.edu Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA.

* These authors contributed equally to this article.



Socioeconomic status (SES) is strongly related to many important life outcomes including cognitive abilities, socioemotional skills, and achievement metrics (Bradley & Corwyn, 2002; Duncan & Brooks-Gunn, 1997; Ursache & Noble, 2016; White, 1982). SES is a multifaceted construct indexing an individual's access to educational and financial resources, and commensurate social benefits. Children from lower-SES backgrounds frequently arrive at kindergarten with significant gaps in core school readiness skills, including language, reasoning, and executive functioning (Garcia & Weiss, 2017; Lee & Burkam, 2002; Ramey & Ramey, 2004), which often snowball into significant academic achievement gaps throughout school and beyond (Duncan et al., 2007; Duncan & Murnane, 2011). These persistent disparities have garnered many efforts to develop effective, scalable interventions to reduce SES achievement gaps at the start of a child's education.

One subset of interventions for children facing early adversity are family-based, or two-generation, approaches, which involve parent training programs alongside a parallel component for children. Rather than exclusively focusing on child enrichment, such programs aim to strengthen the resources and skillsets of parents to support their children's development (Shonkoff & Fisher, 2013). Evaluations of various family-based programs provide a strong evidence base for both short- and long-term effects on children's cognitive and neural development, emotional/behavioral well-being, and ultimate life outcomes (Neville et al., 2015), as well as commensurate improvements in parental stress and mental health and interactive family well-being (Shonkoff & Fisher, 2013). Given such successful outcomes, family-based intervention programs serve as a promising approach for impactful, scalable interventions to improve outcomes for children from low SES backgrounds.

One previously successful family-based intervention program is Parents and Children Making Connections—Highlighting Attention (PCMC-A; Neville et al., 2013). PCMC-A is a nine-week (1 introduction, 8 content) family-based training program for families of preschoolers that combines didactic, interactive, separate small-group parent and child training sessions focused on increasing children's attention and cognition by way of improving family stress, contingency-based discipline, parental language use, and emotion regulation. The theory of change focuses on engaging parents to support child development, broadly construed. Specifically, program creators hypothesized that cultivating positive adult models and adult responsiveness across multiple contexts of parent-child interactions would foster children's attention and self-regulation skills while reducing problem behaviors, which would in turn improve children's learning across a broad range of cognitive domains.

PCMC-A was designed for families with children (3–5 years old) attending Head Start preschools in Oregon and was assessed via a randomized controlled trial in a sample of typically developing (no developmental disorders, psychoactive medications, or services), monolingual English, primarily (62%) white/Caucasian children living at or below the poverty line ($22,050/year for a family of four in 2009) across the state of Oregon (Neville et al., 2013). Neville et al. (2013) found that PCMC-A led to significant improvements in self-reported parental stress, parent-child interactions, children's verbal and nonverbal cognition, children's parent-reported social skills and problem behaviors, and plasticity in electrophysiological measures of children's brain activity during a selective attention task as compared to both active and passive control groups. PCMC-A's reported efficacy and practicality make it a promising intervention to bring to other communities.

To assess the value of expanding PCMC-A at a wider scale, it is important to test whether it is feasible to replicate the intervention with fidelity in new populations. Common pitfalls of intervention programs include effects being specific to certain demographics or programs being logistically difficult to implement at scale. For example, one of the most successful early life intervention programs, Reach Up, a home visiting program in Jamaica, showed impressive educational, social, and health benefits for children 22 years later (Gertler et al., 2014; Walker et al., 2011). However, when researchers tried to implement a similar program in Colombia, the positive effects faded after two years (Andrew et al., 2018). The authors cite lack of piloting, high staff turnover, and implementation barriers as potential reasons behind their null long-term effects. In the United States, researchers have spent the past two decades figuring out the key ingredients for successful scaling of programs such as "The Incredible Years" and "The Nurse-Family Partnership," which have been shown to improve children's socioemotional outcomes. Researchers state gradual expansion, selection of good staff, high quality training, and ongoing program evaluation and monitoring as key to successful scaling of these programs (Olds, 2006; Webster-Stratton & Herman, 2010). Cultural norms can also impact the effectiveness of scaling interventions. In the USA, psychotherapy and health interventions that are culturally adapted to specific communities (e.g. the Latinx community) are more effective than nonadapted interventions, in part because they increase participant engagement and retention (Acevedo-Polakovich et al., 2017; Smith et al., 2011). There has been less research directly comparing standard versus culturally adapted parenting based interventions in the United States. However, studies have found that cultural adaptations for Latinx communities of evidence-based interventions developed for Euro-American families, like the Parent Management Training (the Oregon Model), a clinical and preventative family intervention, have been efficacious (Meija et al., 2017; Parra-Cardona et al., 2017). Before expanding PCMC-A to other sites within the United States or internationally, it is important to test its effectiveness and feasibility in another geographic location with different demographics from the original sample.

The present study aimed to replicate and extend the results of PCMC-A in a different kind of sample, namely Kindergarten and Pre-Kindergarten students from urban charter schools serving predominantly families from lower SES backgrounds in Boston. Families were randomized to either receive the intervention or to a school-as-usual control group, and students were assessed for changes on several metrics of verbal and nonverbal cognition and executive functioning. While participants in this study were also primarily from a lower SES background (79% on free or reduced price lunch, as qualified by household income <185% of the federal poverty line, equivalent to 127% of the regional Supplemental Poverty Measure), sample demographics differed from the original implementation in that children in the present study were on average a year older (4–6 years instead of 3–5 years), were from an urban environment, were primarily from racial/ethnic minority groups (95% non-white), and were not excluded based on home language use (50% non-monolingual English), history of developmental disorders or medication use, or special education/disability status (9% with Individualized Education Plans). These modifications are important because, should the intervention scale to reach families across the United States, it is crucial to know that it provides benefits for lower-SES families from varying backgrounds. Given the wide-ranging benefits across multiple cognitive domains in the original study, and the detailed cultural adaptations made in consultation with local parents, we expected to find similar benefits of PCMC-A within this context.

## Materials and Methods

### *Participants*

Participants were 107 children from 101 families (including four sets of twins and two sibling pairs aged one year apart). Children ($n = 64$ male, 60%) were in pre-Kindergarten ($n = 30$, 28%) or Kindergarten ($n = 77$, 72%), both of which are full-day programs in the metro region of implementation, which offers universal Pre-Kindergarten in an elementary school setting. Children ranged in age from 4 years, 6 months to 7 years, 1 month ($M = 5.78$ years, $SD = 0.59$ years), at the first assessment time point, pre-randomization (all but one child were younger than 6.5 years, with the addition of one 7-year-old who had repeated pre-Kindergarten). Fifty-two (49%) of children participated in additional in-lab cognitive and neuroimaging assessments as part of another study not reported here.

Participants were recruited from eight over-subscribed public charter schools in metropolitan Boston. The eight urban charters schools came from six Charter Management Organizations (three schools were separate sites of one organization), all of whom employed a rigorous approach to urban education with high academic expectations, inquiry-focused and standards-based instruction, and ultimate goals of preparing children for attending college. All children in either pre-Kindergarten ("K1") or Kindergarten ("K2") in each school were invited to participate, through flyers, letters home, and recruitment events held on parent-teacher nights and other family-based school events. Research staff were supported by consultants from a local education-based nonprofit and school-employed family liaisons to assist in personalized outreach efforts. The study we attempted to replicate and extend (Neville et al., 2013) had three groups of children ($n = 66$ intervention, 38 passive control, 37 active control). Since effects were similar for active and passive control groups, the present study only includes a passive control group. An a priori power analysis based on the average of reported effect sizes for cognitive assessments (intervention versus passive control only, $d = 0.4$, power = 0.8, $\alpha = 0.05$) determined a need for 156 (one-tailed) to 200 (two-tailed) participants total using the same analytic approach. Even though the present sample size of $n = 107$ is smaller than that recommended by the a priori power analysis, it is slightly larger than the final sample of the original study (total $n = 104$, intervention and passive control combined). Furthermore, post-hoc sensitivity analyses using the present analytic methods with sample stratification (see below) determined that the present study was powered to detect an effect as small as 0.22 standard deviations for IQ and 0.07 for language, less than half that of the original study.

Data from 101 children (95 families) were able to be linked to the state administrative data including demographic information. The racial and ethnic identities of children were as follows (see Table 1 for data by group): Hispanic ($n = 55$), African American ($n = 36$), White ($n = 5$), Asian ($n = 2$), American Indian ($n = 1$), or multiple races ($n = 2$). Eighty children (79%) qualified for free or reduced-price lunch (FRPL), defined as having an annual household income (before taxes) at or below 185 percent of the federal poverty level threshold, which, given the high cost of living in the region of implementation, is equivalent to 127% of the regional Supplemental Poverty Measure (US Census Bureau, 2016 estimates). Fifty-one children (50%) were identified by their schools as English Language Learners (ELL), and 9 (10%) had an Individualized Education Plan ($n = 4$ for speech/language, 1 for autism, 5 unknown reason).

**Table 1**. Baseline characteristics of participants. (Table view)

| | All students ($N = 107$) | Treatment ($n = 54$) | Control ($n = 53$) | Difference regressions |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Age in months (SD) | 68.82 (7.04) | 68.98 (6.87) | 68.66 (7.29) | 0.17 (1.382) |
| Male | 0.60 | 0.57 | 0.62 | –0.05 (.096) |
| Grade Kindergarten 2 | 0.71 | 0.69 | 0.74 | –0.05 (.089) |
| Race/Ethnicity | | | | |
| White | 0.05 | 0.04 | 0.06 | –0.02 (.044) |
| Black | 0.36 | 0.35 | 0.38 | –0.03 (.097) |
| Asian | 0.02 | 0.02 | 0.02 | <0.01 (.028) |
| Hispanic | 0.54 | 0.58 | 0.50 | 0.08 (.1) |
| Other[a] | 0.03 | 0.02 | 0.04 | –0.02 (.034) |
| Limited English proficiency | 0.50 | 0.54 | 0.47 | 0.07 (.101) |
| Individual education plan | 0.10 | 0.10 | 0.10 | <0.01 (.06) |
| Free/reduced price lunch | 0.79 | 0.79 | 0.80 | <0.01 (.082) |
| Language | 0.00 | –0.09 | 0.09 | –.17 (.185) |
| Nonverbal IQ | 0.00 | –0.10 | 0.11 | –.2 (.243) |
| Matrix reasoning | 0.00 | –0.13 | 0.13 | –.27 (.22) |
| Working memory | 0.00 | –0.08 | 0.08 | –.14 (.255) |
| Processing speed | 0.00 | –0.03 | 0.04 | –.05 (.253) |
| Executive function | 0.00 | –0.07 | 0.07 | –.12 (.209) |
| HKTS | 0.00 | –0.07 | 0.07 | –.13 (.208) |
| H + F Incongru Acc | 0.00 | 0.05 | –0.05 | .12 (.19) |
| H + F Mixed Acc | 0.00 | –0.10 | 0.10 | –.17 (.206) |

*Notes.* In columns 1–3, age is presented as mean (standard deviation) and all other demographic characteristics are proportions. In Massachusetts, Kindergarten 2 is equivalent to Kindergarten in most other states (beginning at age 5 years). For Race/Ethnicity, Limited English Proficiency, Individual Education Plan, and Free/Reduced Price Lunch, the sample includes only the 101 students that could be linked to school administrative data. All cognitive measures were normalized to have a mean of zero and standard deviation of one. In column 4, each row presents an estimate of the coefficient on treatment status from a separate regression model. All regressions include randomization strata fixed effects. We cluster our standard errors at the strata and the errors are reported in parentheses. We also tested the component of the nonverbal IQ and executive function composite, as well as distributional differences and none of these were statistically significant.

[a]Other includes Native American and two or more races. HKTS is the Head-Knees-Toes-Shoulder task. H + F stands for Hearts and Flowers. No estimates were statistically significant.

Figure 1 shows a modified CONSORT flow diagram of participation, and Table 1 reports the participant summary statistics overall and by treatment group. Participants were randomly assigned at the family level to participate in the intervention (Figure 1). Randomization was stratified by whether families participated in the additional cognitive and neuroimaging assessments and whether they were recruited from the school with the greatest number of participants (4 total categories). Table 1 column 4 reports the difference between treatment and control groups. Each row represents a separate regression model where the outcome variable is the child's characteristic, which is regressed on treatment status and a series of dummy variables for the stratification groups. Thus, the coefficient represents the average within stratification difference between treatment and control. Stratification ensured that the experiment assessed the contrast between treatment and control within the four groups because students who volunteered to participate in additional testing might be different than those who did not, and students in the school where we were able to recruit a higher proportion of families could be different than other schools in our sample. For logistical reasons, families were randomized prior to the completion of the baseline assessments scoring. By chance, the treatment group students tended to perform worse on the measures, though neither the mean nor distributional difference were statistically significant. Of the 54 children (51 families) assigned to the intervention group, 32 children (from 30 families) completed the intervention by attending all intervention sessions (59% completion rate). Four children (from 4 families) attended some portion of the intervention (1, 3, 4, or 5 sessions), and 18 children (from 17 families) did not attend any sessions. The elapsed time between pre and post assessment sessions ranged from 74 to 99 days ($M = 84.58$, $SD = 5.10$ days), and did not differ between assigned groups ($t(105) = 0.28$, $p = 0.27$).

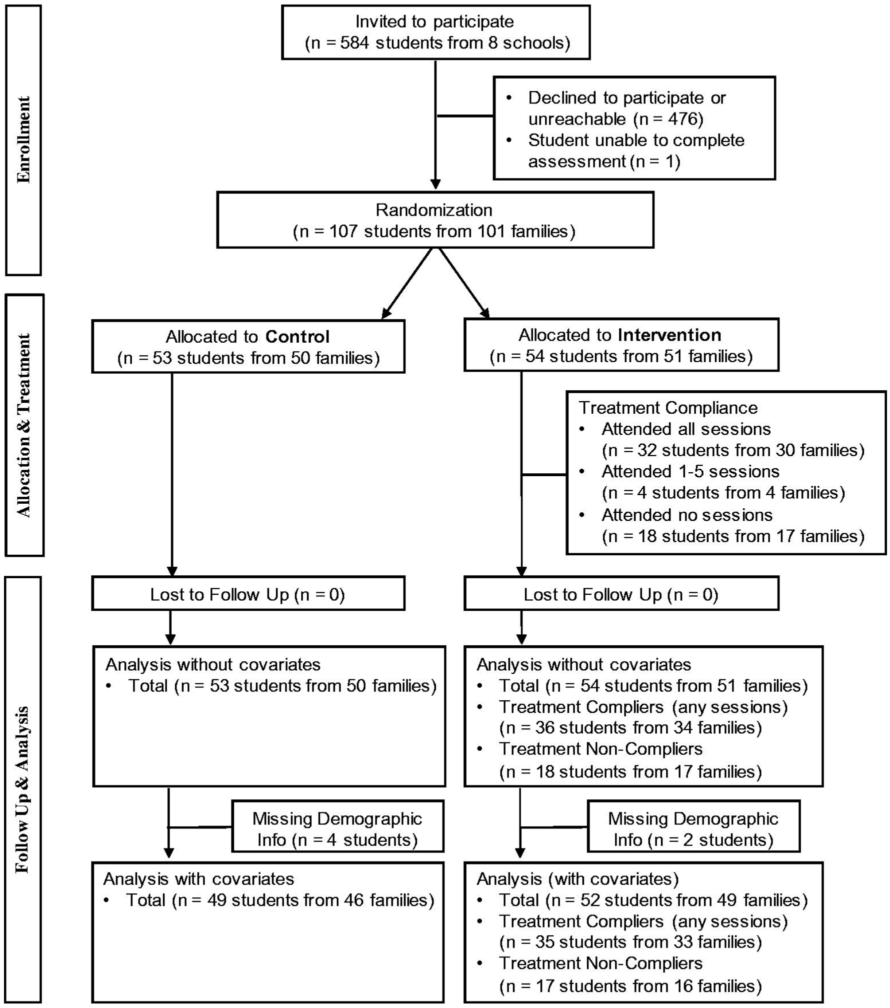


**Figure 1**. Modified CONSORT flow diagram of study participation.

## *Training Program*

The intervention program was a near-exact replication of the Parents and Children Making Connections – Highlighting Attention (PCMC-A) program designed by Neville et al. (for complete description, see Neville et al., 2013, Supplementary Material). Briefly, this program consists of eight weekly 2-hour small group sessions (4–5 families per group) preceded by one introductory session for enrolled children and any parents/caregivers who would like to attend. Parents assigned to the intervention had the choice of attending groups led in either English ($n = 42$) or Spanish ($n = 12$), which occurred on either a weekday evening or Saturday. The child component of the intervention was delivered in English, to match the children's language of instruction at school.

The Spanish version of the intervention was originally adapted for Latinx families, who were primarily of Mexican origin at the program creation site, as well as a separate replication site in Colombia, according to a rigorous cultural adaptation model (for more information, see Pakulak et al., 2017). While there were no changes to curriculum content, the metaphors used to convey the content were occasionally modified. For example, in discussing how to revise the parenting tools you inherit from your parents and grandparents and build them up stronger, the English version used a metaphor of restoring an old family car, while the Spanish version used a metaphor of adapting an old family recipe. Because the Spanish speaking population at the current replication site were primarily of Dominican origin, some word choices for translation were changed, but no additional adaptations were made. No changes were made to the English version of the parent curriculum; however, minor changes were made to the child curriculum to accommodate differences in the availability of materials (e.g., an activity that used snails to represent patience was replaced with worms, which are more prevalent at the replication site). All interventionists were trained to reliability by study staff from the original PCMC-A program (80 hours for parent facilitators, and 20 hours for child facilitators). Additionally, all three parent facilitators were bilingual/bicultural and learned both the English and Spanish versions of the curriculum to ensure consistency across sessions.

Parents and children attended separate but simultaneous sessions at the child's school or a nearby participating school. Trained Parent Facilitators led caregivers in an interactive curriculum of strategies to target family stress regulation, contingency-based discipline, parental responsiveness and language use, and facilitation of child attention through links to child training exercises. The first week was an introductory session, and the remaining eight sessions each focused on a different topic of parent-child interaction. Parents were given "homework assignments" to try strategies at home and report their experiences in the next session. Facilitators called participating parents weekly for a 15-minute "check-in" to discuss implementation and effectiveness of the strategies. Meanwhile, enrolled children participated in a 45-minute child curriculum called "Brain Train," which consisted of multi-sensory activities and games aimed to increase self-regulation of attention and emotion states. As with the parents, each week introduces a new topic and games and was led by pairs of trained Child Facilitators. After completing Brain Train, enrolled children received general childcare (along with any siblings or additional un-enrolled children) until the parent group concluded. Both the parent and child curricula were highly structured, such that it was very difficult for any deviations to occur. During sessions, meals were provided for all family members and other individuals who attended. Child care was provided for siblings and participating children after the Brain Train session. Families received a $10 travel reimbursement each week and were entered into a raffle for $25 awarded to one family per session.

Throughout the intervention, facilitators participated in a fidelity protocol designed by PCMC-A staff. Specifically, the three parent facilitators and the six child facilitators held weekly meetings with the intervention designers to debrief on that week's session, ensure that the curriculum was properly and faithfully delivered and consistent across groups, and to prepare for the following week's sessions. Specifically, the discussion protocol included confirming the presence of principal intervention features, including both the content of the curriculum (e.g., correct information, time allotted for each section), as well as the process of delivery (e.g., the facilitator's affect, promotion of discussion, and handling of challenges). No curriculum deviations were noted in any of the weekly meetings.

### *Cognitive Assessments*

Assessments of cognitive ability were conducted within two weeks immediately before and after the 8-week training period and delivered by testers blind to experimental condition. Cognitive skills were assessed because they were an end target for intervention outcomes, they can be measured objectively, and assessments could be administered during school hours, so as to reduce additional burdens on parents. The theory of change involves indirectly influencing child cognitive skills through parent-mediated improvements to children's attention and self-regulation skills, which were assessed in the original study through parental report of child problem behaviors and electrophysiological measures of selective attention. While the executive functioning assessments described below measure components of attention and behavioral regulation, only 50% of the present sample completed parental report of child problem behaviors in an optional lab-based session. To avoid mixing of samples and sub-samples, the results of the additional lab-based measures will be reported elsewhere.

All testing took place in a quiet private or shared room at each student's respective school, with the exception of one school in which the only place available was a nook outlet off a hallway. Since English was the primary language of instruction at each school, all child assessments were delivered in English. The cognitive assessments were chosen to test similar constructs that showed improvement after the intervention implemented by Neville et al. (2013), including nonverbal intelligence and receptive language. In addition, two measures of executive function were included because the intervention targeted many executive function skills. All data were double scored by coders blind to condition. Discrepancies were resolved by a third blind coder. Scores on all cognitive assessments were highly intercorrelated (Table 2).

To assess nonverbal intelligence (IQ), children completed the Matrix Reasoning, Picture Memory, and Bug Search subtests of the *Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition* (WPPSI-IV; Wechsler, 2012). The WPPSI is a well-validated and standardized test of intellectual functioning designed for children ages 2.6–7.7. Matrix Reasoning measured fluid reasoning ability by having children choose which image completed a visual puzzle, which became progressively more difficult with each trial. Picture Memory assessed visual working memory ability by having children remember pictures for five seconds before identifying them from an increasing number of distractors. Bug Search evaluated processing speed by having children complete as many trials of finding a matching target bug out of five options in two minutes. Due to only one version of the WPPSI-IV being available, the same stimuli were used for pre- and post-testing. To be comparable to the study by Neville et al. (2013), we constructed a composite non-verbal IQ measure by averaging the z-scored raw scores from all three subtests. We also report the individual components for completeness. Bug Search was invalid for two children, one each at pre- and post-testing (valid $n = 106$ at each time point), so for these two children, averages from the other two measures were used to create the nonverbal IQ composite (composite valid $n = 107$).

Language skills were assessed with the *Peabody Picture Vocabulary Test, Fourth Edition* (PPVT-4; Dunn & Dunn, 2017), a well-validated measure of receptive vocabulary. Experimenters read aloud progressively more difficult words, and children pointed to the corresponding picture out of four given pictures. Form A was administered before the intervention, and Form B was administered after the intervention to avoid memorizing target words. The be consistent with the other measures, PPVT raw scores were z-scored to be used as verbal IQ measure. PPVT-4 data was invalid for two children, one each at pre and post testing (valid $n = 105$).

Executive function was measured with both the Head-Knees-Toes-Shoulders (HKTS) task (McClelland et al., 2014) and the Hearts and Flowers version of the dots task (Davidson et al., 2006). HKTS taps into the three main dimensions of executive function: working memory, attentional flexibility, and inhibition, and has been shown to have high inter-rater reliability and validity in assessing children’s self-regulation in diverse samples of children (McClelland et al., 2007, 2014; von Suchodoletz et al., 2013). During the introduction of the task, children were told to respond to a command such as “touch your toes.” In the first part of the task, children were asked to do the opposite of the command; specifically, they were instructed to touch their head when the experimenter told them to touch their toes and vice versa. In the second part, two additional commands are added (touching shoulders and knees) and in the final part, the rules are changed (i.e. now head goes with knees). The measure includes 30 items and children were given a score of 0 for an incorrect response, 1 for a self-correct, and 2 for a correct response bringing the max score to 60. HKTS data was invalid for 3 children, one at pre-testing and two at post-testing (valid $n = 104$).

The second executive function measure, Hearts and Flowers, was chosen because preschool age children have improved on this task after another executive function intervention (Diamond et al., 2007). Like the HKTS, Hearts and Flowers also taps into all three main dimensions of executive function. It was administered on a touch screen computer using the Presentation program by Neurobehavioral Systems (Berkeley, CA). Children responded by using their index finger to touch buttons on the right or left side of the screen below the stimulus images. All children rested their hand on a handlebar that was adjusted so that each child was index finger distance away from the screen. Children completed a practice round of pressing buttons on the screen before starting the experiment to ensure familiarization with the process. The overall structure of the task consisted of three blocks: a congruent block of 12 trials, an incongruent block of 12 trials, and a mixed block of 49 trials. In all conditions, a red heart or flower appeared on the right or left side of the screen. In the congruent trials, children were told to press the button on the same side as the heart as a measure of working memory. In the incongruent block, children were told to press the button on the opposite side of the flower, requiring them to inhibit the tendency to respond on the side where the stimulus appeared and working memory of the new rule. In the mixed block, congruent and incongruent trials were intermixed, requiring all three components of executive function. For all conditions, stimuli were displayed for 500 milliseconds (ms) with 1500 ms to respond and an interstimulus interval of 500 ms. All children received up to 12 practice trials before each block. The two main outcome measures were accuracy (proportion correct) for the incongruent and mixed blocks. Any response faster than 200 ms were considered to be anticipatory (Davidson et al., 2006) and excluded from analyses. The first trial of the mixed block was additionally excluded from analyses. Data from Hearts and Flowers was missing for one child at pre-testing (valid $n = 106$). Similar to the nonverbal intelligence, in order to compare results to Neville et al. (2013) we create an executive function composite measure by averaging the z-scored total scores from HKTS, Hearts and Flowers incongruent proportion accuracy, and Hearts and Flowers mixed proportion accuracy, using two of the measures where one was missing/invalid (composite valid $n = 107$). We also report the results separately by assessment. See Table 2 for relationships amongst all cognitive measures at time 1 controlling for age.

## *Analytic Approach*

A series of multivariate regression models were used to estimate the impact of the intervention on each cognitive outcome measure (the non-verbal IQ, verbal IQ, and executive function composites and each component in post-hoc tests) all of which were standardized to have a mean of zero and standard deviation of one at baseline. We then used the baseline mean and standard deviation for time 2.[1] Each regression model controls for the student's baseline outcome measure.[2] To account for the stratified randomization process, each model also includes indicators of whether the student participated in additional in-lab assessments and came from the school where we recruited the most students. As suggested by Athey and Imbens (2017), we cluster our standard errors at the level of stratification.[3]

What we refer to below as Model 1 includes only those two variables and represents an intent-to-treat estimate of the intervention's effects. To obtain a more precise estimate of this parameter, Model 2 adds controls for the student's age and whether prior to randomization the family requested to receive the training in Spanish or English (noted on consent). Finally, because several families either did not attend the sessions or were unable to attend all the sessions, Model 3 employed a 2-stage least squares regression to estimate treatment-on-treated (TOT) effects. More specifically, we use the random offer of admission as an instrument for a variable measuring the proportion of sessions the family attended. The estimation strategy in Model 3 assumes a linear relationship between the proportion of sessions attended and outcomes. Under that assumption, the coefficient estimate can be interpreted as the effect of attending all sessions as compared to attending none. For Model 3, we also show the first-stage prediction and F-statistic to test for a weak first-stage instrument.

**Table 2**. Partial correlation amongst standardized cognitive measures at time 1 controlling for age. (Table view)

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Matrix reasoning | | | | | | |
| 2. Processing Speed | 0.36** | | | | | |
| 3. Working Memory | 0.39** | 0.33** | | | | |
| 4. PPVT | 0.52** | 0.34** | 0.29* | | | |
| 5. HKTS | 0.49** | 0.38** | 0.28* | 0.48** | | |
| 6. H + F incongruent acc | 0.24* | 0.31* | 0.11 | 0.19 | 0.30* | |
| 7. H + F mixed acc | 0.34** | 0.43** | 0.32* | 0.38** | 0.46** | 0.35** |

Matrix reasoning, processing speed, and working memory come from the Matrix Reasoning, Bug Search, and Picture Memory subscales of the WPPSI-IV. PPVT stands for the *Peabody Picture Vocabulary Test, Fourth Edition* which assess language skills. HKTS stands for Head-Knees-Toes-Shoulders executive function measures. H + F stands for Hearts and Flowers executive function measure. Incongruent acc = incongruent block accuracy, and mixed acc = mixed block accuracy. All $p$ values are FDR corrected. *$p < .05$, **$p < .01$.

## Results

The PCMC-A group did not exhibit significant improvement compared to the control group on the nonverbal IQ composite measure ($b = 0.07$, SE =0.118, $p = 0.61$) or the executive function composite measure ($b = 0.03$, SE = 0.162, $p = 0.85$) (Table 3). The PCMC-A group exhibited slight, non-significant declines in the verbal IQ measure ($b = -0.08$, SE = 0.032, $p = 0.10$). The addition of explanatory variables in Model 2 resulted in little change in the coefficient or increase in the precision of the estimates. Finally, using the randomization as an instrument for session attendance scaled up the coefficients for nonverbal IQ and executive function (although not to a point of statistical significance), and scaled down the coefficient for verbal IQ, such that PCMC-A attendance exhibited a significant *negative* effect on the language measure ($b = -0.12$, SE = 0.048, $p = .014$).

**Table 3**. Effect of PCMC-A on the composite cognitive assessments. (Table view)

| | Beta coefficient (SE) | | | First-stage F-statistic |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Language | –0.08* (0.032) | –0.07 (0.032) | –0.12** (0.048) | 13.18 |
| Nonverbal IQ composite | 0.07 (0.118) | 0.06 (0.108) | 0.11 (0.137) | 12.9 |
| Executive function | 0.03 (0.162) | 0.02 (0.166) | 0.03 (0.231) | 13.38 |

*Note.* Each cell reports an estimated treatment effect from a separate regression; standard errors clustered at the stratification level are reported in parentheses. Model 1 controls for the baseline outcome and randomization strata fixed effects. Model 2 adds controls for age and preferred language of the sessions. Model 3 uses treatment status as an instrument for the proportion of sessions attended and includes all the controls in Model 2. Column (4) reports the first-stage F-statistic. *$p<0.10$, **$p<0.05$.

Follow-up analyses explored whether the intervention had any impact on the individual measures that contributed to the non-verbal IQ and executive function composite measures. There were no effects of the intervention on working memory (Picture Memory; Table 4), processing speed (Bug Search; Table 4) or any executive function measure (HKTS or Hearts and Flowers inhibition or mixed conditions). However, there was a trend for the intervention group improving more than the control group on fluid reasoning (Matrix Reasoning; $b = 0.25$, SE = 0.092, $p = 0.07$; Table 4). This effect was driven by larger improvements in the intervention group than the control group between Time 1 and Time 2. Importantly, as noted in Table 1 there were no differences between groups at Time 1 on all assessments. Model 2 shows these trend level effects became significant with the addition of child age and preferred language of the adult sessions ($b = 0.23$, SE = 0.065, and $p = 0.04$). Among the families that attended more sessions (Model 3), there were improvements on matrix reasoning ($b = 0.38$, SE = 0.146 and $p = 0.01$), but not on the other cognitive measures. Finally, we examined the relation of a child's age in months to treatment outcomes, but none of these models showed any significant relation between age and pre-post differences on test scores.

**Table 4**. Effect of PCMC-A on the component assessments of nonverbal IQ and executive functioning. (Table view)

| | Beta coefficient (SE) | | First-stage F-statistic | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Matrix reasoning | 0.25* (0.092) | 0.23** (0.065) | 0.38*** (0.146) | 12.72 |
| Working memory | –0.14 (0.290) | –0.14 (0.280) | –0.23 (0.437) | 12.25 |
| Processing speed | –0.02 (0.094) | –0.02 (0.096) | –0.03 (0.133) | 13.16 |
| HKTS | 0.02 (0.082) | 0.01 (0.092) | 0.02 (0.125) | 12.35 |

| | Beta coefficient (SE) | | First-stage F-statistic | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| H + F Incong. Acc. | 0.27 (0.219) | 0.27 (0.216) | 0.45 (0.317) | 13.44 |
| H + F Mixed Acc. | –0.19 (0.187) | –0.21 (0.167) | –0.35 (0.304) | 14.54 |

*Note.* Each cell reports an estimated treatment effect from a separate regression; standard errors clustered at the stratification level are reported in parentheses. Model 1 controls for the baseline outcome and randomization strata fixed effects. Model 2 adds controls for age and preferred language of the sessions. Model 3 uses treatment status as an instrument for the proportion of sessions attended and includes all the controls in Model 2. Column (4) reports the first-stage F-statistic. HKTS is the Head-Knees-Toes-Shoulder task. H+F stands for Hearts and Flowers. Statistical significance level: *$p<0.10$, **$p<0.05$, ***$p<0.01$.

## Discussion

We found more limited effects of the PCMC-A intervention on children's cognition than did the original evaluation. Children who were assigned to the intervention improved on fluid reasoning and this effect was larger in those who attended the intervention. However, children who attended the intervention did not exhibit the same increases in language over the 3-months as children in the control group. All other outcomes, including the other nonverbal IQ measures and all executive functioning measures, resulted in null findings, indicating a general failure to replicate the positive effects of PCMC-A under the present conditions. Although we recruited fewer students than our *a priori* power analyses suggested, our sample size was powered to detect effects almost half the size of those reported in the original study for IQ and language (see methods), and thus we do not believe our null findings are driven by sample size alone.

The positive impact on children's fluid reasoning is consistent with results from the original study (Neville et al., 2013), which found that the intervention led to improvement on a composite non-verbal IQ measure comprising fluid reasoning, quantitative reasoning, and working memory. However, the authors did not report results broken down by non-verbal IQ component, so it is unclear if the originally reported positive impact of PCMC-A on non-verbal IQ was driven by positive effects in reasoning specifically. Although this intervention targeted many aspects of non-verbal IQ, it is possible that it was most effective at improving reasoning due to its focus on problem solving and encouraging children to make their own choices, which requires reasoning about alternative options. Interventions that improve reasoning, but not other related non-verbal IQ measures, are not unprecedented (Mackey et al., 2011). Alternatively, the effect of the intervention on reasoning may simply be regression to the mean, since children in the intervention group had slightly lower baseline fluid reasoning than children in the control group (although not significantly so), and after the intervention both groups had matrix reasoning scores in between the two groups' initial scores. Future studies with larger samples and more diverse reasoning measures are needed to explore whether PCMC-A specifically improves fluid reasoning and the mechanisms underlying this improvement.

Given the significant positive effects of the intervention on receptive language in the original study, the present small negative effect was surprising. Although there were no significant decrements in vocabulary scores, the group differences suggest that children in the intervention did not exhibit typical age-related growth in vocabulary, perhaps in service of other cognitive skills. One potential reason for this difference was that the current study assessed only receptive vocabulary, while the original study assessed children's ability to interpret, recall, and execute verbal directions. Because vocabulary is only one component of language knowledge, it is possible that other receptive language measures, such as sentence comprehension or verbal memory, would show neutral or positive effects. Furthermore, the intervention focuses a great deal on parent-child communication and turn-taking and suggested that parents aim to more closely approximate the child's utterance length by "locking in" to the child's level of language use (Neville et al., 2013, supplement page 8). Indeed, the original study reported an increase in parent-child turn-taking during free play, but a decrease in parental mean length of utterance. Other studies have reported that *longer* adult mean length of utterance is correlated with children's vocabulary growth (Hoff, 2003). Thus, it is possible that the parent language strategies supported other components of children's language skills over vocabulary. However, we also encourage caution in interpreting this result too strongly; given the short interval (mean of 12 weeks) between test sessions, it is possible that the family communication strategies may result in positive effects over a longer period of time. Future studies should obtain more comprehensive measures of children's receptive and expressive language skills and track potential long-term changes.

There were a number of differences between the current PCMC-A intervention and the original intervention reported by Neville et al. (2013) that could account for the divergent results. First, our sample included children on average a year older than the original study (4–6-year-olds instead of 3–5-year-olds) because public elementary programs in the implementation region begin with universal pre-Kindergarten at age 4. Although age did not moderate the intervention effects, other research has shown that early interventions benefit younger children more than older children (Gardner et al., 2010; Loeb et al., 2007). However, the PCMC-A intervention was specifically created to target 3–5-year-olds, and thus may not have been as appropriate or effective for children over the age of 5.

Additionally, the present study included a more racially, ethnically, and socioeconomically diverse group of children from an urban environment, which may better reflect the population that this intervention would reach if it were implemented at scale. Furthermore, our sample included children who would have been excluded in the Neville et al. (2013) study such as those with a history of developmental disorders, children on medication, and children enrolled in individualized education plans, although explicit data on the prevalence of these variables (other than IEPs) was not collected. Our findings suggest that benefits of PCMC-A may not generalize to broader, more racially, ethnically, and neurocognitively diverse samples of children. However, it is difficult to disentangle which, if any, of the sample differences may have had the greatest effect on the differing findings.

Furthermore, the present intervention included both English and Spanish-speaking families, whereas Neville et al. (2013) report results for only English-speaking families. Although the intervention was rigorously adapted for Spanish-speaking families, preliminary findings from the intervention creators suggest that the Spanish-language adaptation of PCMC-A may have limited effects on nonverbal IQ measures and little to no effects on children's language or electrophysiological attention skills (Stevens et al., 2017). Thus, the language of intervention delivery and/or the cultural demographics of enrolled participants may impact overall intervention efficacy, such that the intervention may be uniquely beneficial to different cultural groups or populations. Furthermore, for both English and Spanish speaking families, our cultural adaptation of PCMC-A may have missed the mark. Although minor adjustments were made to the PCMC-A curriculum from original implementation in Oregon to the present implementation in Massachusetts (see methods), we did not run focus groups in Boston to make sure the broader goals of the intervention aligned with the goals of the community. For example, one participant was resistant to a lesson encouraging her to let her child make her own choices. This participant valued her child following rules set by adults so as to avoid getting into trouble. Considering how important it is for participants to ideologically align and "buy in" to the intervention (Hall et al., 2016; Smith et al., 2011), we suggest that future adaptations employ a rigorous model for geographical/cultural adaptation that involves a systematic, iterative process in close collaboration with the communities to be served. This in general signifies how hard it is to create "out of the box"/"one size fits all" parenting interventions. We also note that testing bilingual children in English (the language of classroom instruction) may have limited our abilities to detect meaningful intervention effects and underestimated these children's skills. Thus, assessments should also be approached and adapted through a culturally and linguistically sensitive lens.

The intensive nature of PCMC-A also made both recruitment and regular attendance an issue. Our school partners were committed to this research and allowed us to contact all families in their school (584 families total). However, only 18% of families enrolled in the study (note that enrollment rate is not reported in the original Neville et al., 2013 study, limiting our ability to compare rates for external validity). The difference in school structure between the original study (Head Start) and the present study (charter schools) may have contributed to potential differences in family participation and attrition. A major part of the Head Start mission is to establish family-wide supports through an already existing infrastructure, which may have led to increased family buy-in. Even though the participating charter schools strived to integrate families into their work, and study staff and recruitment partners (see Methods) were highly skilled in cultivating personal relationships with prospective families, these efforts might not have been enough to sufficiently reach all eligible families and overcome other barriers to participation. Although recruitment efforts included take-home flyers, direct phone calls, in-person efforts at parent-teacher nights, and word of mouth from enrolled parents, it is possible that some families were still unable to be reached. Furthermore, enrolling in this intervention required a substantial time and travel commitment from parents, and while there are many reasons families may not have attended the intervention, several families reported limited free time and scheduling difficulties as barriers. Notably, many of the enrolled participants attended extend school day programs, thus limiting available time for evening sessions, if weekends were not an option. Although we tried to reduce barriers to access by offering the intervention in English and Spanish at multiple locations and times, and by providing food, childcare, travel reimbursement, and monetary incentives, it was not always enough. The same barriers may have driven noncompliance amongst families assigned to the intervention. Only 36 of the 54 participants assigned to the intervention attended any sessions (67%), and only 32 completed all sessions (59%). Although enrollment issues and noncompliance are not rare for studies of this scope with lower SES participants, the rate of enrollment and nonattendance suggests important considerations for scaling. Future work is needed to establish mutually beneficial, translational research partnerships with lower-income, at-risk families to increase participation and retention (e.g., O’Neil et al., 2019).

Additionally, issues of sample selection and randomization may have dampened the demonstrated efficacy of the intervention. First, while all families at participating schools were invited to attend, the families who opted into the study may not have been the ones in the most need of the intervention curriculum. These families self-selected into the public charter school lottery, and also had the time and resources to attend weekly sessions, and thus may have already been highly motivated to help their children’s cognitive development. Supporting this possibility, the mean score on the standardized IQ and verbal measures were already within the average range before the intervention, such that these students may not have had much room for improvement (Diamond & Lee, 2011). Furthermore, although participants were informed that they would be randomly assigned to either an intervention or control group, most families wanted to receive the intervention. Because families in both groups attended the same schools and many knew each other well, it is possible there was some bleed-over of the intervention curriculum to the control group. Indeed, at least one family in the control group reported learning some of the strategies from a friend in the intervention group. Although randomization by school presents additional challenges, future studies are needed to determine whether cleaner groups may affect between-group differences.

One final difference between the current study and the original study was the choice of standardized assessments. The original study assessed nonverbal IQ with the fluid reasoning, quantitative reasoning, and working memory subtests of the Stanford-Binet 5th Edition nonverbal IQ scale (Roid, 2003), and assessed receptive language with the Sentence Structure and Concepts & Following Directions subtests from the receptive language section of the Clinical Evaluation of Language Fundamentals-Preschool 2nd Edition (Semel et al., 2004). We assessed similar overarching cognitive constructions, but with different standardized assessments (see Methods) and added measures to assess executive functioning given the executive function-related neurophysiological changes reported in the original study. Thus, there is a possibility that effects were unique toward these specific assessments, but we find this alternative unlikely. Furthermore, the original study included parent and teacher self-report of children's social skills and problem behaviors, reports of parenting stress, and electrophysiological measures of child brain functioning. Because the present study aimed to minimize the time needs of parents and teachers, we administered all assessments during school hours, and thus were unable to measure any potential positive impact of the intervention on parents or parent-reported child behaviors.

Although positive effects of the intervention were limited after three months, it is possible that benefits may show up later. So-called "sleeper effects" were observed in the Perry Preschool program, another study that included parent support. Immediate benefits of the program dissipated but became apparent years later in terms of educational attainment, earnings, and social outcomes (Nores et al., 2005). Thus, it may be important to follow up with the children in this intervention to see if beneficial effects are evident later in life.

In summary, although this study did not find broad evidence for beneficial effects of PCMC-A in the present sample, there is likely still promise of scalable family-based training programs for helping children's positive development. Furthermore, the PCMC-A intervention is not an outlier, or a spurious success—other family-based training program success cases include the Chicago Parent-Center Education program (Reynolds & Ou, 2004), the Reach Up home visiting program in Jamaica (Walker et al., 2000, 2005), the Multidimensional Treatment Foster Care for Preschoolers intervention (Fisher et al., 2007; Fisher & Stoolmiller, 2008), as well as many others. However, many of these approaches are multi-faceted, combining cognitive, nutritional, and emotional resources. While we agree that interventions to support child development require a holistic approach, scientifically this makes it hard to untangle which factors are causal and necessary. Thus, moving forward it will be important to run more targeted and tailored interventions to understand which factors enable success, and for which groups of participants. Furthermore, it is critical for future studies to better identify and ameliorate barriers to family participation and ensure appropriate cultural adaptation for all participants, and qualitative analyses of participants' experiences may reveal critical avenues for such advancements. Although difficult to find, the right balance between intervention tailoring and scalability may provide optimal outcomes for the greatest number of children and families.

## Notes

1. Note that Hearts & Flowers incongruent measures at time 2 remained skewed even though we attempted several transformations. To assess if our EF composite was robust to our standardization, we ran several sensitivity checks for each component of the EF composite. These modifications to our measure did not change the results reported.
2. To assess the robustness of this approach, we change the specification to use the difference between the post- and pre-assessment as the dependent variable instead of using it as an explanatory variable. The specification yields similar results with generally larger standard error and does not change any of the conclusions we can draw.
3. As a specification check, we also clustered the standard errors by school and whether the students came to MIT. Clustering at the lower level allowed for more clusters (16), but varying cluster size and a small number of observations in some clusters (<6). The specification increased the standard errors, but did not change the statistical significance of our results for any outcome.

## Acknowledgements

We thank 1647 Families and Transforming Ed for assisting in recruitment and intervention implementation; the University of Oregon Brain Development Lab (including Eric Pakulak, Zayra Longoria, Jimena Santillan, and Helen Neville) for providing the curriculum and training the interventionists to fidelity; Glennys Sanchez, Marta Rivera, Samilla Quiroa, Barney Arnold, Violete Vicario, and Blanca Valentin for delivering the intervention; and Kelly Halverson and Lauren Pesta for testing assistance.

## Declaration of interest statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by a grant from the Walton Family Foundation (Boston Charter Research Collaborative).

## References

Acevedo-Polakovich, I. D., Kassab, V. A., Boress, K. S. J., Clements, K. V., Stout, S., Alfaro, M., Brookins-Fisher, J., & Guerra-Morales, A. (2017). When context matters: Adaptation for high-risk U.S. Latina/o subgroups. *Journal of Latina/o Psychology*, *5*(4), 306–322.

Andrew, A., Attanasio, O., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., & Rubio-Codina, M. (2018). Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in Colombia. *PLoS Medicine*, *15*(4), e1002556.

Athey, S., & Imbens, G. W. (2017). The econometrics of randomized experiments. In A. V. Banerjee & E. Duflo (Eds.), *Handbook of economic field experiments* (Vol. 1, pp. 73–140). North-Holland.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*(1), 371–399.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*(5855), 1387–1388.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959–964.

Duncan, G. J., & Brooks-Gunn, J. (1997). *Consequences of growing up poor*. Russell Sage Foundation.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446.

Duncan, G. J., & Murnane, R. J. (2011). *Whither opportunity? Rising inequality, schools, and children's life chances*. Russell Sage Foundation.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (4th ed.). Pearson.

Fisher, P. A., & Stoolmiller, M. (2008). Intervention effects on foster parent stress: Associations with child cortisol levels. *Development and Psychopathology*, *20*(3), 1003–1021.

Fisher, P. A., Stoolmiller, M., Gunnar, M. R., & Burraston, B. O. (2007). Effects of a therapeutic intervention for foster preschoolers on diurnal cortisol activity. *Psychoneuroendocrinology*, *32*(8–10), 892–905.

Garcia, E., & Weiss, E. (2017). Education inequities at the school starting gate: Gaps, trends, and strategies to address them. *Economic Policy Institute*, *60,* 1–101.

Gardner, F., Hutchings, J., Bywater, T., & Whitaker, C. (2010). Who benefits and how does it work? Moderators and mediators of outcome in an effectiveness trial of a parenting intervention. *Journal of Clinical Child & Adolescent Psychology*, *39*(4), 568–580.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S. M., & Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science*, *344*(6187), 998–1001.

Hall, G. C. N., Ibaraki, A. Y., Huang, E. R., Marti, C. N., & Stice, E. (2016). A meta-analysis of cultural adaptations of psychological interventions. *Behavior Therapy*, *47*(6), 993–1014.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Economic Policy Institute.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*(1), 52–66.

Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Developmental Science*, *14*(3), 582–590.

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*(4), 947–959.

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in Psychology*, *5*, 599.

Mejia, A., Leijten, P., Lachman, J. M., & Parra-Cardona, J. R. (2017). Different strokes for different folks? Contrasting approaches to cultural adaptation of parenting interventions. *Prevention Science*, *18*(6), 630–639.

Neville, H., Pakulak, E., & Stevens, C. (2015). Family-based training to improve cognitive outcomes for children from lower socioeconomic status backgrounds: Emerging themes and challenges. *Current Opinion in Behavioral Sciences*, *4*, 166–170.

Neville, H. J., Stevens, C., Pakulak, E., Bell, T. A., Fanning, J., Klein, S., & Isbell, E. (2013). Family-based training program improves brain function, cognition, and behavior in lower socioeconomic status preschoolers. *Proceedings of the National Academy of Sciences*, *110*(29), 12138–12143.

Nores, M., Belfield, C. R., Barnett, W. S., & Schweinhart, L. (2005). Updating the economic impacts of the High/Scope Perry Preschool Program. *Educational Evaluation and Policy Analysis*, *27*(3), 245–261.

Olds, D. L. (20 06). The nurse–family partnership: An evidence-based preventive intervention. *Infant Mental Health Journal*, *27*(1), 5–25.

O'Neil, L. V., Pakulak, E., Stevens, C., Bell, T. A., Fanning, J. L., Gaston, M., Gomsrud, M., Hampton Wray, A., Holmes, K. B., Klein, S., Longoria, Z., Reynolds, M. M., Snell, K., Soto, A., & Neville, H. (2019). Creating connections between researchers and educators. *Journal of Cognition and Development, 20*(2), 110–133.

Pakulak, E., Hampton Wray, A., Longoria, Z., Garcia Isaza, A., Stevens, C., Bell, T., Burlingame, S., Klein, S., Berlinski, S., Attanasio, O., & Neville, H. (2017). Cultural adaptation of a neurobiologically informed intervention in local and international contexts. In K. R. Pugh, P. McCardle, & A. Stutzman (Eds.), *New directions in peer nomination methodology. New directions for child and adolescent development* (Vol. 158, pp. 81–92). Wiley Periodicals, Inc.

Parra-Cardona, J. R., Bybee, D., Sullivan, C. M., Domenech Rodríguez, M. M., Dates, B., Tams, L., & Bernal, G. (2017). Examining the impact of differential cultural adaptation with Latina/o immigrants exposed to adapted parent training interventions. *Journal of Consulting and Clinical Psychology*, *85*(1), 58–71.

Ramey, C. T., & Ramey, S. L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly*, *50*(4), 471–491.

Reynolds, A. J., & Ou, S.-R. (2004). Alterable predictors of child well-being in the Chicago longitudinal study. *Children and Youth Services Review*, *26*(1), 1–14.

Roid, G. H. (2003). *Stanford binet intelligence scales* (5th ed.). Riverside Pub.

Semel, E. M., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals* (4th ed.). Pearson.

Shonkoff, J. P., & Fisher, P. A. (2013). Rethinking evidence-based practice and two-generation programs to create the future of early childhood policy. *Development and Psychopathology*, *25*(4pt2), 1635–1653.

Smith, T., Rodríguez, M., & Bernal, G. (2011). Culture. *Faculty Publications*. https://scholarsarchive.byu.edu/facpub/2023

Stevens, C., Pakulak, E., Bell, T., Flanning, J., Klein, S., Isbell, E., Longoria, Z., Hampton Wray, A., & Neville, H. (2017). *Expansion and longitudinal assessment of a two-generation intervention for lower SES children: Challenges and implications*. Presentation at the 2017 Society for Research in Child Development Biennial Meeting, Austin, TX.

Ursache, A., & Noble, K. G. (2016). Neurocognitive development in socioeconomic context: Multiple mechanisms and implications for measuring socioeconomic status. *Psychophysiology*, *53*(1), 71–82.

U.S. Census Bureau. (2016). *The supplemental poverty measure: 2016, current population reports*. https://www.census.gov/content/dam/Census/library/publications/2017/demo/p60-261.pdf

von Suchodoletz, A., Gestsdottir, S., Wanless, S. B., McClelland, M. M., Birgisdottir, F., Gunzenhauser, C., & Ragnarsdottir, H. (2013). Behavioral self-regulation and relations to emergent academic skills among children in Germany and Iceland. *Early Childhood Research Quarterly*, *28*(1), 62–73.

Walker, S. P., Chang, S. M., Powell, C. A., & Grantham-McGregor, S. M. (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: Prospective cohort study. *The Lancet*, *366*(9499), 1804–1807.

Walker, S. P., Chang, S. M., Vera-Hernández, M., & Grantham-McGregor, S. (2011). Early childhood stimulation benefits adult competence and reduces violent behavior. *Pediatrics*, *127*(5), 849–857.

Walker, S. P., Grantham-McGregor, S. M., Powell, C. A., & Chang, S. M. (2000). Effects of growth restriction in early childhood on growth, IQ, and cognition at age 11 to 12 years and the benefits of nutritional supplementation and psychosocial stimulation. *The Journal of Pediatrics*, *137*(1), 36–41.

Webster-Stratton, C., & Herman, K. C. (2010). Disseminating Incredible Years Series early-intervention programs: Integrating and sustaining services between school and home. *Psychology in the Schools*, *47*(1), 36–54.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition*.

White, K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461–481.