2022

"What Works for Whom" in Family Support Programs: Leveraging Administrative Data to Improve Precision Matching

Elly M. Miles
University of Denver

"What Works for Whom" in Family Support Programs: Leveraging Administrative Data to Improve Precision Matching

Abstract

Family support programs (FSPs) are designed to stabilize and strengthen families on a range of outcomes to promote well-being. Paradoxically, families with the greatest need are more likely to drop out or experience reduced benefit on average. This study examines patterns of differential benefit for families experiencing elevated stress through a cross-program evaluation of “what works for whom” in FSPs.

Family Resource Center Family Development Services (FDSs), Colorado Community Response (CCR), Promoting Safe and Stable Families (PSSF), Head Start (HS), and SafeCare Case Management Pilot (Safe CMP) contributed data on 15,771 participants enrolled in services from 2014-2020. Program samples were weighted to correct for dropout. Improvements in self-sufficiency, health, and family protective factors were measured using ANCOVA. Main effects and interactions between program and predictors were examined, including cumulative stress score, single-parenthood, poverty, household size, fatherhood, primary language, parent age, race, and ethnicity. Importantly, the study design is pre-post assessments without randomization to program or reference to a comparison group. Thus, pre-post change could not be calibrated against change that would have occurred without program participation.

Dropout ranged from a low of 20% in HS to a high of 75% in PSSF. Latinx caregivers and families with higher stress varied in their pattern of attrition by program. Families enrolled in HS, a center-based model, and CCR, a community-based model, showed the greatest number of differential pre-post improvements. ESL families exhibited greater improvements in their health and child education across program participation. BIPOC families showed greater gains in child education and fewer improvements in family functioning and resilience than white families across program participation. Families with higher stress exhibited greater improvements in maintaining control when disciplining, but fewer pre-post gains in substance abuse, family functioning, child education, and beliefs about the intent of child misbehavior. Homevisiting and community-based programs were linked with the greatest number of differential subgroup benefits. Families with higher stress had the greatest variation in differential benefits by program. This study suggests that no single approach benefited all subgroups equivalently. Findings from this project may increase the equity with which families are supported in FSPs.

Document Type

Dissertation

Degree Name

Ph.D.

Department

Psychology

First Advisor

Sarah E. Watamura

Second Advisor

Julia Dmitrieva

Third Advisor

Angela Narayan

Keywords

Families experiencing high stress, Family support programs, Implementation of evidence-based interventions, Maltreatment prevention, Prevention interventions

Subject Categories

Developmental Psychology | Psychology

"What Works for Whom" in family support programs: Leveraging administrative data to improve precision matching

---

A Dissertation

Presented to

the Faculty of the College of Arts, Humanities and Social Sciences

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Elly M. Miles

August 2022

Advisor: Dr. Sarah E. Watamura

Author: Elly M. Miles
Title: "What Works for Whom" in family support programs: Leveraging administrative data to improve precision matching
Advisor: Dr. Sarah E. Watamura
Degree Date: August 2022

Abstract

Family support programs (FSPs) are designed to stabilize and strengthen families on a range of outcomes to promote well-being. Paradoxically, families with the greatest need are more likely to drop out or experience reduced benefit on average. This study examines patterns of differential benefit for families experiencing elevated stress through a cross-program evaluation of "what works for whom" in FSPs.

Family Resource Center Family Development Services (FDSs), Colorado Community Response (CCR), Promoting Safe and Stable Families (PSSF), Head Start (HS), and SafeCare Case Management Pilot (Safe CMP) contributed data on 15,771 participants enrolled in services from 2014-2020. Program samples were weighted to correct for dropout. Improvements in self-sufficiency, health, and family protective factors were measured using ANCOVA. Main effects and interactions between program and predictors were examined, including cumulative stress score, single-parenthood, poverty, household size, fatherhood, primary language, parent age, race, and ethnicity. Importantly, the study design is pre-post assessments without randomization to program or reference to a comparison group. Thus, pre-post change could not be calibrated against change that would have occurred without program participation.

Dropout ranged from a low of 20% in HS to a high of 75% in PSSF. Latinx caregivers and families with higher stress varied in their pattern of attrition by program.

Families enrolled in HS, a center-based model, and CCR, a community-based model, showed the greatest number of differential pre-post improvements. ESL families exhibited greater improvements in their health and child education across program participation. BIPOC families showed greater gains in child education and fewer improvements in family functioning and resilience than white families across program participation. Families with higher stress exhibited greater improvements in maintaining control when disciplining, but fewer pre-post gains in substance abuse, family functioning, child education, and beliefs about the intent of child misbehavior. Home-visiting and community-based programs were linked with the greatest number of differential subgroup benefits. Families with higher stress had the greatest variation in differential benefits by program. This study suggests that no single approach benefited all subgroups equivalently. Findings from this project may increase the equity with which families are supported in FSPs.

Acknowledgements

I would like to thank Sarah Watamura, who has been an endlessly encouraging, creative, and brilliant mentor. Sarah—your support, your leadership style, the culture you fostered—were instrumental in my journey through this program as a single mother. You have changed how I think and approach the world around me; I am forever grateful.

I would like to thank Julia Dmitrieva, whose patience and kindness has impacted me and many others. Thank you for your many hours solving data challenges in this project. I would like to thank my committee who sharpened the questions and approach in this project. I would also like to thank the incredible partners at the Colorado Office of Early Childhood, the Family Resource Center Association, and the Head Start grantees who made this project possible and were involved from its conception. My deepest gratitude to Jacob Barney, Teri Haymond, Heather Craiglow, Gayle Perryman, Kendra Dunn, and many others. I am also grateful to the Administration for Children and Families for funding this project and their interest in supporting all families enrolled in Head Start.

I want to thank my partner, Andy. Thank you for believing in and encouraging me. I want to thank my mom, who flew internationally to help me keep up with dishes, laundry, and child care. Finally, my biggest thanks go to my children, Tobias and Amos. You began this journey with me as preschoolers and were my reason for starting in the first place. Thank you for opening my eyes to the magic in development. You have been my inspiration and motivation to keep going no matter what.

Table of Contents

List of Figures

List of Tables

Introduction

Family Support Programs (FSPs) are designed to enhance family self-sufficiency and well-being, improve parent-child relationships, and boost parent engagement in high quality child-focused early interventions (Head Start Early Childhood Learning & Knowledge Center [ECLKC], 2018). Nationally, FSPs are funded through a variety of federal, state, and local initiatives. Two of the largest federal funding sources for FSPs include Head Start (HS) and child maltreatment prevention funds (e.g. Social Services Block Grant; Community-Based Child Abuse Prevention), programs which respectively served nearly one million children in 2019 (Head Start Early Childhood Learning & Knowledge Center, 2019) and nearly 2 million children in 2020 (U.S. Department of Health & Human Services, Administration for Children and Families [DHSS], 2020). Family Resource Centers, despite no dedicated federal funding, operate at more than 3,000 sites nationally providing an array of family support programs and services (National Family Support Network, n.d.). Paradoxically, the families most in need of these supports are less likely to consistently engage in or benefit from them (ACF, 2002; Barnes, MacPherson, & Senior, 2005; Raikes et al., 2013; Reyno & McGrath, 2006; Roggman et al., 2008; Lundahl et al., 2006). Few evaluations to date have conducted integrative data analyses across FSPs to identify "what works for whom," limiting the precision with which vulnerable families are supported. Currently, it is unclear whether

families experiencing elevated stress and adversity uniformly engage in and benefit from FSPs or if certain programs more effectively engage and benefit subgroups experiencing greater stress or disadvantage.

Family Support Programs

Historically, FSPs have been broadly defined as

> "Community-based services to promote the well-being of children and families, designed to increase the strength and stability of families...to increase parents' confidence and competence in their parenting abilities, to afford children a stable and supportive family environment, and otherwise enhance child development" (Public Law 103-66; GAO, 1996).

By design, FSPs attempt to act on the pathways by which poverty impacts child development, by intervening in the family's self-sufficiency, health and nutrition, the home environment, parent-child interactions, parent mental health, child access to high-quality early care and education, and neighborhood impacts (Brooks-Gunn & Duncan, 1997; Chase-Lansdale & Brooks-Gunn, 2014). FSPs have heterogeneous designs, but share the goal of boosting family well-being. Some FSPs emphasize family goal-setting, some connect child-centered programs with support for adults, some are adult-centered with additional support for children, some equally integrate adult and child-focused supports, and others function as umbrella programs that support and connect families to stand-alone interventions for both adults and children (Chase-Lansdale & Brooks-Gunn, 2014; Sama-Miller & Baumgartner, 2017). Across FSPs, a core activity includes linking families to an array of needed community resources to support their needs (Azzi-Lessing, 2011) and around half of FSPs provide case management to families (Sama-Miller & Baumgartner, 2017). Programs range from a primary focus on increasing parenting

capacity to those which have other primary foci (reviewed in Layzer et al., 2001). These programs might emphasize, for example, adult job training, self-sufficiency skill building, or housing and food security. Additionally, some programs focus on meeting economic and material needs to stabilize families and prevent later maltreatment (Simon et al., 2021). As two-generation programs, FSPs are simultaneously concerned with building caregiver capacity and in supporting optimal child development. In turn, the services offered to achieve family support goals may mirror those offered to families where positive parenting capacity is the primary focus (Layzer et al., 2001).

FSPs are delivered across a range of service delivery systems, including home-visiting and community-based, center-based, and mixed-methods, serving families in both home and center-based contexts (Office of Planning Research and Evaluation [OPRE] et al., 2006; Layzer et al., 2001; Whittaker & Cowley, 2012). The avenues by which families are targeted and enrolled into these programs vary. Some programs are universally available; in others, families are eligible based on demographic risk, such poverty or low maternal educational attainment (reviewed in Dodge et al., 2021). Other programs target families with elevated stress or clinical risks, including screened-out maltreatment referrals (Simon et al., 2021), parenting difficulties, unsafe home environments, or parent mental health or substance abuse concerns (reviewed in Dodge et al., 2021). Across diverse program designs, FSPs commonly focus on strengthening family well-being and enhancing self-sufficiency.

Program potential

Families experiencing elevated stress and adversity are broadly targeted by preventative family support and early intervention programs due to their potential to leverage cost-effective and long-term gains. As one key example, Early Head Start (EHS) and Head Start (HS), the largest federally funded intervention program for children in poverty, provides high-quality social, emotional, and educational child care and comprehensive family-focused services (ECLKC, 2018). Program evaluations reveal that, in the long-term, children enrolled in HS were more likely to graduate from high school, attend college, receive post-secondary degrees, and exhibit positive parenting and other social-emotional strengths in adulthood (Bauer & Whitmore Schanzenbach, 2016; Bauer et al., 2016). Infants and toddlers in EHS obtain short-term gains in social, cognitive, and language outcomes and their parents improve in parenting behavior and self-sufficiency (Love et al., 2005). Similarly, diverse parent-focused supports, delivered via parenting education programs or home-visiting models also exhibit promising improvements for families experiencing elevated stress and disadvantage. Meta-analyses reveal that these programs can positively impact parenting and reduce future maltreatment risk (Azzi-Lessing, 2011; Chen & Chan, 2016; Michalopoulos et al., 2019; Sweet & Appelbaum, 2004). Community response programs, such as Colorado Community Response and SafeCare, designed to stabilize families referred for and screened out of child protective services, have demonstrated some success with preventing later maltreatment, potentially by supporting the concrete needs of families (discussed in Simon et al., 2021). Similarly, Family Resource Centers (FRCs) act to support families with the complexities posed by

poverty (Lightburn & Warren-Adamson, 2015), while connecting families to additional parent and child-focused interventions. In turn, these services have been linked with increased family well-being across a range of family development outcomes (OMNI & FRCA, 2018).

Reduced benefit

Despite the promise of FSPs, meta-analyses to date reveal small effect sizes (e.g. Layzer et al., 2001; Sweet & Appelbaum, 2004; Chen & Chan, 2016; Gubbels et al., 2021), suggesting that some families benefit more than others across programs, resulting in a reduced average impact (Supplee & Duggan, 2019). Small effect sizes may be linked to the use of universal, or indiscriminate eligibility, into programs. Some FSPs utilize universal designs, which are attractive due in part to their potential to reduce parent stigma (e.g. the perception that bad parents participate in these programs; Leslie et al., 2016), an identified barrier to participation (Axford et al., 2012). Yet, universal programs have been identified as exhibiting smaller effects on parent and child outcomes than programs targeting specific subpopulations (e.g. teenage parents with young children (Layzer et al., 2001). In a recent example, a MIHOPE and MIHOPE-strong evaluation found no effects on birth outcomes in their cohorts that received home-visiting (Michalopoulous et al., 2019); a finding attributed to failing to exclusively target mothers with high-risk health behaviors or limited prenatal care. However, even in the context of more targeted support, families with elevated stress or risk do not on average experience strong improvements. For example, one meta-analytic estimate finds that after accounting

for publication bias, FSPs targeting maltreatment prevention in a sample identified as "at-risk" on average failed to achieve their intended outcome (Euser et al., 2015).

The families served by HS/EHS, FRCs, and other family-serving agencies often experience multiple risk factors, such as family conflict and domestic violence, parent-child relationship problems, poor caregiver mental health, caregiver substance use, housing insecurity, low educational attainment, low social support, unemployment and financial strain, and earlier childbirth (Sidebotham & Heron, 2009; Stith et al., 2009; U.S. Department of Health & Human Services [DHHS] et al., 2019). While families with these characteristics are the target of these interventions, these risk factors may also result in families experiencing fewer intervention benefits. These families may be at greater risk of dropping out of programming prior to completion, reducing access to effective treatment dosage. Indeed, greater disadvantage is widely linked with dropout across parenting-focused programs (Lundahl et al., 2006; Reyno & McGrath, 2006). This observation is paradoxical, given that most FSPs are designed to support families with elevated stress and risk.

Even when families with higher levels of adversity maintain enrollment in programs, they may experience diminished program effects on average. For example, multiple elevated risk factors such as poverty, single parent status, receipt of public assistance, unemployment, or teen parenting, predicted few positive outcomes in EHS (Raikes et al., 2013) and several significant negative effects on parent-child interactions (ACF, 2002; Miller et al., 2016). It may be that the needs of families at the greatest levels of risk are not sufficiently addressed by the services provided, indicating additional needed support

(ACF, 2002). Distressed families with higher levels of need may experience greater barriers to participation in programs. Some work finds a direct relationship between rising cumulative stress and decreased program participation (Mendez et al., 2009). Barriers to participation may be structural, such as limited child care, work schedule conflicts, and transportation challenges (OPRE, 2006; Mendez et al., 2009). Alternately, barriers may be connected to parent wellbeing, such as higher fatigue and health concerns (Mendez et al., 2009). Increasingly, prevention research places emphasis on accurately targeting and tailoring FSPs to the needs of families to bolster engagement and outcomes (August & Gerwitz, 2019). Indeed, as a field, recommendations have been made to shift away from indiscriminate "one size fits all" model designs and increasingly toward models which target subgroups with precision (Supplee & Duggan, 2019).

Program Engagement. By some estimates, close to 50% of parents who are referred to parent support or, relatedly, child behavioral management programs either fail to enroll (Baker, Arnold, & Meagher, 2011; Chacko et al., 2016) or drop out after initially engaging (Chacko et al., 2016; Friars & Mellor, 2009). Engagement in FSPs is a multifaceted construct, with interest or agreement in enrolling, attendance, involvement or motivation, and completion and application of "homework" all potentially unique and meaningful indicators of why families participate in and ultimately benefit from FSPs (Wagner et al., 2003). A variety of structural and demographic family characteristics have been identified as predictive of enrollment and engagement. Low-SES and single-parent families are more likely to seek out formal programs (Friars & Mellor, 2009; Schneider, Gerdes, Haack, & Lawton, 2013; Redmonth et al., 2002). However, families

experiencing these risk factors have repeatedly been identified as more likely to drop out (Baker et al., 2011; Chacko et al., 2016; Fernandez & Eyberg, 2009; Peters, Calam, & Harrington, 2005; Roggman et al., 2008; Schneider et al., 2013), highlighting the possibility that caregivers under stress seek out support, but may experience mismatch between their family's needs and program design or other provider factors.

Caregivers experiencing greater subjective stress, younger caregivers, mothers of first-born children, caregivers with multiple children, and caregivers with more economic hardship and lower education are also more likely to drop out (Axford et al., 2012; Friars & Mellor, 2007; National Committee to Prevent Child Abuse, 1995; Pellerin, Costa, Weems, & Dalton, 2010; Pote et al., 2019; Reyno & McGrath, 2006). Lower levels of education may present a barrier to engagement, as families may feel isolated or have low levels of literacy or verbal skills obstructing engagement (reviewed in Axford et al., 2012). Fathers have also been identified as participating less, with lower levels of enrollment and retention in parenting programs on average (reviewed in Placa & Corlyon, 2014; Pote et al., 2019), although either mother or father participation in an intervention may boost the other parents engagement (McKee et al., 2021).

Latinx ethnicity is inconsistently related to engagement, with some studies demonstrating greater participation and others less (reviewed in Axford et al., 2012). While one body of work has linked minority race and ethnicity to reduced program engagement (McGrath & Reyno, 2006), an assessment of retention by maternal ethnicity and race across several home-visiting FSPs identified that Latinx and Black mothers had higher rates of participation than their majority race and ethnicity counterparts (McCurdy,

Gannon, & Daro, 2004). Families who do not speak English as a first language also exhibit mixed profiles of engagement. Some work finds greater dropout when families do not speak the native language programs are offered in (Eisner & Meidert, 2011) and lower engagement in Spanish-speaking families enrolled in programs such as HS (McWayne et al., 2013; Zill et al., 2003). However, an EHS evaluation found that Spanish-speaking mothers were less likely to drop out (Roggman et al., 2008) and, among attendance categorizations of mothers enrolled in the Nurse Family Partnership nationally, ESL mothers were more likely to stay engaged in programming (Holland et al., 2018). The same evaluation of Nurse Family Partnership found that Black mothers were significantly more likely to drop out, although they also received services from more disorganized programs with higher levels of workforce turnover (Holland et al., 2018). Thus, ethnicity, race, and language appear to have wide variability in their moderation of program engagement, perhaps in part due to heterogeneity within any one demographic group, and in part due to factors that cooccur at greater or lesser rates within and across groups.

Evaluations which have examined family demographic risk cumulatively (e.g. young maternal age, single-parent, unemployment, low education, welfare receipt) indicate that families with greater levels of demographic risk and housing instability are more likely to drop out (Michalopoulos et al., 2019; Roggman et al., 2008). Families with the greatest levels of risk had lower average length of EHS enrollment, were less likely to remain continuously enrolled for a period of two years, and were less likely to be rated as engaged by providers (ACF, 2002). Other risk factors commonly associated with

dropping out of EHS included substance abuse, homelessness, and mental health needs (OPRE, 2006). These disparities in retention are important, as current evaluations of FSP outcomes may not be generalizable to vulnerable families in the context of underrepresentation, especially when the bias introduced by dropout is not controlled for (Whittaker & Cowler, 2012).

Program Outcomes. When families experiencing elevated risk do engage with FSPs, they do not on average experience equivalent levels of benefit. Younger parent age, younger child age, and single-parent status have been identified as undermining outcomes in maltreatment prevention or parent training interventions (ACF, 2002; Lundahl et al., 2006; Reyno & McGrath, 2006). ESL (McConnell, et al., 2013; Wagner & Clayton, 1999), unemployment, the presence of a disability or chronic health condition, the care of a child with a disability or chronic health condition (McConnell et al., 2013), and greater economic disadvantage (Lundahl et al., 2006) are all linked to diminished improvements. However, family income alone may be an insufficient indicator of who will benefit most in an intervention. Pelhamm and colleagues (2017) identified that the majority of families in their sample enrolled in a home-based parenting intervention, while low income, were not "high-risk" and did not benefit in the same way that families with greater levels of distress did.

Minority family ethnicity is known to influence attitudes toward parenting and parenting styles (e.g. Jones et al., 2010) and in some cases has been linked to differential treatment effects (Arnold et al., 2003; Begle et al., 2012). In one study of ADHD treatment modalities, minority race and ethnicity predicted less reduction of symptoms

following treatment (Arnold et al., 2003) and failure to significantly reduce internalizing symptoms (Barrera et al., 2002), suggesting diminished family treatment outcomes. In contrast, other work identifies equivalent treatment effects across ethnicities for parenting outcomes (Jones et al, 2010; Reid et al., 2001), greater response to behavioral treatment for ADHD than in Caucasian families (Arnold et al., 2003), and the treatment of child externalizing symptoms following parent-focused intervention (Berrara et al., 2002). Thus, some evidence points to diminished treatment outcomes in minority ethnicity families, while other work points to equivalence of outcomes. Despite these apparently contradicting data, parenting may differ by ethnicity before intervention (Jones et al., 2010) and programs which recognize and adapt approaches based on differing parenting strategies within and across ethnicities may increase minority families' engagement and benefit (e.g. Kumpfer et al., 2002).

Relatedly, ESL is sometimes negatively associated with outcomes in FSPs. For example, Spanish-speaking Latina mothers did not increase their parenting efficacy in a parenting intervention, while their English-speaking Latina counterparts did (Wagner & Clayton, 1999). Despite lower likelihood of dropout, ESL parents and parents with lower levels of education also did not apply or generalize intervention content in a community-based parent training program at equivalent levels with their counterparts (Eisner & Meidert, 2011). Qualitative work identifies that Spanish-speaking families have greater satisfaction when curricula are culturally adapted (So et al., 2020) and that Spanish-speaking households may have differing foci around child development and parental responsibilities from their English-speaking Latino counterparts. Spanish-speaking

parents may be more concerned with child motor development and completion of homework and less concerned with topics such as ensuring child safety or discipline-focused activities (McWayne et al., 2013). Programs which emphasize these priorities or otherwise emphasize adaptations for acculturating families may be linked with greater benefit for culturally diverse families. For example, cultural adaptation of a parenting intervention for Somali immigrants was linked with improved outcomes (Osman et al., 2019).

Single-parent status and younger maternal age are also linked with reduced treatment response in parenting interventions (Nix et al., 2009; Wagner & Clayton, 1999). These findings are also consistent with Early Head Start outcomes, where families experiencing cumulative risk including single-parent household status and younger parent age did not obtain benefits (ACF, 2002). Finally, meta-analyses find that younger maternal age and single-parent status are linked with reduced parenting treatment effects (Lundahl et al., 2006; Reyno & McGrath, 2006). Thus, mothers who are younger and in single-parent households may less consistently benefit from parenting-focused FSPs.

Family Cumulative Stress. Other highly stressful and destabilizing events in the family, while indicative of greater need, are also linked with reduced benefit. Caregiver mental and physical health problems, substance abuse, domestic violence, housing instability, and transportation have been discussed as barriers for effectiveness in supports aimed at parent self-sufficiency (Chase-Lansdale & Brooks-Gunn, 2014) and, in turn, important targets for subgroups with high levels of need. These experiences in the family are critical and potentially interrelated indicators of need, given that domestic

violence and substance abuse are connected with higher parenting stress (Rutherford & Mayes, 2019; Pinto et al., 2019) and, along with housing instability, are predictive of substantiated maltreatment (Marcal, 2018; Victor et al., 2018). To date, limited work examines subgroup patterns of engagement or benefits in FSPs for families experiencing these types of stressors. Despite a critical need to better serve families with these stressors, providers and home-visitors may be, in fact, less likely to discuss identified substance abuse and domestic violence with families than other concerns (Kanda, 2021). Moreover, providers are least likely to screen, refer, link, and follow-up on substance abuse risk (West et al., 2021). This is a significant problem, as families with these experiences are at greater risk of maltreatment than families typically targeted in FSPs on the basis of demographic risks (e.g. maternal age, education, primiparous birth; Dodge et al., 2021). Failure to engage and support populations with the greatest need reduces the public health benefit and return on investment that is possible in prevention programs (Dodge et al., 2021).

Differential benefit

Despite overall small effects in FSPs, some evaluations reveal the potential for differential benefit in families with the highest levels of need. For example, one state's implementation of Healthy Families America was linked with reduced depression, perceived stress, and harsh parenting techniques in families with the highest levels of risk (e.g. Green et al., 2014). A Nurse Family Partnership long-term follow-up revealed differential benefits in the daughters of mothers with higher-risk (single and in poverty; Eckenrode et al., 2010). Thus, when families with higher levels of risk are effectively

engaged and supported, the potential for improvement may be maximized (Olds et al., 2007; Layzer et al., 2001). This presents an interesting challenge, suggesting that optimized outcomes may require a more precise fit between family, program, and circumstance.

When evidence-based programs do not scale or work across communities, one possible explanation is poor fit between the needs of families and communities and program characteristics. Indeed, programs often become evidence-based with a restricted population and, when scaling, intervention effect sizes appear to reliably decrease (Supplee et al., 2021). Similarly, studies with smaller sample sizes appear to be more likely to find significant program effects (Euser et al., 2015) and, as sample sizes increase, effect sizes drop (Chen & Chan, 2016). In a seminal meta-analysis that identified mostly small effects across 260 FSPs, 73% of reviewed studies were single-site evaluations and only 2% of included studies were statewide multi-site programs (Layzer et al., 2001). This pattern of findings raises the possibility that outcomes which initially appear promising may vary substantially across populations and contexts and have been most effective when tightly tied to communities that effectively engaged diverse families. In turn, diminished benefits may be tied to mismatch between the needs of families, the contexts they exist in, and program features, including chosen delivery systems (Azzi-Lessing et al., 2011). To bolster outcomes in family-level prevention, new attention is needed on broad patterns of benefit, with attention on the question of "what works best for whom and in what contexts" (Supplee et al., 2021).

What works for whom

While a variety of studies examine group-level risk and benefit, attention on programmatic differences contributing to fit and improvements is also critical. However, the programmatic differences driving effects are challenging to capture, leading to a "black box" perception, especially in home and community-based models where interactions are less commonly observed (Hebler & Gerlach-Downie, 2002; Korfmacher et al., 2008). One meta-analysis found that no single program characteristic consistently impacted outcomes across 60 home-visiting programs (Sweet & Appelbaum, 2004). Information on how program content relates to measured outcomes is not always clear (Beatson et al., 2020) and primary areas of focus are not always aligned with areas of primary improvement (Sweet & Appelbaum, 2004). Families may also exhibit growth in areas they reported no motivation to address at program start (FRCA & OMNI, 2018). Further, while some work identifies that longer length of treatment can improve some outcomes (e.g. maternal substance abuse, Duggan et al., 2004), other work suggests that length of treatment or dosage is not meaningfully associated with outcomes (Gubbels et al., 2021), or, surprisingly, even linked with declining effect sizes (Sweet & Appelbaum, 2004).

A body of work documents that differences in service delivery structure may be meaningfully related to program engagement and benefits for subgroups experiencing disadvantage. Families with infants enrolled in center-based programs report challenges in accessing services tied to logistics and their health (Hackworth et al., 2018). Similarly, only Spanish-speaking families reported that reliance on public transportation,

carpooling, or walking to access the programs posed a barrier to participation in center-based programs (So et al., 2020). In contrast, in mixed delivery (home-visits supplementing center-based programs), logistical barriers did not impact retention; instead, high levels of family stress and family-difficulties were linked with lower retention (Hackworth et al., 2018). Dropout in home-visiting programs may additionally be influenced by parent discomfort with providers coming into the home (Hackworth et al., 2018).

Borrowing from the child maltreatment literature, caregivers at high-risk for future maltreatment (identified by having more children and being in greater poverty) had lower future incidence of maltreatment when engaged in programs which emphasized meeting basic needs and when services were delivered in center-based settings as opposed to home-visiting (Chaffin et al., 2001). Similarly, programs emphasizing the provision of concrete needs may be more likely to engage and benefit families at-risk of future maltreatment (Simon et al., 2021). Analyses of Colorado FRC programming further identified that parents with greater poverty and lower family functioning and resilience were less likely to engage in parenting-focused services and more likely to seek out services focused on improving basic needs or adult education (OMNI & FRCA, 2018). However, other work suggests that programs focused on concrete support are less effective in their ability to reduce later maltreatment (Gubbels et al., 2021).

Families with low levels of income are known to experience greater isolation and lower perceived belonging (Stewart et al., 2009), which may impact interactions with service delivery. For example, individuals with lower income may be more likely to

sense community within center-based programs with other families experiencing similar levels of income or marginalization (Stewart et al., 2009). Although not exclusively focused on parents, one analysis of primarily African American women with low income and education revealed that improvements occurred within the context of perceived community and belonging in a center-based program (Brodsky & Marx, 2001). These settings may allow for enriched forms of social support to develop, off-setting the impacts of poor neighborhood climate or other adverse home community impacts (Lightburn & Warren-Adamson, 2015). A variety of other analyses and reviews find that peer support, typically obtained in center-based programs, meaningfully fosters community and bolsters outcomes (Layzer et al., 2001; Whittaker & Cowley, 2012); therefore, social support may improve at a greater rate in center-based programs than home-visiting. Yet, these same findings reveal that limited financial resources may preclude access and participation in center-based interventions (Stewart et al., 2009). Moreover, interpersonal factors such as stigmatization and avoidance may lead to individuals with low-income self-isolating, potentially reducing participation in center-based interventions and auxiliary programming (Ammerman et al., 2006; Stewart et al., 2009). The degree to which site-based support services are viewed as locales for support and community versus challenging to access and sources of stigma or prejudice are currently unclear in family support work.

Home-visiting as a field is largely predicated on its capacity to solve barriers experienced by families with greater need. Qualitative work has identified that low-income parents may experience the additional commitment of attending parenting

education activities as stressful (Gross et al., 2001). In contrast, home-visiting may effectively reach populations with higher levels of need and stress, by reducing barriers connected to the time, location, and cost of accessing and sustaining involvement in center-based programs (Ammerman et al., 2006; Azzi-Lessing, 2011; Pote et al., 2019). In this way, home-visiting holds the promise to promote inclusion of families experiencing marginalization and under-representation (Pote et al., 2019) and may confer advantage when families have high levels of stress, as they may work to identify and reduce the compounded barriers that compromise caregiver ability to engage in center-based programs (Whittaker & Cowley, 2009). Yet, the ability of home-visiting services to sustain engagement or boost strong outcomes among families with high levels of stress has been similarly challenged.

Some work finds that there are no appreciable differences between home-visiting and group-based parenting education models for reduction in child maltreatment risk (Chen & Chan, 2016). Another national evaluation, while much older, found that child outcomes and improvements in parenting behavior were lower on average in family support programs that were delivered in home-visiting models and parent effects were much larger when peer support was received (Layzer et al., 2001), a feature of center-based programs. These findings may be mediated by provider discomfort with directly addressing parent-child interaction in home-visiting (Hebler & Gerlach-Downie, 2002). However, the families served in home-visiting programs also tended to have lower-income than those in center-based programs, suggesting a different level of need in the family (Layzer et al., 2001). Interestingly, when parent self-development was a stated

goal of the program, home-visiting had the strongest effects (Layzer et al., 2001), suggesting that parenting and parent-focused development were bolstered in opposite environments. Therefore, the contexts in which home-visiting may benefit families over center-based models is unclear. Additionally, across modalities, programs that paired case management with intervention activities saw greater improvements (Layzer et al., 2001). Therefore, the inclusion of programs with individualized case management or goal-setting as a component of the intervention may meaningfully impact effect sizes.

Other work suggests that programs with the greatest success utilize mixed methods, flexibly responding to the complex needs of families with high stress and interacting with as many ecological contexts the family exists in as possible (Moran & Van der Merwe, 2004). For example, EHS evaluations identify benefits when occasional home-visiting is tied to center-based services (OPRE, 2006). Parents in center-based programs reported logistical access barriers and showed lower levels of exposure to content, application, and generalization of skills when their health was low (Hackworth et al., 2018). In contrast, these barriers were not impactful in a mixed-delivery model with both center-based and home-based services, suggesting that the addition of home-based services effectively reduced complex barriers faced by families (Hackworth et al., 2018). Yet, lower exposure, application, and generalization of content was not eliminated in the mixed-delivery program; rather, younger parents and ESL parents appeared to apply less content (Hackworth et al., 2018). In summary, it is currently unclear whether families with greater stress experience more benefit overall in home-visiting, center-based, or mixed-delivery programs.

In addition to problems related to mismatch between family need and program design, other individual factors in the family may drive outcomes (McCurdy & Daro, 2001). Readiness to change, a concept investigated heavily through its applications in motivational interviewing (e.g. Miller & Rollnick, 1991), is one potentially meaningful predictor of both FSP engagement and outcomes. Based on Prochaska and DiClimente's (1984) theory of change, readiness to change is a component of the preparation phase in change, supporting subsequent engagement and maintenance stages. As reviewed in McCurdy and Daro's (2001) conceptual framework for FSP engagement, readiness to change is thought to be a meaningful individual determinant of the decision to engage in FSPs. Indeed, both treatment engagement and outcomes for parent and child treatment are boosted by higher readiness to change (Chaffin et al., 2009).

Theoretic framework

This project is informed broadly by Bonfenbrenner's ecological system's theory (1986), capturing that multiple systems and their interactions inform family and child outcomes. Within this model and as explicated in more detail in Conger's Family Stress Model (Conger et al., 2002), rising family stress can impair the quality of available caregiving. Impaired caregiving in the presence of high stress and low resources may lead to stress being experienced as toxic by children (Shonkoff, Boyce, & McEwen, 2009) with consequences for long-term outcomes. Targeted efforts to intervene in this cycle and promote family well-being using whole family two-generation approaches are at the crux of family support designs.

Caregivers represent their children's primary source of buffering and protection under conditions of stress, such as poverty (Shonkoff et al., 2009), and caregiver capacity to provide warm, responsive, safe, and stable home environments can be compromised by limited resources and high parenting stress (Huston & Bentley, 2010). In the absence of externally offered family support in resource-depleted and high stress contexts, children may experience stress to be toxic (Lupien et al., 2009), which is increasingly recognized to contribute to lifelong consequences, including poor health and diminished achievement (Shonkoff et al., 2012; McEwen & McEwen, 2017). Thus, efforts to effectively reduce disadvantage for children experiencing poverty or other early life adversities not only involves early high-quality intervention programs for children experiencing high stress, but also necessarily involves supporting caregivers by reducing experiences of stress, where possible, and promoting skills to facilitate positive and buffering parent-child relationships. FSPs accomplish these goals broadly by enhancing family well-being and promoting positive child outcomes by facilitating "family development" (ECLKC, 2018). The emphasis of particular programs may vary to include, for example, adult job training, self-sufficiency skill building, housing and food security, etc., but all are simultaneously concerned with building caregiver capacity to support optimal child development.

Bonfenbrenner's ecological systems theory (Bonfenbrenner & Morris, 2006) further informs the rationale for FSPs designs and also guides the proposed study framework. Such an ecological framework posits that parenting practices do not occur in isolation of the family's immediate and broader social context, but instead are multiply influenced and determined by transactional processes occurring at different levels in the parent's

environment (Belsky, 1984; Bonfenbrenner, 1986; Bonfenbrenner & Morris, 2006). Crises surrounding resources and structural disadvantages that parents face including their financial wellbeing, housing security, social support, access to mental and physical health care, educational attainment, and other factors broadly targeted by FSPs compromise the quality of available caregiving (Conger et al., 2002). Caregiver ability to sustain involvement and investment in high quality early childhood education may also be compromised (Conger & Donnellan, 2007). Indeed, as barriers are effectively reduced and protective factors strengthened in parents and families, the child's ecology shifts toward greater well-being and parents may increase in their capacity to support their child's involvement in high quality early education like EHS and HS and to provide their child with more sensitive, responsive, and appropriately structured caregiving. When the family's ecological system is most effectively targeted, family development outcomes are bolstered, as maladaptive parenting practices can be replaced and high parent stress, a determinant of maltreatment (Taylor et al., 2009), may be reduced. This theoretical vantage point supports comparisons across FSPs that may promote more precise matching of participants to programs.

The question of what works for whom is not a new one. In fact, papers discussing the importance of this question began emerging in 1967 when Gordon Paul addressed the question "which type of patient, meeting with which type of therapist, for which type of treatment will yield which outcome?" arguing that a more prescriptive approach was needed to bolster patient outcomes (Paul, 1967). Beginning in 1990, calls began for deeper understanding of "what works for whom under what circumstances" in Head Start

(US DHHS, 1990). Yet, progress today in isolating the prevention program that will serve the right family in the right way are still far off. Examination of "what works for whom" are often constrained to single program evaluations (e.g. Healthy Families America, Green et al., 2014; Early Head Start Evaluation, OPRE, 2006) and some cross-program evaluations under shared funding streams (e.g. MIHOPE and MIHOPE-Strong Start; Michalopoulos, 2019). Current efforts to advance the match between program and participant, driven through initiatives such as Precision Home-Visiting (PHV; Supplee & Duggan, 2019; Korfmacher, 2020), focus on the use of responsive research designs (e.g. Multiphase Optimization Strategy, Sequential Multiple Assignment Optimization Randomized Trials) in the intervention testing and development stage. However, such designs are predicated on starting hypotheses in a field which, to date, continues to experience limited consensus on where differential benefit might be expected. Some evaluations have begun to compare and contrast programs on a wider level (see Beatson et al., 2020; Gubbels et al., 2021; West et al., 2021). However, current cross-program evaluations rely on meta-analyses; while useful for establishing evidence of average effects, meta-analyses also pose limitations. For example, meta-analyses have a limited ability to account for subgroup differences, such as who experienced differential benefit or harm both across and within interventions (Brown et al., 2013). This project takes a meaningful step forward by conducting integrative administrative data analysis (Brown et al., 2013), seeking to confirm and explore subgroup differences across a diverse set of FSPs.

Research questions and hypotheses

Family demographic and structural characteristics are found to consistently predict differences in family support outcomes, making their explicit consideration in referral to FSPs necessary. However, the studies reviewed did not compare across diverse FSP designs to illuminate which programs worked *better* for demographically and structurally diverse subgroups; that is, where differential outcomes and systematic disparities in program effectiveness may exist. Identification of the subgroups achieving better or worse family development outcomes could improve the precision with which participants are linked to programs (McConnell, et. al., 2013; Supplee & Duggan, 2019) and critically indicate where program adaptations are needed to more effectively engage indicated subpopulations. This project approaches the highlighted gaps through three main questions:

1) How do participant characteristics predict engagement across an array of FSPs?
2) How do participant characteristics predict differences in family development outcomes across an array of FSPs? and
3) In which service delivery structures (e.g. home-visiting, center-based, or mixed-delivery) do participants experiencing elevated stress and disadvantage demonstrate the greatest pre-post improved outcomes?

Question 1 hypotheses

Family engagement when enrolled within FSPs is hypothesized to be negatively impacted by degree of family poverty, single-parent status, lower caregiver age, fatherhood, family minority race, and higher levels of cumulative stress (Reyno &

McGrath, 2006; Friars & Mellor, 2007; Pellerin, Costa, Weems, & Dalton, 2010; National Committee to Prevent Child Abuse, 1995; Pote et al., 2019; Wagner & Clayton, 1999). Greater attrition of these groups is expected to occur across FSPs and may exacerbate disparities in family development outcomes. ESL families are expected to be less likely to dropout and potential engagement differences in Latinx families are approached as exploratory, given prior conflicting findings. Any family characteristics associated with drop-out will be weighted in subsequent analyses to reduce the bias associated with underrepresentation and increase the accuracy of program effect estimates for target subgroups across FSPs.

Question 2 hypotheses

In addition to exploring which programs are linked with overall pre-post improvements, family demographic and structural characteristics are expected to moderate outcomes overall and across programs (Reyno & McGrath, 2006). Lower parent age, larger family size, single-parent family structure, fatherhood, and lower levels of education, higher poverty, and greater levels of cumulative stress are expected to decrease treatment gains on average across programs. However, based on prior work, families with greater cumulative stress may make greater improvements in economic self-sufficiency (OMNI & FRCA, 2018). Minority race and ethnicity families and those who speak English as a second language may experience less benefit following FSP participation, on average (Reyno & McGrath, 2006). However, given a mixed literature (e.g. Reid et al., 2001), this is approached as exploratory and unlikely to be a consistent finding across FSPs. Differences in ESL families' outcomes may relate to FSPs' degree

of cultural adaptation or emphasis on child and parenting priorities that overlap with those of Spanish-speaking families (e.g. McWayne et al., 2013), although these factors are unmeasured in this analysis.

Question 3 hypotheses

While subgroups experiencing stress and disadvantage are expected to display reduced pre-post improvements overall, differential patterns are also expected by program. For the most part, these differences are approached as exploratory, given the mixed literature reviewed. However, several subgroup differences are approached with confirmatory hypotheses (Bloom & Michalopoulos, 2013). First, families experiencing greater levels of stress, indicated by accumulation of limited income, domestic violence, substance abuse, housing instability, food insecurity, and concurrent or former child welfare involvement, are expected to demonstrate improved outcomes in home-based versus center-based service delivery systems. Because a variety of sources point to greater social isolation and barriers in these instances, it is expected that home-based delivery systems may more effectively support complex family needs in an environment with fewer barriers to accessing services.

When parents experience poverty, families in center-based programs are expected to demonstrate greater pre-post improvements on a select number of outcomes. In particular, improvements may be strongest for social support and parenting-focused outcomes in center-based programs, where peer support is possible (Brodsky & Marx, 2001; Chaffin et al., 2001; Whittaker & Cowley, 2012). Although exploratory, programs which utilize mixed-delivery systems, both at home and in center-based settings, may be

especially supportive of improvements on average, due to efforts to engage the family across ecological contexts. Other subgroup differences (e.g. race, ethnicity, family language) are approached as exploratory, given limited consensus in the literature to inform hypotheses. Finally, across structural and demographic factors, caregiver readiness to change is hypothesized to be a meaningful moderator of outcomes (McCurdy & Daro, 2001).

The Present Study

The present questions are examined within the state of Colorado, utilizing administrative records on a suite of Family Support Programs, spanning Head Start, Family Resource Center Family Development Services, SafeCare Case Management Pilot, Promoting Safe and Stable Families, and Colorado Community Response. Across Colorado, an estimated 48,000 children received community and family support in 2019 (Snyder, 2019), including approximately 8,152 in Head Start (Child Care Technical Assistance Network, n.d.), 2,989 families in Family Development Services (FDS) within FRCs (FRCA & OMNI, 2020), and 3,449 in Promoting Safe and Stable Families (DHHS, 2022). The programs examined in this study shared the features of family-driven goal setting and individualized comprehensive case management, although areas of emphasis, routes of referral, and eligibility had several differences. Head Start adds family comprehensive support onto child-focused center-based services, Safe CMP adds family comprehensive support onto parenting-focused home-visiting, FDS integrates family comprehensive support in a center-based program, CCR integrates family comprehensive support in a community-based program, and PSSF adds family comprehensive support

into a mixed array of services delivered in home and center-based programs. At the time of the study, referral to CCR was the most restrictive, as referrals primarily originated from screened out Child Protective Service (CPS) referrals (Allan et al., 2018). All other programs assessed included a range of self, provider, or CPS referrals. Eligibility in CCR was determined only by referral from CPS. Safe CMP and HS required children to be within a particular age range and in families experiencing risk criteria (Beachy-Quick et al., 2017). Eligibility for PSSF is determined locally, reducing insight into common eligibility criteria. In contrast to other programs, no eligibility criteria were established for FDS, suggesting a universal model.

The present study utilizes administrative data to integrate across FSPs supporting families experiencing elevated stress and disadvantage. As a major strength of this analysis, disparate programs in Colorado coordinated their use of shared assessment tools and contributed data to better understand where indicators of differential pre-post gains or losses existed for the families with the greatest levels of need. However, important limitations include that no random assignment or comparison group was identified, fidelity measurement and monitoring varied widely by program, and missing data posed unique challenges for integrated data analysis. Despite these inherent limitations, knowledge gained in the present study may improve the precision with which families experiencing elevated stress in Colorado, and even nationally, may be engaged and supported. The findings in this analysis may enhance the precision with which subgroups may be matched to programs or with which programs may identify and differentially

support families experiencing fewer benefits and provide insight to support subsequent rapid and responsive research designs.

Methods

The characteristics of each FSP were examined, followed by review of the participants available for integrated data analysis. Next, measures are reviewed, including demographic predictors and outcome measures. Finally, the data analytic plan describes the process of pooling, preparing, and analyzing data for analysis.

Family Support Programs

Five family support programs (FSPs) contributed data for analysis. Each program is briefly described, including their reported practices to measure and promote high-fidelity services. *Table 1* compares programs' core features.

Table 1. Comparison and Contrast of Included Family Support Programs

Program Characteristics	**FDS**	**CCR**	**HS**	**PSSF**	**Safe CMP**
Is this program offered at multiple sites/statewide?	■	■		■	■
Comprehensive case management	■	■	■	■	■
Funding support for immediate needs	■	■			■
Parenting-focused programming	■		■	■	■
On-site therapy			■		
On-site resources (e.g. food pantry, community assistance, benefits specialist	■		■		
Fatherhood assistance	■		■	■	
Respite Care				■	
Adoption/Foster-care support	■			■	
Family reunification services	■			■	
Family preservation	■			■	

Head Start (HS)

HS is a center-based program, with two home-visits per year. In addition to providing academic and social-emotional development, HS utilizes a comprehensive service model to promote whole family well-being (ECLKC, 2018). Families are eligible to enroll in HS on the basis of child age (3-6) and the presence of low income or the presence of other risk factors (e.g. homelessness, eligibility for public assistance, foster care; (ECLKC, 2022). Beginning in 2018 and as part of the county's two-generation approach, a HS grantee located in Jefferson County, Colorado, began administering the CFSA to parents in the fall and spring of the academic year to measure the efficacy of their family support efforts over time. This HS program offers an array of comprehensive family-focused services on site including father assistance, parenting classes, family therapy, job and resource fairs, a food pantry, a co-located WIC office and other forms of community assistance. This site collects CFSAs as part of their commitment to attune their outcome measurements with that of other community-based FSPs. *Fidelity monitoring.* This HS reported reviewing the quality of data being captured from families as well as other relevant system metrics on an ongoing basis. Monthly program evaluations are conducted, and outcomes analyzed to ensure programmatic fidelity and compliance.

Family Resource Center (FRC) Family Development Services (FDS)

FRC's are center-based programs that act as a "single point of entry for providing comprehensive, intensive, integrated, and collaborative community-based services for vulnerable families, individuals, children and youth" (OMNI & FDS, 2018, p. 7). Vulnerable families who enter FRCs are screened for presenting needs and are offered an

opportunity to learn more about family support services. Families may enroll in parenting education, job training, education classes, health insurance enrollment support, financial emergency assistance, and more intensive family-centered goal setting and case management through Family Development Services (FDS). All families engaged in FDS have a coordinated case manager who assists the family in settings goals, matching to services, and linking families with additional referrals. Although FDS services are initially accessed in a center-based program, FDS service delivery is mixed in that families select their preferred service setting, spanning center, home, or community-based session with their case manager. Across 32 FRCs, 52 of 64 Colorado counties are served by FRCs (FRCA & OMNI, 2020). *Fidelity monitoring.* Guided by implementation science, FDS co-developed, with Member Center staff, intermediary and practice-level implementation drivers to support the provision of effective Family Development Services with fidelity. A fidelity monitoring assessment and a standard process to assess and improve fidelity were implemented in 2019. Prior to 2019, staff meeting training requirements, assessment follow-up rates, and family engagements around goal progress were used to know if a center was implementing Family Development according to guidance.

Colorado Community Response (CCR)

CCR is a home and community-based program, formed in 2013 to mitigate the risk of child maltreatment by strengthening families' protective factors, building social capital, increasing financial stability and self-sufficiency, and improving family functioning and well-being (Allan et al., 2018). In particular, CCR emphasizes improvement of family

concrete and social supports (Everson et al., 2021). Families referred to CCR were reported to Child Protective Services (CPS) but were either screened out, closed assessments, or cases without child welfare services. Families voluntarily engaged in these services, involving comprehensive case management for 12-16 weeks in addition to funding support to address immediate needs. In a four-year evaluation period (2014-2018), 1,900 families engaged in CCR services across 28 counties. At the time data was collected for the present study, participants only accessed CCR via CPS; however, the program is currently available to any eligible family in Colorado and no longer requires a CPS referral. *Fidelity monitoring.* CCR programming is currently developing implementation fidelity metrics, although none were formally in place at the time of this data collection.

SafeCare Case Management Pilot (Safe CMP)

The SafeCare® model, first implemented in Colorado in 2009, is a home-visiting model developed by Georgia State University and currently supported by the National SafeCare Training and Research Center. SafeCare is recognized as an evidence-based Home Visiting intervention by the Health Resources and Services Administration. As of 2020, 30 counties and 1 Tribe currently implement SafeCare (Colorado Department of Human Services, 2020). Currently, referrals to SafeCare are made through a variety of entry points including screened-out Child Welfare referrals, closed out Child Welfare assessments, provider-referrals, and self-referral. Despite widened eligibility criteria, close to half of families enrolled in SafeCare have a prior child welfare touchpoint (e.g. screened out referrals, Colorado Department of Human Services, 2020). Families are

eligible to participate in SafeCare if they have a child under the age of 6 and either prior child welfare involvement or the presence of three risk factors (e.g. single parent, parent age under 20, low caregiver education, housing instability, substance use, and violence exposure; Beachy-Quick et al., 2017). SafeCare emphasizes parent-child interaction, home safety, and child health for families with children under age 5. Four sites, beginning in 2018, piloted an adapted version of SafeCare, called the SafeCare Case Management Pilot (Safe CMP), that integrated components of Colorado Community Response (CCR) case management. All participants in this sample were enrolled in Safe CMP. Presently, all participants enrolled in SafeCare can access this case management service. *Fidelity monitoring.* Implementation fidelity is monitored for SafeCare on an ongoing basis through an intermediary organization.

Promoting Safe and Stable Families (PSSF)

PSSF, funded through Title IV–B, Subpart 2 of the Social Security Act, provides intensive case management and family support services to families experiencing elevated risk. PSSF focuses on four services areas: family support, family preservation, time-limited family reunification, and adoption promotion and support (Casey Family Programs, 2011). Families may enter services anywhere on this service spectrum and receive increasingly intensive levels of service depending on their level of need. The family support provided via PSSF is tightly tied to local communities, resulting in varied service delivery systems. Because an approximately even mix of site-based and home-based services are conducted statewide, PSSF is considered a mixed delivery system. Although it is unknown if families received home-based, center-based, or both forms of

services while enrolled, all families received individualized goal-setting and case management. In Colorado 23 county-administered sites provided PSSF services within 36 counties, including in both of Colorado's recognized tribes. Approximately 20% of the program's resources are devoted to each service area and services range in their intensity from family support with no required risk factors present to adoption promotion and support, which provides support to families preparing to adopt or following adoption. In 2018, 1,553 families received family support services and 599 received family preservation services (Colorado Office of Early Childhood, 2019). PSSF emphasizes the use of long-term parenting focused services over provision of concrete supports. Thus, in addition to receiving intensive case management, families are often linked to the following parenting programs: Incredible Years Parenting Program, Nurturing Fathers Program, and Nurturing Parenting. Families who received six hours or more of case management supports were required to measure progress using the CFSA. In the present sample of families who completed at least one CFSA, 615 families (46%) received family support services; 117 (8.8%) received family preservation services; 106 (8%) received adoption support; 25 (1.9%) received time-limited reunification support; 260 (19%) received family support, preservation, or time-limited reunification, but further designation from local providers was unspecified; and 208 (16%) did not have service designation information available. *Fidelity monitoring.* Implementation fidelity has varied in PSSF by parent education program. Implementation fidelity has been measured for Incredible Years, but not yet for other types of parenting education or intensive case management.

Participants

Participants were identified through case records embedded within each FSP across three partner organizations. Dates and labels on the case records of completed CFSAs did not consistently align (e.g. the order of dates did not always correspond to the labels 'Baseline' and 'Follow-up'); thus, a variety of data cleaning steps were conducted to organize and prepare data in a series of timepoints using the best available information (see Appendix A for data file preparation details). Typically, records were organized based on the date of assessment. Only cases with at least one timepoint of either the CFSA or the PFS were included. When CFSA records had less than one week between timepoints (n=22), secondary timepoints were treated as duplicates and only one was retained (see Appendix A). For the purpose of this analysis, records with more than 18 months between the first two CFSA timepoints (n=222) only had the first timepoint analyzed and were treated as if they did not have a second timepoint (a gap over 18 months may have signified a new or restarted treatment episode). In comparison to other cases, participants with greater than 18 months in between CFSA timepoints were all enrolled in FDS (n=191) or PSSF (n=31) and on average had higher mean CFSA scores at timepoint 1, were less likely to be Latinx and more likely to report English as a second language. CFSA records were additionally obtained from a funding stream identified as Community-Based Child Abuse Prevention; these participants (n=31) had high levels of missing data and were removed from the file. At timepoint 1, 248 participants received services without completing the CFSA or PFS (FDS, 205; HS, 8; CCR, 30; PSSF, 8; Safe

CMP, 1). Sixteen participants received services without completing the CFSA or PFS at either timepoint (FDS, 11; HS, 5) and five cases had no demographic characteristics reported (CCR, 4; Safe CMP, 1); these cases were removed. Following these data cleaning steps, a total of 15,771 participants had a minimum of 1 timepoint on the CFSA and a total of 9,626 had a minimum of 2 timepoints on the CFSA (see *Table 2*).

Table 2. Participants by Timepoint

Family Support Program	Timepoint 1	Timepoint 2
Family Resource Center Association (FDS)	9,589	5,974
Colorado Community Response (CCR)	4,010	2,692
Promoting Safe and Stable Families (PSSF)	1,331	334
Head Start, Jefferson County (HS)	581	463
Safe CMP + CCR combo (Safe CMP)	260	163
Total	*15,771*	*9,626*

Participants were organized by cohort year, following removal of outliers (e.g. the CFSA did not exist prior to 2014, so dates below 2013 were treated as data entry errors and not counted, n=6). The average year of enrollment was 2017 (SD=1.4 years; Range=2014-2020).

Measures

Demographic predictors

The following demographic predictors were used in analyses: family income, parent education, household size, parent gender, parent age, ethnicity, minority race, family language, and household structure. Child age is reported but not used in analyses due to

high levels of missing data. An additional index of high stress family events was created for analysis. The number of demographic variables available differed by FSP (see *Table 3).* Only 940 (6%) of cases had complete demographics on all 10 indicators; however, when child age was removed (a demographic with high levels of missingness), 3,070 (19.5%) had complete demographic data (FDS, 1,994 cases [20.8%]; HS, 553 cases [95%]; CCR, 424 cases [10.6%]; PSSF, 63 cases [4.7%]; Safe CMP, 36 cases [13.9%]).

Table 3. Data by FSP

FSP	Inc	Pnt Edu	Hsld Size	Pnt Gndr	Pnt Age	Ethn	Race	Lang	Hsld Stcr	C Age
FDS	9,369	9,166	9,360	9,562	9,275	9,425	6,535	3,496	3,604	--
CCR	3,875	3,912	3,992	2,342	1,943	1032	955	2,524	565	1,423
PSSF	1,303	1,257	1,330	896	888	628	840	1,140	102	577
JCHS	581	573	581	580	578	581	568	577	581	564
Safe CMP	260	260	260	118	100	76	71	118	49	52
Total	***15,388***	***15,168***	***15,523***	***13,498***	***12,784***	***11,742***	***8,969***	***7,855***	***4,901***	***2,616***

Note: Inc=Income, Pnt Edu=Parent education, Hsld Size=Household size, Pnt Gndr=Parent gender, Pnt Age=Parent age, Ethn=Ethnicity, Lang=Primary language, Hsld Stcr=Household structure, C Age=child age

Family income. Reported income was pulled from demographic screening when available and, when unavailable, was pulled from the CFSA form (sub-item under Domain 1). Data were initially available on 15,392 participants (M=23,423, SD=105,520). However, data were not normally distributed (Skewness=105.1, SE=0.02; Kurtosis=12,121.6, SE=0.04) and three participants had extreme outliers that exceeded the interquartile range by more than 33 times. These outliers were removed (along with an apparent data entry error of -1), resulting in an average income of $22,376

(SD=$25,151, Range=0-$494,000, see *Table 4*) and somewhat improved data normality (Skewness=3.84, SE=0.02; Kurtosis=33.84; SE=0.04). Although kurtosis continued to exceed recommended ranges, data were not further transformed or cleaned, as the data nonnormality was considered to accurately reflect the income distribution of the population served (11% of the sample reported $0 annual income and 25% reported income under $6,000). For analyses, income was examined at a low value of $0, the 50th percentile of $18,000 and the 75th percentile of $30,000. Data were missing on 2.4% of the full sample, ranging from a low of no missing data in HS to a high of 3.4% missing in CCR.

Parent Education. This demographic variable was available on a subset of participant demographics (n=675) and was supplemented using the information collected on the CFSA timepoint 1 on the level of adult education in the home, resulting in a final sample of 15,168. In 227 cases, information provided in the CFSA did not align with provided participant demographics on caregiver education. In these cases, demographic information was retained as primary, due to less precision on the CFSA (the CFSA measures all adults in the household versus the primary caregiver assessed in the demographics). Due to the larger amount of available data from the CFSA timepoint 1, all data were recoded across programs into the following groups that aligned with the CFSA: 0=Less than a High School Education (n=2,643, 17.4%), 1=GED/High School Diploma (n=5,005, 33%), 2=Some college with no degree or current student working on college degree or specialized training/certification (n=2,093, 13.8%), 3=Has a professional certification/training, associates, bachelors, or higher degree (n=5,427,

35.8%). See *Table 4* for education by program. For analyses, education was examined continuously with low education examined as 'less than a high school education', mid-levels of education as 'GED or High School Diploma' and at high levels as 'completion of a professional certification or college degree'. Data were missing on 3.8% of the full sample, ranging from a low of 0.4% missing in Safe CMP to a high of 5.6% missing in PSSF.

Household size. This demographic variable was available on a subset of participant demographics (n=4,543) and additionally was collected on the CFSA (n=15,538) through the following question: "How many people are in your family including yourself?" In 2,680 cases (17%) of cases with household size information, size reported on demographics did not align with household size reported on the CFSA. Given that household size may have changed between initial recording of family demographics at intake and CFSA completion, the CFSA values were used first when available and when unavailable were supplemented with demographic intake information on household size (n=158). Initially, data were available on 15,538 cases (M=69.7, SD=5,607), but were not normally distributed (Skewness=106.1, SE=0.02; Kurtosis=11,877.9, SE=0.04), which appeared to be caused by 10 outliers which exceeded the interquartile range by 164 times or more (i.e., likely to be data entry errors, e.g. 650,000 household members). These values were removed (along with five data entry errors of '0'), resulting in a sample of 15,523 (M=3.7, SD=1.5, Range=1-16, see *Table 4*) and a normal data distribution (Skewness=.89, SE=0.02; Kurtosis=1.7, SE=0.04). For analyses, household size was examined by average household size and higher versus lower household sizes were

examined as one SD above and below the mean, rounded to the closest discrete value. After accounting for outliers and 'NA' items, data was missing on 1.6% of the full sample, ranging from a low of 0% missing in HS to a high of 2.4% missing in FDS.

Parent Gender. This demographic variable was collected on 14,155 participants, including 657 'NA' responses. On OEC cases, when primary caregiver designation was unavailable, primary contact gender was used if it was determined the primary contact for a case was not a child (see Appendix A). After updating 'NA' entries to missing, 13,498 participants had information on Gender (83.4% female, see *Table 4*). Responses were coded as 0=Female, 1=Male. Data were missing on 14.4% of the full sample, ranging from a low of 0.2% missing in HS to a high of 54.6% missing in Safe CMP.

Parent Age. This demographic variable was recorded on 14,127 participants either through designation as a primary caregiver or a primary contact on a case. Data were not normally distributed (Skewness=-.27, SE=0.02; Kurtosis=4.5, SE=0.04) and 38 outliers were identified that exceed the interquartile range by more than 8.5 times (outliers appeared to be data entry errors, e.g. 120). An additional 1,180 participants had values at or below 0, indicating data entry errors. Finally, 122 participants had parent ages reported between the ages of 1 and 14. Because the parental age could not be verified, a cut-off was set at age 14 and all values below 14 were treated as missing data. Following removal of outliers, 12,787 participants had parent age (M=35.15, SD=10.06, Range=14-

Table 4. Sample Demographics

Demographics	**FDS**	**HS**	**CCR** *M (SD)*	**PSSF**	**Safe CMP**	**Whole Sample**	*Range*
Family Income	$21,320 (22,636)	$14,885 (16,540)	$23,598 (22,948)	$28,232 (34,829)	$19,127 (12,913)	$22,376 (25,152)	0-$494,000
Household Size	3.7 (1.5)	4 (1.4)	3.7 (1.5)	3.8 (1.5)	3.9 (1.6)	3.7 (1.5)	1-16
Parent Age	34.3(9.7)	32.4 (7.5)	38.8 (10.2)	36.7 (11.7)	33.5 (6.4)	35.15 (10.06)	14-101
Child Age	--	3.6 (.62)	10.1 (4.8)	7.5 (4.9)	4.4 (2.3)	7.4 (5.07)	0-18.9
Stress Index	.46 (.73)	.73 (.87)	.99 (.91)	.68 (.88)	.46 (.68)	.62 (.83)	0-5
			valid %				
Parent Education							
< High School	18%	9.9%	17%	17.3%	21.5%	17.4%	--
GED/High School Diploma	32.2%	35.3%	35.9%	27.6%	38.5%	33%	--
Some college/current student	13.5%	8.7%	15.3%	13.3%	14.6%	13.8%	--
College/professional degree	36.3%	46%	31.8%	41.8%	25.4%	35.8%	--
Single-parent household	48.5%	60.2%	62.7%	61.8%	40.8%	51.7%	--
Male caregiver	17.1%	9.7%	14.6%	21.7%	5.1%	16.6%	--
BIPOC	19.9%	30.8%	12.7%	20.6%	12.7%	19.8%	--
Latinx	40.4%	54.2%	45.1%	56.4%	59.2%	42.5%	--
ESL	16.6%	10.1%	10.8%	14.4%	24.6%	14.1%	--

101.2, see *Table 4*) and data reflected a normal distribution (Skewness=0.82, SE=0.02; Kurtosis=0.91, SE=0.04). For analyses, parent age was examined as mean age with older versus younger parent age examined as one SD above and below the mean, respectively. Data were missing on 18.9% of the full sample but ranged from a low of 0.5% missing in HS to a high of 61.5% missing in Safe CMP.

Parent Ethnicity. This demographic variable was available on 14,743, with 3,001 (20.4%) of the responses listed as 'NA' or 'Declined to Answer'. For OEC cases, when parent ethnicity was unavailable, primary contact ethnicity was used or, in some cases, child ethnicity was used as a proxy for caregiver ethnicity (see Appendix A). Although it is possible that child ethnicity did not match parent ethnicity in all cases, these families likely also experience need tied to their child's race or ethnicity. For example, transracial and transcultural caregivers report needed resources tied to their child's race and ethnicity along with limited available resources and provider support (de Hayes & Simon, 2003). Data in these cases accounted for a very small proportion of available data, less than 1% of data in the OEC. 'NA' and 'Declined to Answer' responses were recoded as missing for analysis, resulting in a final sample of 11,742 participants (0=non-Hispanic/Latinx; 1=Hispanic/Latinx ethnicity). In the final sample 42.5% of participants (n=4,988) identified as Hispanic/Latinx ethnicity (see *Table 4*). Data were missing on 25.5% of the full sample, ranging from a low of 0% in HS to a high of 74.3% missing in CCR.

Parent Race*.* This demographic variable was provided on 11,714 participants, with 2,745 (23.4%) responses coded as 'Unspecified/Declined to Answer.' These responses were recoded as missing for analysis, resulting in a final sample of 8,969 participants. For

OEC cases, when parent race was unavailable, primary contact race was used or, in some cases, child race was used as a proxy for caregiver race (see Appendix A). As with ethnicity, instances in which child race was used as a proxy for family needs and experiences accounted for less than 1% of the OEC's data. Responses were recoded across programs into the following groups: White (80.2%), Black or African American (7.7%), Multi-racial/Biracial (3.4%), American Indian or Alaskan Native (6.6%), Asian (0.9%), Native Hawaiian/Other Pacific Islander (0.6%), and other (0.7%). See *Table 4* for breakdown by program. For the purpose of analysis and to manage small sample sizes outside of the Black and American Indian or Alaskan Native groups, race was recoded into a variable identifying Black, Indigenous, and People of Color (BIPOC) participants in contrast to White (majority race) participants (0=White, 1=BIPOC). In the final sample, 19.8% of participants (n=1,780) were identified as BIPOC. This variable was missing on 43.1% of the full sample, ranging from a low of 2.2% in HS to a high of 76.2% in CCR.

Family Language. This demographic variable was available on 8,694 participants, with 839 (9.7%) of responses coded as 'NA'. These responses were recoded as missing for analysis, resulting in a final sample of 7,855 participants. For OEC cases, when parent language was unavailable, primary contact language was used or, in some cases, child language was used as a proxy for caregiver language. As with race and ethnicity, this accounted for less than 1% of available data. Responses were recoded across programs into the following groups: English (86%), Spanish (13%), and other (1%). For analysis and to manage the small 'other' sample size, family language was recoded into English as

a Second Language (ESL, 0=Family Language is English; 1=Family Language is Spanish/Other). In the final sample, 14.1% of participants (n=1,106) were identified as ESL (see *Table 4*). This variable was missing on 50.2% of the full sample, ranging from a low of 0.7% missing in HS to a high of 63.5% in FDS.

Household Structure. This demographic was available on 5,407 participants and, when unavailable, was supplemented through marital information (N=1,714). Marital status was used to approximate household structure for 716 OEC Cases and for 998 FDS cases. Marital status was recoded in the following way: Divorced, separated, single, and widowed primary caregivers were coded as 'single-parent household'. Living together and married primary caregivers were coded as 'two-parent household'. No data were available on re-partnered or remarried households. This coding scheme was examined for accuracy within a subsample of FDS cases with both household structure and marital status information (n=2,141), where coding was identified as aligning with household structure 93% of the time. Values of 'other' and 'NA' (N=2,220) were coded as missing data. Data were coded as 0=two-parent households and 1=single-parent household. In total, household structure data was available on 4,901 participants, 52% (n=2,536) of whom were single-parent households (see *Table 4*). Information on household structure and/or marital status was missing on 68.9% of the full sample, ranging from a low of 0% missing in HS to a high of 92.3% missing in PSSF.

Child Age*.* This demographic variable was available on 2,681 participants (M=8.0, SD=7.7, Range=0-120.8), either through designation as a primary child or a primary contact. When primary contacts were between the ages of 14 and 19 coding as child

versus adult was decided based on the presence of other adult contacts in the home (see Appendix A). Data were not normally distributed (Skewness=7.43, SE=0.05; Kurtosis=100.2, SE=.10), which appeared to be related to seven outliers which exceeded the interquartile range by more than 12 times. An additional 58 child ages ranges from 19-40. Because child age in these cases could not be verified and this analysis is primarily concerned with dependent children, all child ages over the value of 19 (n=65) were removed from analysis. Following cleaning, data were available on 2,616 participants in total (M=7.4, SD=5.07, Range=0-18.9, see *Table 4*) and data normality was improved (Skewness=.61, SE=0.05; Kurtosis=-0.78, SE=.10). Data were missing on 83.4% of the full sample, ranging from a low of 2.9% of cases in HS to a high of 100% missing in FDS. Although FDS had collected this data, it was unretrievable at the time of analysis due to a software change. Child age was not used in program change score analyses due to high missingness and limited alignment across programs (e.g. eligibility criteria in one program excludes children under the age of three and two programs exclude children over the age of five).

High Stress Events Index. A composite variable was created, capturing a cumulative index of up to six high stress events or experiences: no reported income, insufficient food, homelessness or transitional/shelter housing, substance use requiring prevention or intervention efforts, domestic violence, and child abuse or neglect. Across all programs, information collected from the CFSA documenting income, housing needs (homeless or in a shelter), substance use (moderately to severely impairing), and food sufficiency (insufficient quantity of food) were utilized. Other forms of information varied by

program and supplemented information pulled from the CFSA. In the OEC programs (PSSF, Safe CMP, CCR), intake questionnaires included information on the family housing, substance use, food sufficiency, and domestic violence and case closure notes indicated whether intake problems had been addressed and/or which resources were offered at the close of the treatment session. These notes were coded for reports of substance use (e.g. terms including substance, alcohol, drugs, relapse), domestic violence, and child welfare concerns (e.g. terms including CPS, child abuse, sexual abuse, neglect). In the HS program, family service needs were identified for every participant, including domestic violence, child welfare, substance use treatment, substance use prevention, and housing services. Information on risk for domestic violence was also used based on participant responses of 'No' to the PFS prompt, "I feel safe in my relationships." In FDS, service utilization was used to identify participants who had likely experienced the above stressors. For example, utilization of housing assistance and supervised visitation were used to identify families with likely housing issues and child welfare concerns. Additionally, information on risk for domestic violence was based on participant responses of 'No' to the PFS prompt, "I feel safe in my relationships". In total, the risk index averaged 0.60 out of a possible 6 (SD=0.80; Range=0-5, see *Table 4*), with means ranging from 0.46 in Safe CMP (SD=0.68) and FDS (SD=.73) to 0.99 in CCR (SD=0.91). For analyses, the stress index was examined as a low, indicated by no stress events, average, indicated by one stress event, and high, indicated by two or more stress events.

All missing demographics were determined to be Missing At Random (MAR) using Little's MCAR test, $\chi 2$=636.9, DF=67, p<.000. Missingness varied by FSP and, per correspondence between the researcher and site administrators, it appears that staff training for data entry and mandatory fields within data entry software varied widely, which likely contributed to variation in patterns of missing data by FSP.

Colorado Family Support Assessment

The CFSA, developed by the Family Resource Center Association (2014) to measure self-reliance and well-being in families, is a three-part measure, capturing family self-sufficiency, protective factors, and family readiness to change. The following three sub-sections of the CFSA are reviewed in turn.

Part A. Self-sufficiency matrix (Richmond, Pampel, Zarcula, Howey, & McChesney, 2015). The self-sufficiency matrix underwent reliability testing in 2015 and confirmatory factor analysis (CFA) in 2018 (OMNI & Family Resource Center Association, 2018). The scale measures self-sufficiency across 14 domains, covering income, employment, housing, transportation, food security, child care, child education, adult education, cash savings, debt management, health coverage, physical health, mental health, and substance abuse. A prior Confirmatory Factor Analysis identified a two-factor structure in the measure: economic self-sufficiency (income, employment, housing, transportation, food security, adult education, cash savings, and health insurance coverage domains) and health (physical and mental health; Richmond et al., 2017). Debt management, child education, and child care were analyzed separately, as they didn't load onto a factor. A reliability study on the CFSA found good inter-rater reliability, with intraclass correlation

coefficients ranging from .79 - .96. with average accuracy at 84.4% (Richmond et al., 2017).

Each of the domains are rated on a scale of 1 to 5 using domain specific indicators ranging from “thriving” to “in crisis”. Families who score a 3 or above in each domain cross a “prevention” line and are considered to be in safe, stable, or thriving environments. For example, food security (a domain within the economic self-sufficiency factor), is defined as, “a family’s level of food security based on USDA definitions.” Scores range from 1, “Very low food security: food intake reduced for one or more family members because the household lacks money or other resources for food,” to 5, “high food security: family members have no problems, or anxiety about, accessing enough quality food with variety.” Mental health (a domain within the health factor) is defined as, “the degree to which any family member’s mental health issues interfere with life activities.” Scores range from 1, “Family member’s mental health concerns prohibit important life activities (including work, school, caring for children, managing a household, or reaching developmental milestones for young children),” to 5, “Family member(s) have no known ongoing mental health problems.”

Per technical guidance on the use of the self-sufficiency subscale (OMNI & Family Resource Center Association, 2018), a mean score was calculated for participants with a minimum of 6 out of 8 responses. On the health subscale, a mean score was calculated for participants with a minimum of 1 out of 2 possible responses. At timepoint 1, the self-sufficiency subscale was available on 15,309 participants (M=2.89, SD=0.76, Range=1-5) and the health subscale was available on 15,357 participants (M=4.0, SD=1.06,

Range=1-5). At timepoint 2, the self-sufficiency subscale was available on 8,758 participants (M=3.05, SD=0.72, Range=1-5) and the health subscale was available on 8,777 participants (M=4.13, SD=0.97, Range=1-5). Subscales were normally distributed at all timepoints. At timepoint 1, missing data on subscales in the full sample was at or below 2.9% and, at timepoint 2, was at or below 9%.

Three domains are not captured within the health or self-sufficiency factors and will be separately analyzed. First, child care is assessed as "the family's ability to obtain reliable, affordable, and quality childcare." Scores range from 1, "needs child care, but none is available or accessible or child is unsupervised and may be unsafe," to 5 indicating all of the following, "child care is reliable, child care is affordable without subsidies, child care is quality, reliable back-up childcare options are available when needed." Child education is assessed as "school-aged children's access to and engagement in educational institutions." Scores range from 1, "any child in the family is not enrolled in school," to 5 "No child in the family has truancy / disciplinary actions at school AND all children are meeting academic achievement expectations AND any child is exceeding academic achievement expectations." Finally, debt management is assessed as "the degree to which a family is managing debt." Scores range from 1, "inability or limited ability to pay down debt (may be making payments but cannot meet minimum required payment)" to 5, "family is debt-free." Missing data on child care, child education, and debt domains ranged from a low of 6.8% on debt at timepoint 1 to a high of 36.2% on childcare at timepoint 2. Missing data on child care and child education was

determined to be missing not at random (MNAR), due to valid response options of 'not applicable' which significantly impacted rates of missingness (see Appendix B).

A total mean CFSA score was additionally calculated, excluding scores for Child Care and Child Education given high levels of valid missingness, if a minimum of 9 out of 12 possible domains were available (75% of data points). At timepoint 1, the total CFSA mean score was available on 15,165 participants (M=3.22, SD=0.64; Range=1.25-5) and, at timepoint 2, the total CFSA mean score was available on 8,736 participants (M=3.37, SD=0.61, Range=1-5). At timepoint 1, the CFSA mean score had 3.8% missing data and, at timepoint 2, had 9.2% missing data. When examining missing data at timepoint 2 by FSP (see Appendix B), missing data on the CFSA subscales and total mean CFSA score varied from a low of 0.6% in Safe CMP to a high of 12.8% missing in FDS. In fact, no individual FSP exceeded 5% missing data on subscales or total mean score at timepoint 2 outside of FDS. Thus, missing data were only analyzed further within FDS. A missing total mean score on the CFSA at timepoint 2 (indicating less than 75% of the data points were complete) was associated with lower income, $t(1190)=6.7$, $p<.000$; smaller family size, $t(971)=2.5$, $p<.05$; single-parent households, $\chi 2$ (1, N=2631)=29.2, $p=<.001$; fathers, $\chi 2$ (1, N=5962)=14.6, $p=<.001$; English as a primary language in the household , $\chi 2$ (1, N=2592)=27.9, $p=<.001$; and BIPOC families, $\chi 2$ (1, N=3974)=15.2, $p=<.000$. Little's MCAR test confirmed that data was Missing at Random (MAR), $\chi 2=74.04$, DF=21, $p<.000$.

Part B. Protective Factors Survey (PFS; Counts, Buffington, Chang-Rios, Rasmussen, & Preacher, 2010). The PFS, developed by the FRIENDS Network in

collaboration with the University of Kansas Institute for Educational Research and Public Service, is a well-validated tool for maltreatment prevention. It has undergone three national field tests and has been found to have strong content, construct, and criterion validity. The survey measures five family protective factors: family functioning/resiliency (a = .92), social support (a = .91), concrete support (a = .74), nurturing and attachment (a = .83), and child development/knowledge of parenting (deemed unscalable and measured through five questions). Questions are rated on a Likert scale from 1-7 with scores ranging from "never" to "always." Questions include, "in my family, we talk about problems," "there are many times when I don't know what to do as a parent," "My child misbehaves just to upset me," "my child and I are very close to each other," and "I am able to soothe my child when he/she is upset."

Per technical guidance on construction of the subscales (FRIENDS National Center for Community Based Child Abuse Prevention, 2020), the family functioning and resilience subscale is only calculated if a minimum of 4 out of 5 possible questions are completed; the social support subscale is only calculated if a minimum of 2 out of a possible 3 questions are completed; the concrete support subscale is only calculated if a minimum of 2 out of 3 possible questions are completed; and the nurturing and attachment subscale is only calculated if a minimum of 3 out of 4 possible questions are completed. At timepoint 1, the family functioning and resiliency subscale was available for 12,719 participants (M=4.44, SD=1.2, Range=0-6); the social support subscale was available for 12,740 participants (M=4.52, SD=1.50, Range=0-6); the concrete support subscale was available for 12,734 participants (M=4.14, SD=1.66, Range=0-6); and the

nurturing and attachment subscale was available for 12,002 participants (M=5.27, SD=0.86, Range=0-6). At timepoint 2, the family functioning and resiliency subscale was available for 6,978 participants (M=4.63, SD=1.1, Range=0-6); the social support subscale was available for 6,980 participants (M=4.76, SD=1.3, Range=0-6); the concrete support subscale was available for 6,983 participants (M=4.49, SD=1.6, Range=0-6); and the nurturing and attachment subscale was available for 6,699 participants (M=5.33, SD=0.79, Range=0-6). Data distribution was normal on all subscales at all timepoints, except for timepoint 2 on the nurturing and attachment subscale where kurtosis exceeded normal distribution (Kurtosis=3.9, SE=0.03), but still fell within a range of moderate normality (Hair et al., 2010) and, given large sample sizes, was not anticipated to violate assumptions of normality. At timepoint 1, subscale data on the full sample were missing from a low of 19.2% on the social support subscale to a high of 24% on the nurturing and attachment subscale. At timepoint 2, subscale data on the full sample were missing from a low 27.5% on all subscales except for nurturing and attachment, which was 30.4% missing. Missingness at timepoint 1 varied by FSP, however (see Appendix B for details), from a low of 2.9% missing in Head Start to a high of 31.8% missing in CCR. At timepoint 2, missingness varied from a low of 4.3% missing in Safe CMP to a high of 35.3% missing in CCR.

Only FSPs with missing data exceeding 5% were examined collectively further (PSSF, CCR, and FDS at both timepoints and Head Start at timepoint 2). Missing data at timepoint 1 on all four PFS subscales was associated with English as primary language, white race families, non-Latinx ethnicity, completion of high-school or some college

(except the nurturing and attachment subscale), fathers, lower income, older parent age, and smaller household sizes (see Appendix B *Table 1*). Missing data at timepoint 2 on all four PFS subscales was associated with single-parent households (except nurturing and attachment subscale), English as primary language, white race (except for nurturing and attachment subscale), non-Latinx ethnicity, fathers, lower income, older parent age, and smaller household size (see Appendix B *Table 1*). Data were examined next with Little's MCAR test individually by FSP to determine pattern of missing data, confirming that data were MAR for FDS at timepoints 1, $\chi 2$=8458.8, DF=839, p<.000, and 2, $\chi 2$=4752.2, DF=539, p<.000; data were MAR for CCR at timepoints 1, $\chi 2$=1141.91, DF=70., p<.000, and 2, $\chi 2$=841.9, DF=614, p<.000; data were MAR for PSSF at timepoints 1, $\chi 2$=959.8, DF=510, p<.000, and timepoint 2, $\chi 2$=394.8, DF=309, p<.01; and data were MAR for HS at timepoint 2, $\chi 2$=118.2, DF=83, p<.01.

Part C. Readiness to Change. Readiness to change was collected at each CFSA on an array of 18 treatment domains (including the 14 domains on the CFSA, parenting skills, child development, and social support). By each domain, participants indicated with a checkbox their desire to make change in a given area and then estimated how ready they felt to make change in each checked area on a scale of 1 to 10. For analyses, readiness to change at treatment baseline was included as a covariate. For subscales readiness to change represented a mean score at baseline, calculated as the average rating (sum of values 1 through 10) divided by the number of areas of indicated desired change (up to 8 on the self-sufficiency subscale). Readiness to change on the self-sufficiency subscale was available on 14,553 participants (M=5.75, SD=4.2) and on the health subscale was

available on 14,501participants (M=2.16, SD=3.83). Readiness to change data for child care was available for 9,633 participants (M=1.67, SD=.43); debt was available on 9,695 participants (M=2.12, SD=3.74); on substance abuse was available on 9,104 participants (M=.35, SD=1.71); on child education was available on 14,501 participants (M=1.12, SD=3.02); on child development was available on 9,591 participants (M=1.97, SD=3.73); on parenting skills was available on 10,277 participants (M=3.04, SD=4.22); and on social support was available on 9,688 participants (M=1.35, SD=3.10). Readiness to change scores were subsequently matched to all subscales and standalone outcomes on the CFSA (self-sufficiency, health, child care, child education, and substance use), as well as all measures on the PFS (utilizing child development, parenting skills, or social support as covariates for the 4 PFS subscales and 4 standalone outcomes, depending on the closest content match). Data were missing on 8.1% of participants across the full sample, ranging from no missing data in PSSF, CCR, and Safe CMP to 13% missing in FDS.

Data Analysis Plan

A series of steps were taken to prepare data, examine missing data across samples, and select a strategy for management of missing data. To address question 1, patterns of dropout within and across samples were calculated. To proceed to questions 2 and 3, samples were then weighted to correct for dropout bias. In the final integrated data analysis, whole sample bivariate and multivariate associations were calculated, followed by examination of moderation between program and predictor.

Data Preparation

A series of steps were taken to manage missing data and weight the sample to correct for dropout bias. Preliminary analyses revealed the data to be generally MAR and generally above 5% missing (at both timepoints for the PFS). Multiple Imputation (MI) has been consistently identified as a best practice for missing data, capable of lowering bias and standard errors introduced by missing data when missing data is not severe (under 50%; Buhi et al., 2008; Schafer & Olsen, 1998; Schlomer et al., 2010), when models are well-specified, and when auxiliary information is provided (Madley-Dowd et al., 2019). Furthermore, multiple imputation has been demonstrated as successful even in 'difficult' applied datasets with up to 60% missing data (Pampaka et al., 2016). Another method, Inverse-Probability Weighting (IPW; Little, 1986; Schmidt & Woll, 2017), is a method employed to adjust for the bias introduced by attrition in longitudinal designs, wherein unequal sampling can bias results toward the participants who did not drop out. In accordance with Seaman and colleagues (2012), IPW and MI were used in conjunction to address dropout and missing data, with MI only used on outcome data for participants who participated at timepoint 2 and IPW correcting for bias in dropout prior to timepoint 2. This method was selected given an increasing number of studies (discussed in Seaman et al., 2012) that have successfully combined IPW and MI to adjust for attrition and missing data. The combination of IPW and MI has advantages over use of either one method alone and does not result in an inflated missing data parameter estimation, parameter variance, or standard errors (see Seaman et al., 2012). In this approach, IPW is calculated first and then applied to MI analyses.

Inverse Probability Weighting

Prior to calculating IPW, missingness was corrected for in the demographic predictors used to calculate likelihood of dropout. Bias is introduced when predictors of missingness themselves are missing; thus, MI was first was calculated on predictor variables prior to developing weights, increasing the number of complete cases available for the Inverse Probability Weights (IPW) analysis (see Seaman & White, 2014; Thomas et al., 2007). Imputations were only conducted on variables with a maximum of 50% missing data (following guidance by Garson, 2015) and values were predicted by demographic characteristics, interaction terms between FSPs and demographic characteristics (all continuous variables were centered prior to creating interaction terms), associated covariates, and outcome subscales at timepoint 1, in order to provide auxiliary information (see Madley-Dowd et al., 2019). All imputations used fully conditional specification method (Markov Chain Monte Carlo) with 10 imputed data sets.

Next, logistic regression was used to calculate IPW (IPW; Little, 1986; Schmidt & Woll, 2017), where the inverse of the subpopulation's estimated probability of participation at timepoint 2 is used as a weight. Following guidance from Seaman and White (2013), each program's logistic regression was calculated separately and included covariates associated with missingness (*Table* 6) and covariates associated with outcome variables at timepoint 1, the combination of which are likely to reduce bias and increase model efficiency. Variables were entered simultaneously and model fit was examined using the Hosmer-Lemeshow's test (Hosmer & Lemeshow, 1989), along with examination of the Fraction of Missing Information for pooled imputation values.

Because not all indicators had been imputed (those with high levels of missingness), listwise deletion resulted in a constricted sample size for weights with the most available information. Thus, 1-2 additional weights were calculated per sample, progressively dropping variables with higher levels of missingness until a predicted probability representing the most available information existed for each participant. In Safe CMP, logistic regression could not be completed until household structure was dropped due to extremely low variance. In two samples (PSSF, Safe CMP) zero probability values emerged. Following guidance from Seaman & White (2013), dropping predictor(s) with high missingness (household structure, child age) corrected zero probability values. In PSSF, very low probabilities of retention additionally emerged (<10% in 29 participants in the original data file). Demographics were observed in these cases, child age was dropped, and interaction terms were added to the model to allow the joint effects of variables to be less multiplicative.

There were no indications of poor model fit on logistic regression models (indicated by 5 or more significant, $p<.05$, goodness of fit tests on imputed datasets), except in the third logistic regression sample in CCR (following removal of household structure and child age from models). To correct poor fit, nonsignificant predictors with dropout were removed and significant interaction terms among covariates were added. IPW weight values were estimated by saving predicted values in SPSS and the inverse calculated as 1/the predicted value. While this weighting technique can control for responder bias on target variables, it can also result in increased estimator variance (Schmidt & Woll, 2017). Thus, as a final step, weights were trimmed (Potter, 1990) at the 1st and 99th

percentiles. Per recommended practices, weights and trimming were separately calculated and applied for each FSP as a separate stratum (Schmidt & Woll, 2017).

Weights were pooled (aggregated) following MI and used in subsequent analyses. Pooled weights ranged from 1.11 to 11.98 (M=1.78, SD=1.06), with the 5th and 95th percentiles at 1.23 and 4.12. Fifty-one cases in PSSF had weights over 10. In these cases, the mean predicted probability of completing timepoint 2 was 0.07; similarly, only 4 of these participants (7.8%) had completed timepoint 2.

Multiple Imputation

Of cases with timepoint 2 CFSAs, 804 (8.4%) participants received services without completing the CFSA or PFS (FDS, 718; HS, 17; CCR, 66; PSSF, 3). These cases were dropped in timepoint 2 analyses (Garson, 2015). MI was only conducted on dependent variables identified as MAR or MCAR (Employment, Child Care, and Child Education were not imputed given data was NMAR). All models included all available demographic variables as predictors of imputation parameters, covariates associated with outcome variables at timepoint 2, and all interaction terms between demographic indicators and FSP. Multiple imputation was conducted on outcome data with missingness below 50% (missingness in the full sample at timepoint 2 ranged from a low of 0.5% on the CFSA Transportation domain to a high of 24.5 on PFS items 15 and 20). Multiple imputation was conducted (incorporating weighting from the previous step) using predictive mean matching (this approach was selected over linear regression due to its imputation of observable values, an important step for calculation of change scores, and its heightened robustness to model misspecification; Morris et al., 2014), a fully conditional

specification method (Markov Chain Monte Carlo), and 10 imputed data sets for dependent variables at both baseline (Jakobsen et al., 2017) and at timepoint 2 (Seaman et al., 2012).

Analytic Plan

Baseline differences in outcome measures were examined by program (*Table* 7) and by subgroup (see Appendix D Tables 1-3). For each outcome of interest, the following steps were employed: bivariate correlations (*Table 5)* identified demographic, stress indicators, and other covariates (e.g. length of treatment, readiness to change) associated with the outcome (change scores). All significantly associated indicators were entered simultaneously into an Analysis of Covariance model (ANCOVA) model to identify variables driving main effects in change scores. Interaction terms were further entered between demographic covariates by program in the model and Bonferroni corrections were used to correct for multiple comparisons. If interactions were insignificant, they were dropped from the model and, if both covariates and interactions were insignificant, they were both dropped from the model. In two models (substance abuse, social support improvement), missing data on the readiness to change covariate resulted in an extremely restricted sample size in Safe CMP (n=2-5), obstructing power. These models report results without Safe CMP (Appendix C includes models that retain SafeCMP and drop readiness to change for these two outcomes). Across models, significant interactions were probed further using Factorial Analysis of Variance (ANOVA) when covariates were categorical (all other covariates remained in these models to control for the effects of covariates) and when covariates were continuous, various sample points were pulled

based on the data's distribution to identify how the model looked at low, medium, and high levels of the continuous predictor. Due to utilizing multiple imputation and limited accessibility to pooled *F* statistics, all reported results include the range of *F* statistic estimates across the 10 imputed datasets. Across imputations, if 5 or more imputations conveyed an insignificant F statistic ($p>.10$), significant results were not reported.

Results

Family Engagement

Participant drop-out rates across all programs averaged 39%, ranging from a low of 20% in HS to a high of 75% in PSSF (*Table 2*). Initial bivariate correlations (*Table 5*) revealed a preliminary relationship between each demographic predictor of interest and dropout in at least one program, except for parent education which was not associated with dropout.

Table 5. Correlations with Dropout at Timepoint 2

FSP	Inc	Sing Par	P Age	Latinx	Hsld Sz	C Age	Male	ESL	BIPOC	Par Edu	Strss
FDS	-.06**	.07**	.03**	-.10**	-.05**	--	.09**	-.06**	.08**	-.01	.05**
HS	-.08†	.05	.03	.02	-.05	-.04	.05	-.02	-.02	-.04	.03
CCR	.01	.04	-.07**	-.06*	.03*	-.02	-.02	-.04*	-.00	.01	-.05**
PSSF	.06*	-.07	-.04	-.11*	.03	-.12**	-.09**	-.04	.12**	.03	-.12**
Safe CMP	-.01	.03	-.14	.11	-.03	.04	-.10	-.01	.05	-.08	.07

Note: Inc=income, Sing Par=single parent, P Age=parent age, Hsld Sz=household size, C Age=child age, Male=male caregiver, Par Edu=parent education, Strss=stress index. *†p<.10, *p<.05, **p<.01, ***p<.001*

Next, logistic regression was used to calculate the likelihood of retention, accounting for a wider array of predictive covariates per program. These models revealed a more select set of demographic factors with impacts on three programs (FDS, PSSF,

SafeCare CMP), several of which operated differently to predict retention at the program level (*Table 6*).

Table 6. Odds Ratio (OR) of Participation at Timepoint 2

FSP	Single Parent Exp(B)	Parent Age Exp(B)	Latinx Exp(B)	House-hold Size Exp(B)	BIPOC Exp(B)	Male Caregiver Exp(B)	Stress Index Exp(B)
FDS	.83*	.99***	1.41***	1.08*	-	.75**	.81***
HS	-	-	-	-	-	-	-
CCR	-	-	-	-	-	-	-
PSSF	-	-	-	-	.09**	-	1.4***
Safe CMP	-	-	.01*	-	-	-	-

Note: All continuous variables were centered before conducting analyses. An OR over 1 indicates greater probability of participation at timepoint 2. Only ORs from initial weights are presented here. Only variables with $p<=.05$ are shown in this table. *$p<.05$, **$p<.01$, ***$p<.001$

Overall, participation in FDS was associated with the greatest variability in retention as a function of demographic factors. In FDS, single-parents were 17% more likely to drop out, male caregivers were 25% more likely to drop out, Latinx families were 41% less likely to drop out, the addition of each family member reduced dropout by 8%, each additional year of parent age increased dropout by 1%, and the addition of each stress event in families increased drop out by 19%. In Safe CMP, Latinx families were 99% more likely to drop out. Finally, in PSSF, each additional year of child age reduced dropout by 9% and the addition of each stress event decreased dropout by 40%.

Family Development Outcomes

The needs of participants enrolled in programs significantly differed at baseline (*Table 7*). Programs also differed significantly in the characteristics of participants served

at baseline (Appendix D *Tables 1-3*). Of self-sufficiency items, economic self-sufficiency and debt had the lowest means across programs, indicating high level of economic need and rising debt. Families enrolled in CCR and Safe CMP had lower baseline self-sufficiency scores than in HS, FDS, and PSSF. Families enrolled in CCR also had lower baseline debt scores, indicating greater disparities in financial well-being, than in HS, FDS, and PSSF. Of family protective factors, "knowing what to do as a parent", an indicator of parenting skill, had the lowest means across programs. Families enrolled in FDS, CCR, and Safe CMP had lower baseline "knowledge of what to do as a parent" than in HS.

Preliminary correlations between subgroups and family development outcomes in the whole sample (*Table 8*) revealed that a higher number of high stress events, lower levels of parent education, and lower levels of income were associated with greater gain in an array of outcome areas. Other subgroup characteristics were less frequently associated with outcomes. Caregiver reported male gender was linked with less gain in one outcome area, losing control when disciplining. Higher baseline scores were consistently linked with less change and higher readiness to change was linked with greater improvements across most outcomes. Greater length of time between assessments was linked to both greater and lesser improvement in several areas.

Table 7. Baseline Needs by Family Support Program

	Program					
Family Development Outcomes	**FDS** [a] M (SE)	**HS** [b] M (SE)	**CCR** [c] M (SE)	**PSSF** [d] M (SE)	**Safe CMP** [e] M (SE)	*F value*
Economic Self-Sufficiency	2.94 (0.2)^{bcde}	3.09 (.03)^{acde}	2.83 (.02)^{abd}	3.31 (.04)^{abce}	2.60 (.09)^{abd}	38.93***
Mental/Physical Health	4.11 (.02)^{bce}	4.34 (.04)^{ace}	3.77 (.03) ^{abd}	4.17 (.05)^{abce}	3.71 (.12)^{abd}	41.16***
Debt Management	2.86 (.03)^{cd}	2.84 (.06)^{cd}	2.62 (.05)^{abd}	3.17 (.08)^{abce}	2.61 (.18)^{d}	10.77***
Child Care	3.62 (0.03)^{be}	4.03 (0.06)^{acde}	3.50 (0.04)^{be}	3.71 (0.072)^{be}	2.211(0.17)^{abcd}	31.46***
Child Education	3.86 (0.02)^{bce}	4.30 (0.05)^{acde}	3.53 (0.04)^{abd}	3.86 (0.06)^{bce}	3.41 (0.15)^{abd}	40.25***
Substance Abuse	4.77 (0.02)^{cd}	4.85 (0.03)^{cd}	4.61 (0.02)^{ab}	4.60 (0.04)^{ab}	4.82 (0.10)	14.64***
Family Functioning/Resilience	4.55 (0.02)^{c}	4.66 (0.05)^{e}	4.35 (0.03)^{ab}	4.49 (0.06)	4.21 (0.14)^{b}	9.91***
Social Support	4.52 (0.03)^{bc}	5.00 (0.06)^{ac}	4.35 (0.05)^{abd}	4.73 (0.08)^{ce}	4.14 (0.18)^{bd}	20.41***
Concrete Support	4.10 (0.03)^{bd}	4.78 (0.07)^{ace}	4.06 (0.05)^{bd}	4.63 (0.084)^{ace}	3.68 (0.20)^{bd}	30.24***
Nurturing/Attachment	5.32 (0.02)^{bcd}	5.55 (0.04)^{acd}	5.06 (0.03)^{abe}	5.14 (0.04)^{ab}	5.37 (0.10)^{c}	38.88***
Knowledge of Child Development/ Parenting Skill						
Question 1: Know what to do	3.52 (0.04)^{b}	3.99 (0.08)^{acde}	3.33 (0.06)^{b}	3.49 (0.10)^{b}	3.21 (0.23)^{b}	11.24***
Question 2: Help child learn	4.72 (0.03)^{bc}	4.95 (0.06)^{ac}	4.53 (0.05)^{ab}	4.68 (0.08)	4.85 (0.18)	7.81***
Question 3: Misbehavior to upset	4.20 (0.04)^{c}	4.34 (0.08)^{c}	3.87 (0.06)^{ab}	3.98 (0.10)^{b}	3.69 (0.23)	9.01***
Question 4: Praise for good behavior	5.21 (0.02)^{b}	5.50 (0.05)^{acd}	5.14 (0.03)^{b}	5.13 (0.05)^{b}	5.25 (0.13)	12.18***
Question 5: Do not lose control	5.01 (0.03)^{b}	5.30 (0.06)^{acd}	4.98 (0.04)^{b}	5.05 (0.07)^{b}	5.07 (0.16)	6.62***

Note: Lower baseline scores indicate greater need. Subscripts indicates a significant difference (p<.05) between program means.
**p<.05, **p<.01, ***p<.001*

Table 8. Correlations among Subgroup Predictors and Family Development Outcomes

Family Development Outcomes	Strss	Par edu	Inc	Hsld Sz	Sing par	P Age	Male	ESL	BIPOC	Ltx	Stg scr	Cht	Wks btw	RtC
Self-Sufficiency	.19**	-.12**	-.14**	-.03**	.07**	-.00	-.01	-.02	.03*	-.03*	-.36**	.00	-.04**	.10**
Health	.06**	-.01	.00	.01	.03	.04**	-.01	-.03*	.00	-.01	-.45**	.01	-.01	.08**
Debt Management	.06**	-.00	-.03**	.03**	.03	.02	-.01	-.03*	.01	-.01	-.46**	.01	-.02*	.13**
Child Care	.11**	-.04**	-.04**	-.02	.03	-.02	.01	-.03	.01	-.00	-.54**	.00	-.05**	.24**
Child Education	.03**	-.02*	-.01	-.01	.00	.05**	.01	-.04**	-.03*	-.02	-.52**	.01	-.01	.11**
Substance Abuse	.12**	-.03**	.00	-.01	-.03	-.00	.01	-.02	-.02	.02	-.53**	-.01	.00	.19**
Functioning/Resilience	.05**	-.03*	-.03**	-.01	.04*	.00	-.01	.00	-.05**	.02	-.46**	.02	-.03*	--
Social Support	.07**	-.04**	-.04**	.03*	.03	.02	-.03	.01	.00	.03*	-.55**	.03*	-.01	.07**
Concrete Support	.03*	.00	-.02	.02	.04*	.01	-.01	-.01	-.04**	-.00	-.56**	.02	.02	--
Nurturing/Attachment	.05**	-.01	-.01	.02	.02	.01	.01	-.02	-.00	-.03*	-.46**	.01	-.00	.03*
Knowledge of Child Development/Parenting Skills														
Q1: Know what to do	.02*	.02	-.00	-.00	.01	.00	-.01	.04**	.01	-.01	-.51**	-.01	.05**	.08**
Q2: Help child learn	.01	-.01	-.01	.00	-.02	.01	.01	-.01	-.03*	-.02	-.58**	.00	.03*	.07**
Q3: Misbehavior	.02*	.01	-.03**	.01	.02	.02	-.02	-.01	-.02	-.02	-.51**	-.00	.01	.02
Q4: Praise behavior	.01	-.01	-.01	-.00	.02	-.01	.01	.01	-.02	-.01	-.52**	.00	.02	.01
Q5: Do not lose control	.04**	.00	-.02	-.00	.03	-.01	-.04**	-.02	.01	.00	-.56**	.02	.02	.04**

Note: Strss=High Stress Index, Par edu=parent education, Inc=income, Hsld Sz=household size, Sing par=single-parent household, C Age=Caregiver age, Male=Male caregiver gender, Ltx=Latinx, Stg Scr=starting score, Wks btw=weeks between timepoint 1 and 2, Cht=cohort year, RtC=readiness to change. Readiness to change variables: Nurturing/Attachment, Q1, Q4, and Q5 use readiness to change in parenting skills while Q2 and Q3 use readiness to change in knowledge of child development. Child care and child education coefficients are provided with nonimputed data. *$p<.05$, **$p<.01$, ***$p<.001$

Self-Sufficiency. There was a main effect of program on self-sufficiency improvement (*Figure 1*), $F(4,6967)$=[Range=3.9-5.8], [p<.01-p<.001], [ηp2=.002-.003], such that families enrolled in Head Start (M=.23, SE=.02) and CCR (M=.19, SE=.02) had greater pre-post improvements than families enrolled in FDS (M=.13, SE=.01) and PSSF (M=.13, SE=.02). There were main effects of parent education, $F(1, 6967)$=[Range=5.4-10.1], [p<.05-p<.01], ηp2=.001, and income, F (1, 6967)=[175.5-203.9], p<.001, ηp2=.03, such that greater levels of income and education were linked with greater pre-post improvements. There was also a main effect of starting score, F (1, 6967)=[904.3-954.7], p<.001, ηp2=.12, such that higher starting scores were linked with fewer pre-post gains. Finally, there was a marginal effect of readiness to change (Appendix C).

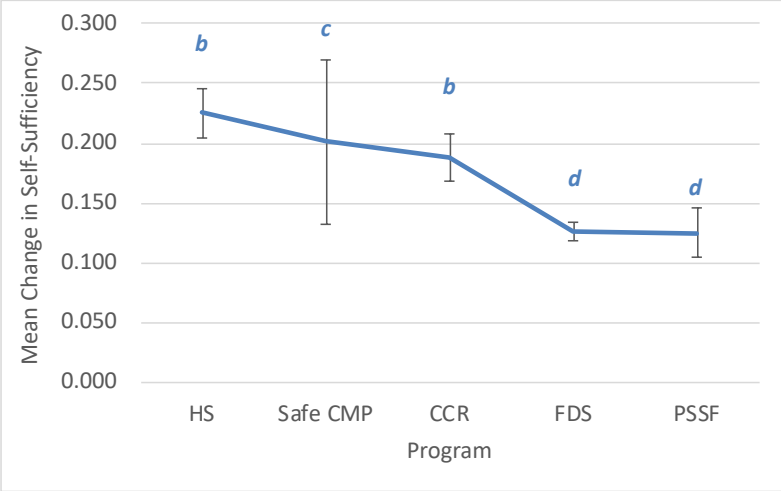


Figure 1. Mean Self-Sufficiency Improvements by Program

Note: a=significantly higher (p<.05) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three or more programs.

A series of interactions was identified between program and subgroups (see *Table 9* and Appendix E *Figures 1a-1b*).

Table 9. Characteristics Associated with Self-Sufficiency Improvement

	Education			BIPOC	
Order	**Low**	**Med**	**High**	**White**	**BIPOC**
Program #1	Safe	Safe	HS*	HS*	Safe
Program #2	PSSF	HS	CCR*	CCR	HS*
Program #3	HS	CCR	FDS	PSSF	CCR
Program #4	CCR	PSSF	Safe	FDS	FDS
Program #5	FDS	FDS	PSSF	Safe	PSSF

Note: Safe=Safe CMP; Education, Low=less than a high-school education, Medium=High School Diploma or GED, Higher=professional certification or college degree (associates or above). =significantly higher ($p<.05$) than three or more programs, =significantly higher than one program, * =significantly higher than two programs, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, =significantly lower than three or more programs, =participants had a negative change score.

An interaction was identified between program and parent education, (4, 6967)=[3.7-5.1], [$p<.01$- $p<.001$], [ηp2=.002-.003], such that at lower levels of education, families enrolled in Safe CMP (M=.24, SE=.08) and PSSF (M=.12, SE=.03) had greater pre-post gains than families enrolled in FDS (M=.03, SE=.01); whereas, at higher levels of education families enrolled in Safe CMP did not differ from any other program (M=.17, SE=.11) and families enrolled in HS (M=.31, SE=.03), CCR (M=.26, SE=.03), and FDS (M=.19, SE=.01) had greater gains than families enrolled in PSSF (M=.13, SE=.03).

Finally, a marginal interaction was observed between BIPOC status and program, F (4, 6967)=[1.2-3.8] [$p=.32$-$p<.05$], [ηp2=.001-.002], such that at in White families enrollment in CCR (M=.18, SE=.01) was associated with greater pre-post gains than FDS (M=.13, SE=.01), but this effect was not present for BIPOC families (M=.19, SE=.04). Additionally, in White families enrollment in HS (M=.20, SE=.02) was linked with greater pre-post gain than both FDS and Safe CMP (M=.08, SE=.06); whereas, in BIPOC

families enrollment in HS (M=.25, SE=.03) was linked with no differences from Safe CMP (M=32, SE=.11) and to greater gains than families enrolled in PSF (M=.10 SE=.04).

Mental and Physical Health. There was a main effect of English as a Second Language (ESL), $F(1,6661)$=[Range=15-17.9], $p<.001$, [ηp2=.002-.003], such that ESL families made great improvements on average (M=.27, SE=.02) than English as a Primarily Language (EPL) families (M=.17, SE=.01). There was also a main effect of starting score, $F(1, 6661)$=[Range=1762-1824], $p<.001$, [ηp2=.21-.22], such that participants with higher starting scores had fewer pre-post gains.

Child Care. There was a main effect of program (*Figure 2*), $F(4,5584)=4.7$, $p<.01$, ηp2=.003, such that families who participated in CCR (M=.40, SE=.03) had greater improvements in access to stable high-quality childcare than families who participated in PSSF (M=.25, SE=.04), FDS (M=.24, SE=.02), and Safe CMP (M=-.04, SE=.10); in contrast. There was also a main effect of starting score,

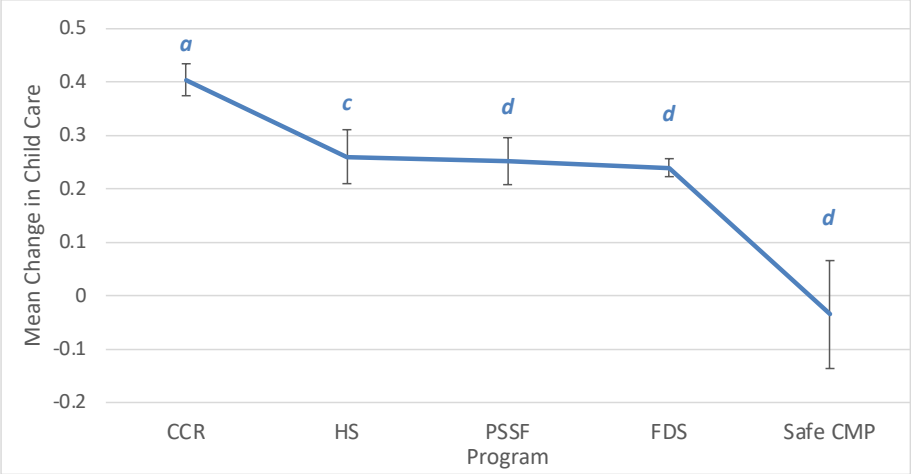


Figure 2. Mean Child Care Improvements by Program

Note: a=significantly higher ($p<.05$) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three or more programs.

F(1, 5584)=2144, *p*<.001, ηp2=.28, such that participants with higher starting scores had lower improvements.

An interaction was observed (see *Table 10* and Appendix E *Figures 2a-2b*) between program and family stress, *F*(4, 5584)=3.7., *p*<.01, ηp2=.003, such that at low levels of stress families that participated in HS had greater improvements in access to high quality child care (M=.34, SE=.06) than Safe CMP (M=-.05, SE=.12), whereas at medium and higher levels of stress, participation in HS was linked with less improvement (M=.23, SE=.06 and M=.12, SE=.10, respectively) than CCR (M=.41, SE=.03 and M=.44, SE=.04) and no significant differences from participation in Safe CMP (M=-.03, SE=.12 and M=-.02, SE=23).

Table 10. Characteristics Associated with Child Care Improvement

	Stress			Income		
Order	**Low**	**Med**	**High**	**Low**	**Medium**	**High**
Program #1	CCR*	CCR	CCR	CCR*	CCR	CCR*
Program #2	HS	PSSF	PSSF	HS	HS	PSSF
Program #3	FDS	HS	FDS	PSSF	PSSF	FDS
Program #4	PSSF	FDS	HS	FDS	FDS	HS
Program #5	Safe	Safe	Safe	Safe	Safe	Safe

Note: Safe=Safe CMP; Stress, Low=0 experiences, Med=1 experience, High=2 experiences; ; Income, Low=$0, Medium=$18,000, Higher=$30,000; =significantly higher (p<.05) than three or more programs, =significantly higher than one programs, * =significantly higher than two programs, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score

An additional marginal interaction was identified between program and family income, *F*(4, 5584)=2.2, *p*<.10, ηp2=.002, such that participation in Safe CMP was linked with adverse pre-post gain at moderate (M=-.13, SE=.09) and higher (M=-.06, SE=.12) levels

of income, significantly less gain than families who had enrolled in CCR (M=.40, SE=.03) and HS (M=.28, SE=.05) at moderate income and CCR at higher income (M=41, SE=.03). However, at low levels of income, enrollment in Safe CMP was not linked to adverse change (M=.06, SE=.14) or to significant differences from any other program.

Child Education. There were main effects of program on improvements in child education (*Figure 3*), $F(4,2775)=5.9$, $p<.001$, $\eta p2=.008$, such that families who participated in HS (M=.40, SE=.06) had greater improvements than those who participated in CCR (M=.09, SE=.06) and PSSF (M=.05, SE=.05). Main effects of stress, $F(1,2775)=16.2$, $p<.001$, $\eta p2=.006$, and starting score, $F(1,2775)=1038.8$, $p<.001$, $\eta p2=.27$, were also found such that higher starting scores and stress were linked with fewer improvements. Main effects of ESL, $F(1,2775)=13.6$, $p<.001$, $\eta p2=.005$, and BIPOC families, $F(1,2775)=4.2$, $p<.05$, $\eta p2=.002$, such that ESL (M=.44, SE=.08) and BIPOC (M=.40, SE=.10) made greater gains than English as Primary Language (EPL)

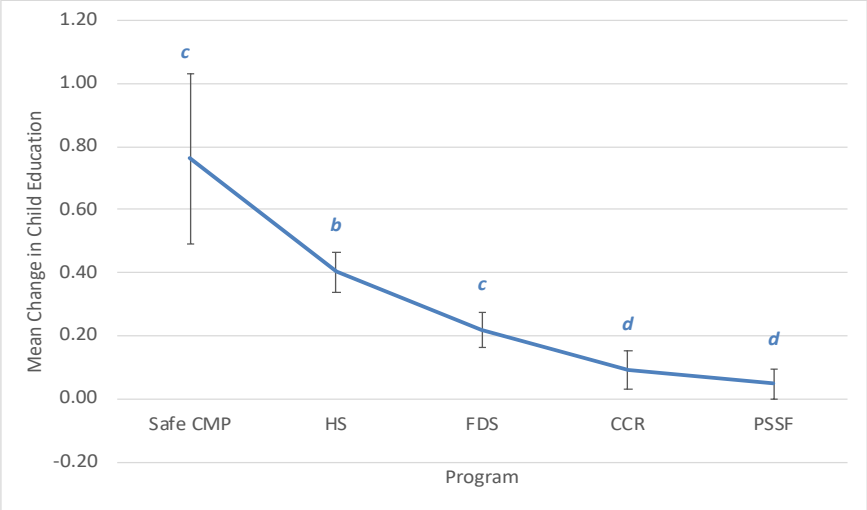


Figure 3. Mean Child Education Improvements by Program

Note: a=significantly higher ($p<.05$) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three or

families (M=.17, SE=.05) and White families (M=.21, SE=.04). A marginal effect of parent education was additionally found (Appendix C).

Table 11. Characteristics Associated with Child Education Improvements

	Stress			Education			ESL		BIPOC	
Order	**Low**	**Med**	**High**	**Low**	**Med**	**High**	**EPL**	**ESL**	**White**	**BIPOC**
Program #1	Safe*	Safe	HS*	HS	Safe	Safe	HS*	Safe*	HS	Safe*
Program #2	HS	HS*	FDS	PSSF	HS	HS	Safe	HS	FDS	HS*
Program #3	PSSF	FDS	Safe	FDS	FDS	FDS	PSSF	FDS	CCR	FDS
Program #4	FDS	CCR	CCR	Safe	PSSF	CCR	FDS	CCR	PSSF	CCR
Program #5	CCR	PSSF	PSSF	CCR	CCR	PSSF	CCR	PSSF	Safe	PSSF

Note: Safe=Safe CMP; Education, Low=less than a high-school education, Medium=High School Diploma or GED, Higher=professional certification or college degree (associates or above); EPL=English is Primary Language, ESL=English is Second Language; BIPOC=Black, Indigenous, People of Color; Stress, Low=0 events, Med=1 event, High=2 events. =significantly higher (p<.05) than three or more programs, * =significantly higher than two programs, =significantly higher than one programs, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score.

Significant interactions were identified between program and ESL status, parent education, BIPOC status, and stress on improvements in child education (see *Table 11* and Appendix E *Figures 3a-3d*). First, an interaction was found between program and ESL status, $F(4,2755)=6.3$, $p<.001$, $\eta p2=.009$, such that ESL families that participated in PSSF (M=-.06, SE=.08) had lower pre-post gains than families enrolled in Safe CMP (M=1.31, SE=.36), HS (M=.51, SE=.12), and FDS (M=.30, SE=.10); whereas, EPL families who participated in PSSF (M=.15, SE=.05) had no differences from families who participated in other programs. There was an interaction between program and parent education, $F(4,2755)=9.9$, $p<.001$, $\eta p2=.01$, such that only at higher levels of education families who participated in Safe CMP (M=1.23, SE=.41) and FDS (M=.28, SE=.06) had greater pre-post improvements than families in PSSF (M=-.14, SE=.06). At

low levels of education, enrollment in HS (M=.38, SE=.09) and PSSF (M=.33, SE=.07) were linked with greater pre-post gains than CCR (M=.03, SE=.07).

There was an interaction between program and BIPOC status, $F(4,2775)=3.3$, $p<.05$, ηp2=.005, such that BIPOC families who participated in Safe CMP (M=1.45, SE=.45) had greater improvements than those who participated in CCR (M=.01, SE=.09), while White families had no significant differences between enrollment in Safe CMP (M=.08, SE=.19) and CCR (M=.17, SE=.05). Lastly, an interaction was found between program and stress, $F(4,2755)=9.5$, $p<.001$, ηp2=.01, such that only at high levels of stress, families enrolled in CCR (M=.05, SE=.07) had greater pre-post improvement than PSSF (M=-.33, SE=.08), where pre-post changes were negative on average. Additionally, at low levels of stress, enrollment in Safe CMP (M=1.07, SE=.29) was linked with greater improvements than enrollment in FDS (M=.22, SE=.05) and CCR (M=.12, SE=.07); while at greater levels of stress Safe CMP was not linked to differences from any program (medium stress, M=.63, SE=.28, higher stress, M=.19, SE=.36).

Debt. A man effect of program was found (*Figure 4*), $F(1,6630)$=[Range=3.4-5.2], $p<.001$, [ηp2=.002-.003], such that enrollment in CCR (M=.38, SE=.04), HS (M=.37, SE=.09), and FDS (M=.31, SE=.03) were linked with greater improvements than in PSSF (M=.16, SE=.05) and Safe CMP (M=-.18, SE=.12), though enrollment in Safe CMP was also linked with lower gains than PSSF.

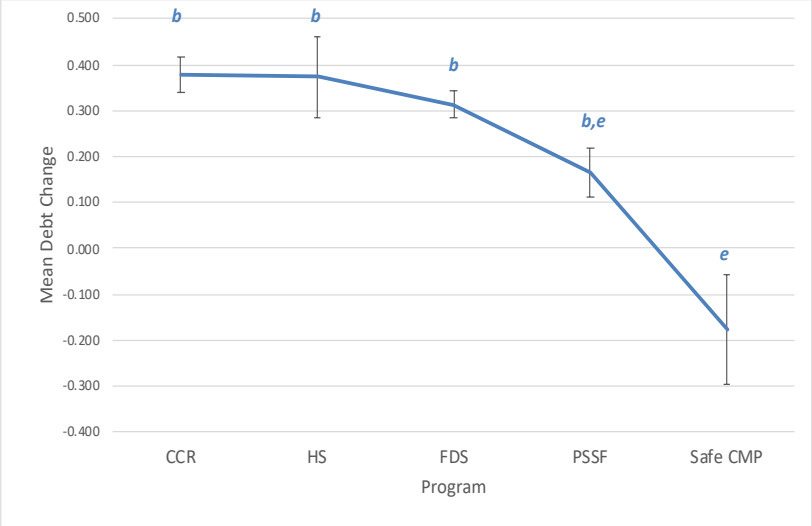


Figure 4. Mean Debt Improvements by Program

Note: a=significantly higher (p<.05) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three

A main effect of ESL was found, F(1,6630)=[Range=3.9-6.9], [p<.05-p<.01], ηp2=.001, such that ESL families made greater gains (M=.28, SE=.06) than EPL families (M=.14, SE=.03). Finally, a main effect of starting score was found, F(1,6630)=[Range=1768.8-1919.4], p<.001, [ηp2=.22-.23], such that higher starting scores were associated with fewer improvements and higher income was associated with greater improvements. A marginal interaction was observed between program and ESL (see *Table 12* and Appendix E *Figure 4a*), F(1,6630)=[Range=1.5-2.9], [p=.21-p<.05], [ηp2=.001-.002], such that EPL families enrolled in CCR (M=.26, SE=.04) made greater gains than families in FDS (M=.16, SE=.04), but ESL families enrolled in CCR (M=.54, SE=.08) did not differ from those enrolled in FDS (M=.48, SE=.06). Additionally, enrollment in FDS was linked with greater gains than PSSF in ESL families (FDS, M=16, SE=.06, M=.09, SE=.06) but not EPL families (FDS, M=.48, SE=.06, PSSF, M=.13, SE=.11).

Table 12. Characteristics Associated with Debt Improvement

Order	Family Language	
	EPL	ESL
Program #1	HS	CCR
Program #2	CCR	HS
Program #3	FDS	FDS
Program #4	PSSF	PSSF
Program #5	Safe CMP	Safe CMP

Note: EPL=English is Primary Language, ESL=English is Second Language. =significantly higher ($p<.05$) than three or more programs, * =significantly higher than two programs, =significantly higher than one programs, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score.

Substance Abuse. A main effect of program was found on improvements in substance abuse (*Figure 5*), $F(4,7613)=[4.3\text{-}5.5]$, $p<.01$, $\eta p2=.002$, such that families who participated in CCR (M=.27, SE=.08) had greater improvements than families who participated in HS (M=.06, SE=.03) and FDS (M=.04, SE=.01). A main effect of stress, $F(1, 7613)=[10.4\text{-}12.6]$, [$p<.01$- $p<.001$], [$\eta p2=.001\text{-}.002$], and of starting score were also observed, $F(1,7613)=[2216.4\text{-}2531.8]$, $p<.001$, [$\eta p2=.23\text{-}.25$] such that increased stress and starting scores were

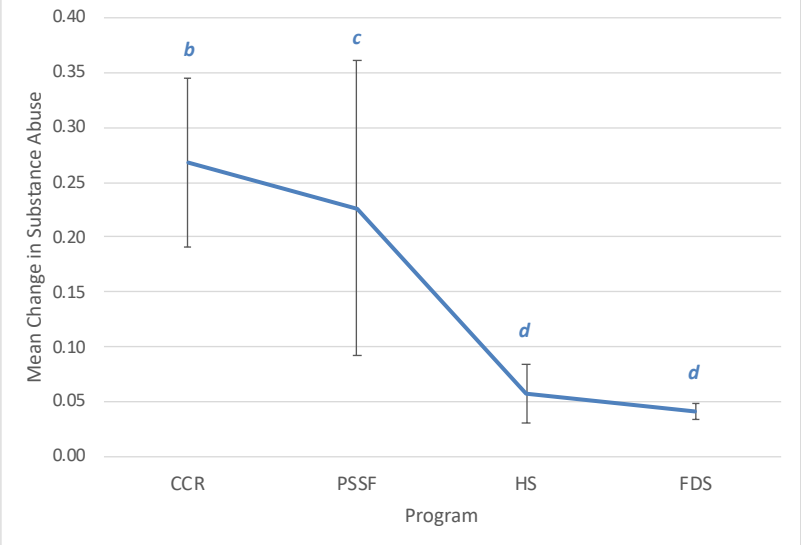


Figure 5. Mean Substance Abuse Improvement by Program

Note: a=significantly higher ($p<.05$) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three or more programs.

linked with fewer improvements. Readiness to change also predicted pre-post change, $F(1, 7613)=[23.3\text{-}46.9]$, $p<.001$, [ηp2=.003-.006], such that higher reported readiness to change scores was linked with fewer gains. An interaction was found between stress and program (see *Table 13* and Appendix E *Figure 5a*), $F(3,7613)=[6.4\text{-}7.3]$, $p<.001$, ηp2=.003, such that the relatively greater improvements made by families in CCR applied when there were fewer stress events in the family. At 0 stress events, enrollment in CCR was linked with greater gains than HS (HS, M=.13, SE=.04) and FDS (M=.06, SE=.01). At higher levels of stress (2, 3, or 4 events), all participants had decreasing change scores and had significantly lower pre-post gains than families enrolled in PSSF (2 events: M=.31, SE=.11, 3 events: M=.37, SE=.12, 4 events: M=.43, SE=.17). Of families with greater levels of stress (3 and 4 events), families who participated in FDS also had fewer pre-post negative gains (3 events: M=-.04, SE=.03, 4 events: M=-.07, SE=.03) than those that participated in HS (3 events: M=-.25, SE=.08, 4 events: M=-.36, SE=.11) or CCR (4 events: M=-.42, SE=.14). A secondary model was also recalculated

Table 13. Characteristics Associated with Substance Abuse Improvement

	High Stress Events				
Order	**0 Events**	**1 Event**	**2 Events**	**3 Event**	**4 Events**
Program #1	CCR*	PSSF	PSSF	PSSF	PSSF
Program #2	PSSF	CCR*	FDS	FDS	FDS*
Program #3	HS	FDS	CCR	CCR	HS
Program #4	FDS	HS	HS	HS	CCR

Note: Safe=Safe CMP; =significantly higher (p<.05) than three or more programs, * =significantly higher than two programs, =significantly higher than one program, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score.

without readiness to change and with the inclusion of Safe CMP. This model revealed a similar pattern of effects, with main effects of program and stress (Appendix C), although the interaction between program and stress was no longer significant.

Family Functioning and Resilience. A main effect of minority race was observed $F(1,7311)=[16.4\text{-}46.3]$, $p<.001$, [ηp2=.002-.006], such that BIPOC families had lower pre-post gains (M=.03, SE=.03) than White families (M=.16, SE=.01). A main effect of stress was also found, $F(1,7311)=[13.8\text{-}30.9]$, $p<.001$, [ηp2=.002-.004], such that greater levels of stress were linked with fewer gains. Finally, a main effect of starting score, $F(4, 7311)=[1254.5\text{-}1726.4]$, $p<.001$, [ηp2=.18-.19], was observed, such that higher scores were linked with less improvement. A marginal main effect of household structure was found (Appendix C). Finally, a marginal interaction was observed between household structure and program (*Table 14* and Appendix E *Figure 6a*), $F(4, 7311)=[1.6\text{-}3.7]$, [$p$=.18-$p$<.01], [ηp2=.002-.005], such that in both two and single-parent households, families enrolled in PSSF (two-parent: M=-.36, SE=.14, single-parent: M=-.30, SE=.19) had lower scores than all other programs except FDS in single-parent homes (M=.05, SE=.04). Additionally, single-parent households enrolled in FDS were linked with lower change scores than Safe CMP (M=.62, SE=.20) and CCR (M=26, SE=.06), but this relationship was not observed in two-parent households.

Table 14. Characteristics Associated with Family Functioning and Resilience Improvement

Order	Household Structure	
	Two-Parent	**Single-Parent**
Program #1	HS*	Safe CMP*
Program #2	Safe	CCR*
Program #3	CCR	HS*
Program #4	FDS	FDS
Program #5	PSSF	PSSF

Note: Safe=Safe CMP; Income, Low=$0, Med=$18,000/year, High=$30,000/year; =significantly higher ($p<.05$) than three or more programs, =significantly higher than one program, * =significantly higher than two programs, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score.

Social Support. A main effect of starting score was identified, $F(1,7971)=[2522.1\text{-}3270.8]$, $p<.001$, [ηp2=.28-.29], such that higher starting score was linked with lower improvements. A secondary model was also recalculated including Safe CMP and dropping readiness to change. This model revealed marginal main effects of program and parent education (Appendix C). The secondary model additionally revealed a marginal interaction between program and parent education (*Table 15* and *Appendix E Figure 7a)*, $F(1, 12804)=[1.1\text{-}3.8]$,[$p=.35$-$p<.01$], [ηp2=.000-.001], such that families with greater education enrolled in Safe CMP (M=.55, SE=.13) and HS (M=.53, SE=.07) had greater gains than families enrolled in CCR (M=.26, SE=.05), PSSF (M=.23, SE=.05), or FDS (M=.21, SE=.02); however, this relationship was not present at lower levels of education. Families with less education had greater gains when enrolled in CCR (M=.28, SE=.04) than in FDS (=.13, SE=.02).

Table 15. Characteristics Associated with Social Support Improvement

	Education		
Order	**Low**	**Medium**	**High**
Program #1	Safe	Safe	Safe
Program #2	CCR	HS	HS
Program #3	PSSF	CCR	CCR
Program #4	HS	PSSF	PSSF
Program #5	FDS	FDS	FDS

Note: Safe=Safe CMP; Stress, Low=0, Medium=1, Higher=2; Education, Low=less than a high-school education, Medium=High School Diploma or GED, Higher=professional certification or college degree (associates or above). =significantly higher ($p<.05$) than three or more programs, * =significantly higher than two programs, =significantly higher than one program, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score.

Concrete Support. A main effect of program was found on improvements in concrete support (*Figure 8*), $F(4,7450)=[2.7\text{-}6.1]$, [$p<.05$- $p<.001$], [$\eta p2=.001\text{-}.003$], such that families enrolled in Safe CMP (M=.73, SE=.20), HS (M=.54, SE=.07), and CCR (M=.45, SE.08) had greater improvements than families enrolled in FDS (M=.25, SE=.03). Enrollment in Safe CMP and HS was additionally associated with greater gain than enrollment in in PSSF (M=.18, SE=.15). A main effect of starting score was identified, $F(1, 7450)=[3175.7\text{-}4139.6]$, $p<.001$, [$\eta p2=.33\text{-}.36$], such that higher starting score was linked with lower improvements.

A series of interactions (*Table 16* and Appendix E *Figures 8a-8c)* between predictor and program were identified. First, program interacted with family stress events, $F(4, 3776=[3.3\text{-}10.9]$, [$p<.05$-$p<.001$], [$\eta p2=.002\text{-}.006$], such that at higher levels of stress,

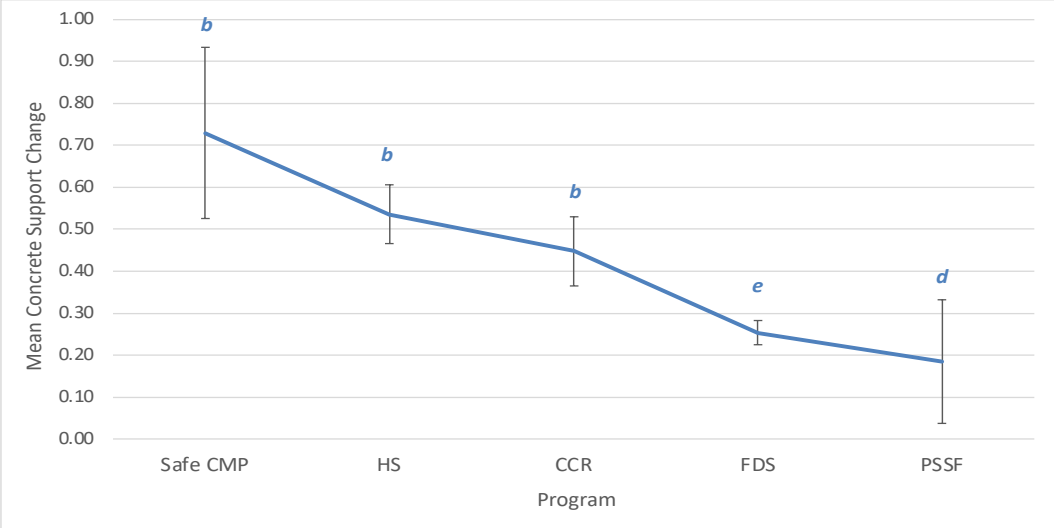


Figure 6. Mean Concrete Support Improvements by Program

Note: a=significantly higher (p<.05) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three or more programs.

families enrolled in PSSF (M=.-.14, SE=.17) exhibited less pre-post gain than families in all other programs (Safe CMP, M=.73, SE=.34, HS, M=.70, SE=.12, CCR, M=.41, SE=.09, FDS, M=.26, SE=.05). In contrast, at low levels of stress enrollment in PSSF (M=.33, SE=.16) did not differ from other programs. Additionally, the gains associated with enrollment in HS increased with rising stress, such that enrollment when stress was low (M=.46, SE=.09) was associated with greater gain than one program (FDS, M=.25, SE=.03), but with greater gain than three programs (CCR, FDS, PSSF) when stress was higher. Second, program interacted with minority status, $F(4, 7450)=[6.7\text{-}16.5]$, $p<.001$, [ηp2=.004-.009], such that White families enrolled in Safe CMP (M=-.04, SE=.17) had fewer improvements in concrete supports than all other programs (HS, M=.67, SE=.08, PSSF, M=.50, SE=.08, CCR, M=.45, SE=.06, FDS, M=.41, SE=.02); whereas, BIPOC

families enrolled in Safe CMP (M=1.50, SE=.37) had greater gains than families enrolled in all other programs (CCR, M=.45, SE=.14, HS, M=.40, SE=.12, FDS, M=.10, SE=.05, PSSF, M=-.14, SE=.26).

Table 16. Characteristics Associated with Concrete Support Improvement

	Stress			BIPOC		Household Structure	
Order	Low	Med	High	White	BIPOC	Two-parent	Single-parent
Program #1	Safe	Safe*	Safe	HS	Safe	PSSF	Safe
Program #2	CCR	HS*	HS	PSSF	CCR*	HS	CCR
Program #3	HS	CCR*	CCR	CCR	HS	CCR	HS
Program #4	PSSF	FDS	FDS	FDS	FDS	FDS	FDS
Program #5	FDS	PSSF	PSSF	Safe	PSSF	Safe	PSSF

Note: Safe=Safe CMP; Stress, Low=0, Medium=1, Higher=2; Income, Low=$0, Med=$18,000/year, High=$30,000/year. =significantly higher (p<.05) than three or more programs, * =significantly higher than two programs, =significantly higher than one program, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score

Finally, a marginal interaction was identified in a subsample between program and household structure, $F(4,2903)=[1.5\text{-}14.2]$, [$p=.20$-$p<.001$], [$\eta p2=.002\text{-}.02$], such that no significant differences existed between two-parent households, but in single-parent households enrollment in HS (M=.58, SE=.09) was linked with greater gain than enrollment in FDS (M=.37, SE=.06).

Nurturing and Attachment. A main effect of starting score was identified, $F(1,9383)=[1649.61\text{-}2553.4]$, $p<.001$, [$\eta p2=.19\text{-}.21$], such that higher starting score was linked with lower improvements. Additionally, a marginal main effect of program was identified (Appendix C). A marginal interaction (*Table 17* and *Appendix E Figure 9a*) was identified between program and ethnicity, $F(4,9383)=[2.1\text{-}10.9]$, [$p<.10$-$p<.001$],

[ηp2=.001-.005], such that Latinx families enrolled in PSSF (M=-.12, SE=.04) had significant declines in pre-post scores, significantly lower than families enrolled in all other programs (HS, M=09, Se=.04, FDS, M=.08, SE=.01, Safe CMP, M=.08, SE=.08, CCR, M=.06, SE=.03). In contrast, there were no differences between non-Latinx families enrolled in PSSF from other programs.

Table 17. Characteristics Associated with Nurturing and Attachment Improvement

Order	Ethnicity	
	Non-Latinx	Latinx
Program #1	HS	HS
Program #2	Safe CMP	FDS
Program #3	PSSF	Safe CMP
Program #4	FDS	CCR
Program #5	CCR	PSSF

Note: Safe=Safe CMP; =significantly higher ($p<.05$) than three or more programs, , * =significantly higher than two programs, =significantly higher than one program, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score

Child Development and Knowledge of Parenting. This construct was assessed via improvements in five standalone questions, each assessed in turn.

Question 1. I know what to do as a parent. A marginal main effect of program was identified (Appendix C). A main effect of starting score was also identified, $F(1,13094)=[3178.6-4600.4]$, $p<.001$, ηp2=.26, such that higher starting score was linked with lower improvements. A marginal interaction between program and stress was identified (*Table 18* and *Appendix E Figure 10a)*, $F(1,13094)=[1.5-4.4]$,[$p=.21$-$p<.001$], ηp2=.001, such that at low stress enrollment in CCR (M=.43, SE=.05) and at moderate

stress enrollment in PSSF (M=.47, SE=.08) were linked with greater improvements than FDS (low stress, M=.28, S=.03; moderate stress, M=.27, SE=.03). At higher levels of stress, there were no differences between programs.

Table 18. Characteristics Associated with Question 1 Improvements

	Stress		
Order	**Low**	**Medium**	**High**
Program #1	CCR	Safe	Safe
Program #2	PSSF	PSSF	PSSF
Program #3	Safe	HS	HS
Program #4	FDS	CCR	FDS
Program #5	HS	FDS	CCR

Note: Safe=Safe CMP; =significantly higher ($p<.05$) than three or more programs, * =significantly higher than two programs, =significantly higher than one program, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score

Question 2. I know how to help my child learn. A main effect of program was identified (see *Table 21*), $F(4,7369)=[4.9\text{-}11.3]$, [$p<.01$-$p<.001$], [ηp2=.003-.006], such that families enrolled in Safe CMP (M=.58, SE=.19) and HS (M=.40, SE=.07) had significantly greater improvements than families enrolled in FDS (M=.13, SE=.03) and PSSF (M=.02, SE=.13). A main effect of length of time between CFSAs was identified, $F(4,7369)=[13\text{-}23.7]$, $p<.001$, [ηp2=-.002-.003], such that greater lengths of time were linked with greater improvements. Finally, a main effect of starting score was also identified, $F(1,7369)=[2825.1\text{-}3939.2]$, $p<.001$, [ηp2=.34-.35], such that higher starting score was linked with lower improvements. An interaction was identified between program and minority status (*Table 19* and Appendix E *Figure 11a*), $F(1,7369)=[2.4\text{-}5.1]$, [$p<.05$-$p<.001$], [ηp2=.001-.003], such White families enrolled in Safe CMP

(M=.49, SE=.16) had greater gains than families enrolled in CCR (M=.16, SE=.05). However, in BIPOC families, there were no differences between Safe CMP and other programs. Rather enrollment in HS (M=.49, SE=.11) and CCR (M=.35, SE=.12) were linked with greater gains than enrollment in PSSF (M=-.10, SE=.20).

Table 19. Characteristics Associated with Question 2 Improvements

Order	Race	
	White	BIPOC
Program #1	Safe CMP	Safe
Program #2	HS	HS*
Program #3	FDS	CCR*
Program #4	CCR	FDS
Program #5	PSSF	PSSF

Note: Safe=Safe CMP; BIPOC=Black, Indigenous, or Person of Color. =significantly higher ($p<.05$) than three or more programs, * =significantly higher than two programs, =significantly higher than one program, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score

Question 3. My child does not misbehave to upset me. A main effect of program was identified (see *Table 21*), $F(4,12985)=[6.3\text{-}17.7]$, $p<.001$, [ηp2=.002-.005], such that families enrolled in PSSF (M=.28, SE=.07) and FDS (M=.26, SE=.02) had significantly greater improvements than families enrolled in CCR (M=.07, SE=.05) and Safe CMP (M=-.08, SE=.10), where families demonstrated a slight decline over time. A main effect of starting score was also identified, $F(1,12985)=[3316.4\text{-}4852.1]$, $p<.001$, [ηp2=.26-.27], such that higher starting score was linked with lower improvements. Finally, a main effect of income was found, $F(4,12985)=[8.4\text{-}13.0]$, [$p<.01$-$p<.001$], ηp2=.001, such that greater income was linked with fewer improvements. A marginal main effect of stress was also found (Appendix C). Interactions were identified (see *Table 20* and Appendix E

Figure 12a-12b) between program and income, $F(4,12985)=[8.4\text{-}13.0]$, [$p=<.01$-$p<.001$], ηp2=.001, such that at lower levels of income, families enrolled in HS (M=.36, SE=.10) had significantly greater improvements than families enrolled in CCR (M=.06, SE=.06), an effect that was absent at moderate (HS, M=18, SE=.07) and higher (M=.06, SE=.10) levels of income. At these higher levels of income, HS displayed no difference from any other program. Finally, an interaction between program and stress was identified, $F(4,12985)=[2.4\text{-}5.8]$, [$p=<.01$-$p<.001$], ηp2=.001, such that at low (M=.35, SE=.10) and moderate (M=.25, SE=.07) levels of stress enrollment in PSSF was linked with greater gain than enrollment in Safe CMP (low, M=.02, SE=.11, moderate, M=-.14, SE=.14) and CCR at low levels of stress (M=.04, SE=.07). However, at greater levels of stress there was no difference between PSSF (M=.14, SE=.11) and other programs. Rather,

enrollment in FDS (M=.26, SE=.05) was linked with greater gain than HS (M=-.07, SE=.15) and Safe CMP (M=-.30, SE=27).

Table 20. Characteristics Associated with Question 3 Improvements

	Stress			Income		
Order	**Low**	**Med**	**High**	**Low**	**Medium**	**High**
Program #1	PSSF*	FDS	FDS*	PSSF*	PSSF*	FDS*
Program #2	FDS*	PSSF	CCR	HS	FDS*	PSSF
Program #3	HS	CCR	PSSF	FDS	HS	CCR
Program #4	CCR	HS	HS	CCR	CCR	HS
Program #5	Safe	Safe	Safe	Safe	Safe	Safe

Note: Safe=Safe CMP; Stress, Low=0 experiences, Med=1 experience, High=2 experiences; Income, Low=$0, Medium=$18,000, Higher=$30,000; =significantly higher ($p<.05$) than three or more programs, =significantly higher than one programs, * =significantly higher than two programs, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score

Question 4. I praise my child when he/she behaves well. A main effect of program was identified (see *Table 21*), $F(4,13099)=[6.3\text{-}11.7]$, $p<.001$, [ηp2=.002-.004], such that families enrolled in Safe CMP (M=.26, SE=.06) and HS (M=.22, SE=.04) demonstrated significantly greater improvements than families enrolled in FDS (M=.07, SE=.01), CCR (M=.04, SE=.02), PSSF (M=.04, SE.05). A main effect of starting score was also identified, $F(1,13099)=[3276.2\text{-}5294.9]$, $p<.001$, [ηp2=.26-.29], such that higher starting score was linked with lower improvement.

Question 5. I do not lose control when I discipline my child. A main effect of program was identified (see *Table 21*), $F(4,13094)=[3.7\text{-}11.2]$, [$p<.01$- $p<.001$], [ηp2=.001-.003], such that families enrolled in HS (M=.32, SE=.06) had greater improvements than families enrolled in CCR (M=.13, SE=.04), FDS (M=11, SE=.02), and PSSF (M=-.03, SE=.06). Enrollment in PSSF was linked with a fewer pre-post gains than all other programs, including Safe CMP (M=.25, SE=.08). A main effect of starting score was also identified, $F(1,13094)=[4080.3\text{-}6252.7]$, $p<.001$, [ηp2=.31-.32], such that higher starting score was linked with lower improvement. A marginal main effect of stress score was also identified (Appendix C). Finally, a marginal interaction was identified (*Table 21* and *Appendix E Figure 13*a) between program and stress, $F(1,13094)=[1.0\text{-}6.6]$, [$p=.41$-$p<.001$], [ηp2=.000-.002], such that at high levels of stress enrollment in PSSF (M=-.04, SE=.08) was associated with fewer pre-post gains than all other programs (Safe CMP, M=.43, SE=.19, HS, M-.31, SE=.10, FDS, M=.23, SE=04, CCR, M=.17, SE=.04); whereas, at low levels of stress, Safe CMP (M=.16, SE=.09) was not linked with differences from PSSF or other programs. Additionally, at low levels of

stress, HS (M=.31, SE=.07) was linked with greater gain than CCR (M=.11, SE=.07), FDS (M=.05, SE=.02), and PSSF (M=-.02, SE.07); however, at high levels of stress, enrollment in HS was only linked with greater gain than PSSF.

Table 21. Characteristics Associated with Question 5 Improvements

	Stress		
Order	**Low**	**Medium**	**High**
Program #1	HS	HS	Safe CMP
Program #2	Safe CMP	Safe CMP	HS
Program #3	CCR	FDS	FDS
Program #4	FDS	CCR	CCR
Program #5	PSSF	PSSF	PSSF

Note: Safe=Safe CMP; Stress, Low=0 experiences, Med=1 experience, High=2 experiences. =significantly higher (p<.05) than three or more programs, , * =significantly higher than two programs, =significantly higher than one program, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score

Table 22. Programs Associated with Child Development and Knowledge of Parenting Improvement

	Question			
Order	**Q2. Help child learn**	**Q3. Misbehave to upset**	**Q4. Praise my child**	**Q5. Lose control**
Program #1	Safe*	PSSF*	Safe	HS
Program #2	HS*	FDS*	HS	Safe
Program #3	CCR	HS	FDS	CCR
Program #4	FDS	CCR	CCR	FDS
Program #5	PSSF	Safe	PSSF	PSSF

Note: Safe=Safe CMP; =significantly higher (p<.05) than three or more programs, * =significantly higher than two programs, =significantly higher than one program, =no differences with any programs, =significantly lower than one or more programs, =significantly higher than one or more and significantly lower than one or more programs, = significantly lower than three or more programs, =participants had a negative change score

Discussion

This project focused on identification of "what works for whom" in family support programs. Overall differences in benefit by program and family characteristic were identified and differential patterns of engagement explored for families facing elevated stress or other forms of structural adversity. Broadly, these data support that a variety of families with higher levels of need are at risk for early dropout and inconsistent benefit across programs. Documented patterns in these data are suggestive of relative benefit or disadvantage, but should not be interpreted as causal given the brief longitudinal design utilized, the exploratory nature of subgroup analysis, and the lack of a comparison group. No single program benefited all families consistently over others, although broad themes emerged indicating the types of programs which may differentially support families in conditions of higher stress or disadvantage.

Participant characteristics and engagement

Dropout rates in this sample were varied by program, with the strongest pattern of engagement displayed in HS (20% attrition) and the weakest in PSSF (75%). Drop-out rates in PSSF, while exceedingly high, have been documented at similar levels in some home-visiting evaluations (Holland et al., 2018). Family poverty, family stress, single-

parent status, lower caregiver age, and BIPOC status were all hypothesized to predict earlier dropout across programs based on prior literature (Reyno & McGrath, 2006). While an initial scan revealed support for this hypothesis, associations were not consistent across programs or in their direction of effect. After accounting for all covariates that significantly predicted dropout and baseline family development scores in multivariate analyses, fewer demographic predictors accounted for attrition. Only FDS, with a large sample well-powered to identify small effects, identified the hypothesized relationship with single-parent households and cumulative family stress. Interestingly and unexpectedly, poverty was not associated with increased dropout in any program. This result is surprising, given the consistency with which very low family income has predicted drop out (Reyno & McGrath, 2006). However, Brand and Jungmann (2014) similarly demonstrated that family socioeconomic status became insignificant when examined in a multivariate model accounting for other participant characteristics and service experiences.

While poverty itself did not emerge as significant in these data, cumulative family stress (including no reported income, homelessness, and substance abuse) did relate to dropout in FDS. Given stressors that can cooccur with, cause, or be caused by low income, such as housing instability (Roggman et al., 2008) and substance abuse (Damashek et al., 2011), have been linked previously with program dropout, it may that the contributing factors or consequences of economic strain account for dropout more acutely than level of income itself. Alternately, dropout could relate to the service design of FDS. Services provided by FRCs are primarily center-based (e.g. services to meet

basic needs, parenting education) and these activities bidirectionally connect to FDS (e.g. generating referrals to and from FDS). Although FDS may occur in a variety of settings, including home-based contexts, it may be that families with higher levels of stress have greater difficulty engaging in center-based referrals and resources embedded within FRCs. Additionally, it is unknown how often families were engaged in FDS on site versus in the community. Center-based services have been described as more challenging for families experiencing stress to access and consistently attend (for discussion see Ammerman et al., 2006). Interestingly, the relationship between cumulative stress and dropout was reversed in PSSF, with greater stress predicting greater retention. It may be that PSSF was better able to consistently engage families with elevated stress or that service content or delivery-system systematically differed from other programs. Alternately, the population served by PSSF may systematically differ in such a way that engagement rises with stress (e.g. adoptive or foster-care families, Rees & Selwyn, 2009).

Parent age was associated with program engagement in the FDS program, but in the opposite direction than hypothesized, with younger parents showing a small but significant increased likelihood of participation. In another divergence, FDS appeared to effectively engage Latinx caregivers, who were less likely to stay engaged in Safe CMP. This pattern suggests that FDS differentially engaged young parents and Latinx parents, although there is no clear explanation as to why. Yet, these mixed findings are reflected in the field of home visiting and family support. Both maternal age (Brand & Jungmann, 2014; Holland et al., 2014) and Latinx ethnicity (Ammerman et al., 2006; Daro et al.,

2003; Mauricio et al., 2018) are linked with inconsistent patterns of attrition across home-visiting studies. Understanding and capturing how younger parents and Latinx families were effectively engaged in FDS may support efforts to increase the engagement of families facing greater vulnerability. It may be that unmeasured factors at the program, provider, or community level account for these engagement differences. For example, McCurdy and Daro (2001) argue that parent participation in family support programs is influenced by the synergy of participant, neighborhood, provider, and program-level factors. While speculative, it could be that families in FDS were more likely to have a provider who was racially matched, a factor which predicted greater treatment duration and a higher total number of visits in one home-visiting evaluation (Daro et al., 2003). Alternately, FDS may adapt their approach for Latinx families; for example, one of the FRCs included in this sample had designed its entire program around supporting Latinx families, emphasizing cultural attunement with the Latinx community (The Family Center / La Familia, n.d.). Similarly, other FRCs have demonstrated strong and successful outreach with the Latinx community (Lightburn & Warren-Adamson, 2015). Previous work links cultural adaptations in curriculum for Latinx families (e.g. by incorporating respect for family traditions) to an increase in retention from 65% to 98% (Kumpfer et al., 2002). Although the present study could not isolate particular FRCs from the whole sample or analyze curricula, future work should examine how cultural adaptations and outreach strategies may drive increased engagement. Finally, other work links process and intervention-focused factors (e.g. low maternal engagement during a home visit) to more variance in attrition than participant demographics (Brand &

Jungmann, 2014). Future examinations of differential patterns of engagement should account for the effects of the wider ecological context in which they occur.

Programs and family development outcomes

Among improvements at the full sample level, several programs demonstrated shared benefit. In a variety of other outcome areas, enrollment in multiple programs was linked to differential pre-post gain. For example, self-sufficiency improvements occurred at equivalent rates across HS and CCR and debt improved at equivalent rates across CCR, HS, and FDS. Social support improvements occurred at equivalent levels across Safe CMP and HS (see Appendix D *Table 4*). This finding only partly aligns with hypothesized patterns, where center-based programs were expected to improve social support, potentially via access to peer support. Rather, improvements in social support were also exhibited in a home-based programs (Safe CMP). A variety of other outcomes (concrete support and four child development and knowledge of parenting items) improved at equivalent levels in two FSPs, in all cases revealing improvement in a combination of center, mixed-delivery, and home-based programs. In two instances (mental and physical health and family functioning and resilience), improvements occurred equivalently across all programs. Thus, in some family development areas, improvements appear to uniformly occur across service delivery systems. This finding reflects some work demonstrating that parent-focused outcomes did not vary on average across home-visiting and group-based approaches (Chen & Chan, 2016). Furthermore, the largest effect sizes across these analyses were tied to starting score, indicating that families with greater need in each outcome area made the greatest gains. However,

differential patterns of advantage were identified at the program and subgroup level over and above the impact of baseline need.

Among programs demonstrating improvement at the full sample level, enrollment in HS and CCR demonstrated a trend toward strong improvements on a wider array of outcomes for families on average than other programs (see Appendix D *Table 4*). Families enrolled in HS exhibited greater improvements in 9 out of 15 family development outcomes, while families enrolled in CCR exhibited greater improvements in 5 out of 15 family development outcomes. The overall benefit connected to HS may be partly explained by the nature of the sample. Rather than accounting for implementation variety across sites, the HS in this sample comprises a high-fidelity program at a single site in comparison to a suite of alternate programs offered statewide. A meta-analysis focused on outcomes across diverse home-visiting models identified that single-site interventions tended to have larger effect sizes than multisite interventions (Sweet & Appelbaum, 2004), a difference which may be explained by variation in implementation factors (Casillas et al., 2016). Thus, HS effects in this analysis may not reflect outcomes in HS at a statewide level.

Alternately, the design of family support services at this HS site may explain its benefit. Although most comprehensive family support services and programs were offered on site, families typically also received two home-visits per year. In this way, HS may reap several of the benefits inherent to home-visiting models (e.g. a trusting relationship), while engaging families in a structured set of services available on site. Until future work can compare the impacts of HS on a statewide level or better control

for variation in implementation fidelity across models, these results should be interpreted with caution.

CCR, as a home-visiting model, exhibited slightly more (by outcome count) differential benefits for families overall in comparison to Safe CMP's home-visiting and FDS and PSSFs' mixed delivery systems (see Appendix D *Table 4*). Families enrolled in CCR experienced the greatest improvements in access to child care and benefit in the parenting skill reflecting "knowing what to do as a parent". At the time this data was collected, every CCR case had previously been screened out by CPS. Given that the majority of CPS referrals are tied to concerns about parenting or the home environment (Simon et al., 2021), it may be that improvements in knowing what to do reflected a greater underlying need for parenting knowledge in this population. However, CCR programming does not provide direct parenting education, so these improvements may be leveraged via effective linkages to other community supports.

Every subgroup (with the exception of large families and male caregivers who exhibited no differential benefits in any programs) experienced differential advantage in at least one domain when enrolled in CCR (see Appendix D *Table 5*). Most notable among CCR's unique features are the combination of home-based support with access to tangible support (e.g. immediate cash assistance) that is most typically accessible via structured center-based programs. In this way, CCR reaps some of the benefit of structured services in a format conducive to trust and individualized support for families with higher levels of need (Ammerman et al., 2006). Because Safe CMP includes access to tangible support but does not exhibit a similar overall profile to CCR, it may be that

families who enter services with a primary parenting-focus (via Safe CMP) do not on average experience as many improvements in areas of secondary focus for family stability and wellbeing. In support of this idea, enrollment in Safe CMP was generally linked with greater improvements in two parenting skills, helping children learn and praising positive behavior, and building concrete and social support (Appendix D *Table 4*). A similar pattern of effect is reflected in evaluations of MIHOPE and MIHOPE-Strong Start. In these programs, parenting skills and supports improved, but no positive effects were demonstrated in self-sufficiency or health (Michalopoulos et al., 2019). Conversely, families enrolled in Safe CMP on average exhibited decreased understanding that misbehavior was not designed to upset parents (*Table 17*). It is unclear why families enrolled in Safe SMP exhibited a pattern of both differential pre-to-post gain and loss in various family protective factors.

In contrast to HS and CCR and contrary to hypothesis, FDS and PSSF showed the lowest number of overall improvements, as mixed-delivery programs. Enrollment in FDS, the program implemented at the largest number of sites, was typically related to more moderate improvements overall. FDS displayed limited benefits overall for improvements in self-sufficiency, a surprising finding, given that FRCs act as hubs with access to myriad resources and other interventions on site (Colorado Family Resource Center Association [CFRCA] & OMNI, 2019, 2020). In fact, FRCs provided far more services to meet the basic needs of families than other services such as those focused on parenting, ECE, or adult education (CFRCA & OMNI, 2019, 2020). Yet, these results (an average improvement of .13 points) are consistent with evaluations from prior years

within FRCs, where average self-sufficiency improvements ranged from .13-.15 points between 2019-2020 (all data prior to the COVID pandemic; CFRCA & OMNI, 2019). Thus, it appears that improvements in self-sufficiency for FDS are significantly lower than in other programs in this sample on average. Whether this lower improvement is connected to the much larger population serviced across programs with varied levels of implementation fidelity or to a systematic difference in the services provided through FDS is unclear, but a diluted impact due to population size and diversity as well as variation in implementation fidelity is a possible interpretation. Many programs experience reduced efficacy when taken to scale in this way (Supplee et al., 2021). Another possible explanation for relatively lower improvements may be connected to FRCs role as a hub for connection to other family support programs, such as CCR and PSSF. For example, 15 out of 21 currently contracted CCR programs operate under the umbrella of an FRC (personal correspondence between the statewide program manager in the OEC and the author). This role as a comprehensive service hub reduces the burden on families to identify "right service, right time" within the service landscape array. While speculative, it may be possible that some families with high levels of need are funneled to intensive home-based supports prior to engaging FDS. Alternately, families could have opted into CCR following a child welfare referral, made significant gains in family wellbeing and, to sustain ongoing growth, later enrolled in FDS where longer term support services may be housed. However, the ability to test the effects of concurrent or consecutive service enrollment was constrained by lack of shared identifiers across administrative datasets.

In three instances, FDS emerged as displaying differential benefit (Appendix D *Table 4*): debt improvement, promoting the knowledge that misbehavior is not intentionally designed to upset parents, and, marginally, promoting nurturing and attachment (Appendix C). Enrollment in PSSF, a community-based model with a mixed delivery system focused on family strengthening, stabilization, and preservation, was similarly linked with greater knowledge that child misbehavior is not intended to upset parents and of what to do as a parent (Appendix D *Table 4*). This shared pattern of benefit in FDS and PSSF may indicate enrollment in group-based parenting education classes; however, this should be examined in future studies measuring service activities on a granular level.

Participation in services with PSSF was not associated with differential benefit for BIPOC, Latinx, or single-parent households across any outcomes. Moreover, BIPOC families enrolled in PSSF experienced adverse pre-post change in 'knowing how to help my child learn' and concrete support, Latinx families experienced adverse pre-post change in nurturing and attachment, and single-parents experienced adverse pre-post change in concrete support. Enrollment in PSSF was also, on average, associated with negative change in nurturing and attachment and losing control when disciplining (*Tables 17* and *22*), a direction of effects not seen in other programs. Declining scores should be interpreted with caution, as these could be tied to various issues that are obscured in these data and study design. For example, seemingly negative gain may reflect protection from greater loss and/or greater insight or comfort with disclosure of existing issues by the second assessment timepoint. It is also possible that adverse changes may reflect differences in the population served by PSSF, including adoptive and foster-care families.

It may be possible that families enrolled in PSSF experienced a worsening emotional connection to the children in their care; for example, a significant group of children in foster care exhibit worsening trajectories of mental and behavioral health over time (Tarren-Sweeney & Goemans, 2019). In one study of adopted children, 75% of children had attachment problems and 88% of adoptive families reported seeking support due to the level of difficulty experienced in their first year of placement (Rees & Selwyn, 2009). As an additional factor, subgroup analysis revealed that negative changes were not universal in PSSF. Only Latinx families appeared to experience declining scores in nurturing and attachment and only in PSSF for reasons which are not clear (Appendix E *Figure 9b*). It may be that unmeasured factors, such as participants' cultural perception of services, mismatch with provider language or culture, or concerns regarding immigration status (Doshi et al., 2020) impacted the acceptability of services and parent trust for the provider. For example, low levels of trust at initial engagement may have driven delays in disclosing problems in the parent-child emotional connection or disciplinary practices. It is also possible that engagement with programming content built parent awareness of concerning behaviors or issues in the parent-child relationship to consider and report. Future comparative studies should continue to make efforts to isolate underlying needs, including the cultural needs, of the populations served to better understand apparent outcome disparities.

Participant characteristics and family development outcomes

Within this diverse set of results, several notable patterns emerged. While most family characteristics demonstrated differential patterns of benefit in at least one outcome

area, other characteristics demonstrated less variability in their relations with outcomes across programs. Several characteristics which had been hypothesized to reduce improvements overall, such as younger parent age, larger household size, and fatherhood, as measured by reported caregiver gender, did not emerge as differentially benefited by any program following multivariate analyses. These subgroups had fewer preliminary associations with outcomes. For example, only one outcome area indicated an association with male caregivers and outcomes; that is, less improvement in the parenting skill of losing control when disciplining. The overall lack of fatherhood-related disparities identified here may reflect literature on fatherhood interventions. Similar to the small and positive effect sizes found in these data, fatherhood interventions also generally exhibit small and positive effects (e.g. Henry et al., 2020; Holmes et al., 2020). However, in these data fathers and mothers benefited equally in self-sufficiency improvements, diverging from other analyses suggesting that men didn't on average improve their economic wellbeing across fatherhood-focused interventions (Holmes et al., 2020). The FSPs evaluated here, in contrast to this review, were not specifically focused on fathers as a population and, additionally, did not measure primary versus secondary caregiver status or the intersection between caregiver role and reported gender, limiting the degree to which assumptions may be made specifically about fathers. Therefore, future work should investigate if there are meaningful differences in how fathers in primary, secondary, and those in non-traditional caregiver roles are engaged in family-focused versus fatherhood-focused interventions.

As hypothesized, families with lower income (and education) overall tended to make fewer improvements in self-sufficiency. This issue may be driven by challenges to increasing income itself. For example, families enrolled in FDS with low-income had high levels of difficulty improving their access to income. Only 1% of families overall and 2% of families reporting readiness to change moved above the risk-prevention line on their family's income between timepoint 1 and 2, as reflected by income that was above 200% of the poverty line (CFSA & OMNI, 2018).

Families with elevated levels of stress generally experienced fewer improvements, as hypothesized. Substance abuse, family functioning and resilience, child educational success, and the belief that behavior is not designed to upset parents exhibited lesser pre-post improvements in the context of rising stress. However, stress did not consistently relate to reduced benefits across programs and outcomes, as expected. Rather, contrary to hypothesis, families with higher stress made greater improvements overall in not losing control when disciplining children. This complements other home visiting evaluations utilizing cumulative stress scores demonstrating that when stress is high, parenting practices may improve at greater rates (Green et al., 2014).

ESL families obtained greater benefits in child education, reflecting findings for ESL children enrolled in HS on some academic outcomes (Raikes et al., 2013). BIPOC families also reflected this trend (Appendix C) with greater improvements in child education. Interestingly, both BIPOC and ESL families had higher baseline scores in child education, suggesting they may have reported less need (Appendix D *Table 1*). In one program, Safe CMP, BIPOC families made greater improvements in child education

than in PSSF. It may be that home-visitors serving BIPOC families with young children are more effectively able to support child success and engagement with preschool and kindergarten programming than in mixed-delivery system models, although this idea should be examined in future studies.

More ESL families benefited from services focusing on mental and physical health overall. This may be explained by the barriers ESL families report to accessing social and physical healthcare. The challenges of an already highly complex system are underscored by limited access to linguistically and culturally-matched professionals (Doshi et al., 2020). Adding to this, in instances where ESL families are immigrants, deportation concerns may add barriers to accessing and utilizing healthcare resources. Many ESL families do not receive needed healthcare services or treatment for health concerns due to healthcare access barriers (Marshall et al., 2005), especially healthcare services that are more preventative in nature. In this context, providers across service delivery systems appear to be effective in their shared capacity to support ESL caregivers in identifying and accessing needed preventative treatment to improve functioning and quality of life.

Data revealed that BIPOC families made fewer improvements overall in family functioning and resilience from the pre-to-post assessments, regardless of program. One possible explanation for this is that less need exists in the underlying population. However, this seems unlikely given that BIPOC families did not differ from white families in baseline needs at the start of treatment (see Appendix D *Table 1*). BIPOC communities have additionally been identified as experiencing ongoing inequities and disadvantages, factors which can adversely impact family functioning (Patterson, 2002).

This trend may also reflect limitations in the degree to which BIPOC families are effectively engaged and served. Some work outlines the ways in which BIPOC families may distrust traditional care settings due to experiences of historical persecution and racism, compounded by ongoing experiences of racism and discrimination (Santiago et al., 2013). Indeed, analyses of the experiences of BIPOC families in human services nationally reveal that when enrolled, treatment may systematically differ. Black families are disproportionately likely to be enrolled in lower quality HS sites and BIPOC families receive fewer referrals than white families for child care, employment, and education supports when enrolled in TANF (McDaniel et al., 2017). Therefore, nontrivial differences in the experiences of BIPOC families across services-systems may exist and these results warrant further attention.

Readiness to change, theorized to impact outcomes, did not appear to meaningfully impact improvements in most of the outcomes examined. Although this construct initially emerged as significant in 11 out of 13 domains where it was measured, its effect became insignificant in 9 outcomes following multivariate analyses that accounted for baseline scores and other predictors. Greater readiness to change did, notably, marginally enhance improvements in self-sufficiency and, unexpectedly, predicted fewer substance use improvements. Because a whole field of study is devoted to the granular examination of the stages of change and their relationship with substance use improvement (e.g. DiClemente, 2018), this finding should be examined in future evaluations of family support programs with greater attention to stages of change (e.g. contemplation versus action).

The broader insignificance of readiness to change is surprising, given that readiness to change was associated with greater need and greater improvements in FRCA evaluations (CFRCA & OMNI, 2019). However, this pattern of findings is reflected in other home-visiting evaluations, where bivariate analyses supported readiness to change as meaningfully connected to intent to engage in services. However, multivariate analyses no longer supported readiness to change as a crucial predictor; rather, perceived need, helpfulness of the intervention, and comfort with the home visitor were predictive (McCurdy et al., 2006). Future work should explore how readiness to change may contribute toward or be reflected by other key variables impacting change.

Service delivery systems conferring advantage when adversity is greater

What Works for Whom. The overall patterns of change obscured differential benefits experienced by families with greater levels of adversity. In some cases, apparent greater benefits clustered based on the program's primary service delivery system (see Appendix D *Table 5*). The greatest number of differential subgroup benefits were found in home-visiting and community-based programs (Safe CMP and CCR), where differential benefits were identified in 15 subgroup by outcome areas (*Appendix D Table 5*). Families experiencing elevated adversity may face greater difficulty engaging in center-based supports and home-visiting may resolve a variety of structural barriers faced (Smith et al., 2018) via eliminating the need for transportation, promoting the acceptability of services in the participant's home environment, facilitating a trusting relationship with caregivers facing isolation, and active efforts on the part of the home-visitor to engage the family (versus relying on family initiative; Ammerman et al., 2006; Azzi-Lessing, 2011).

Families with low income in these samples displayed an even mix of benefits between home, center-based, and mixed-delivery services. The hypothesis that families with greater poverty would make greater gains in center-based or mixed-delivery services in social support was unsupported, but in parenting was partially supported. Families enrolled in mixed-delivery programs made greater gains in understanding that misbehavior is not designed to upset parents (Appendix D *Table 5*). Although center-based programs did not consistently differentially benefit families with very low income, as expected, they shared a meaningful role with home and community-based programs in accelerating well-being in families facing disadvantage across an array of needed supports. These results suggest that collaboration among center and home-based programs that build upon strengths may advance effective family support services.

HS, the only fully center-based service delivery system in this evaluation was linked with differential benefits in 12 instances across subgroup by outcome areas (Appendix D *Table 5*). This strong pattern of benefits may be explained by assessments of effective two-generation interventions, which suggest that parent-focused outcomes (e.g. self-sufficiency) may be more likely to improve in contexts where children are simultaneously receiving intervention (Chase-Lansdale & Brooks-Gunn, 2014). However, a pattern was also noted such that the greatest pre-post improvements in HS were sometimes found in families with lower levels of adversity. Families with higher levels of education in HS made greater improvements in self-sufficiency and social support relative to families in other programs, an effect that was absent at lower levels of education (*Appendix E Figure 1a, 7a*). Similarly, families with less stress revealed greater improvements in child care

and when enrolled in HS, an effect that was diminished at higher levels of stress (*Appendix E Figure 2b*). Finally, families in HS demonstrated negative pre-post change in beliefs about child misbehavior, but this effect was only seen at higher levels of stress (*Appendix E Figure 12a*). This pattern is consistent with other literature linking the greatest child and parenting-focused benefits of Early Head Start and HS to families with moderate levels of adversity (e.g. Raikes et al., 2013). However, in other instances HS did benefit families with greater adversity. For example, families with greater stress made greater improvement in concrete support when enrolled in HS (Appendix E *Figure 8a*). Families with elevated stress (e.g. domestic violence, homelessness) were hypothesized to differentially benefit in home-based service delivery settings. However, these families revealed the greatest variability, with differential patterns of benefit by program in 6 out of the 15 investigated outcome areas (Appendix D *Table 5*). When cumulative stress rose, families exhibited high variation in the array of needed programs to optimize well-being, showing differential pre-post gain in every program at least once. This trend may reflect a greater level of complexity in these family's service needs and a greater need for a well-coordinated system of care.

Some families with greater levels of stress appeared to experience better than expected outcomes. Families with higher levels of stress made greater pre-post gains in losing control when disciplining when enrolled in Safe CMP, an effect that was absent at low levels of stress (Appendix E *Figure 13a*). Similarly, at high, but not low levels of stress, families enrolled in Safe CMP made differential improvements in concrete support (Appendix E *Figure 8a*). Families facing other structural adversities also benefited in

Safe CMP. Families with low levels of education differentially benefited when aiming to improve self-sufficiency (Appendix E *Figure 1a*). ESL families made significantly stronger gains in their children's educational achievement and engagement when enrolled in Safe CMP and BIPOC families made greater gains in concrete support (Appendix D *Figure 3a, 8b)*. Thus, Safe CMP's approach targeting both structured parenting-focused needs in combination with individualized case management in home-visitation was especially effective with improving indices of parent well-being (self-sufficiency, concrete support), child well-being (child education), and the parent-child relationship (not losing control when disciplining) in families with elevated stress or those facing systematic and structural forms of adversity. However, it is also notable that in some instances Safe CMP was linked with opposite patterns. For example, Safe CMP supported low stress, but not higher stress, families in improving child education (Appendix E *Figure 3d*). Additional investigation of Safe CMP's impacts are needed in a larger sample.

In some instances, a pattern of worse than expected changes were observed when family stress was high. For example, when stress was higher, families enrolled in PSSF appeared to have greater difficulty improving in several domains. In particular, enrollment in PSSF was linked with adverse pre-post change in concrete support, child education, and losing control during discipline (see Appendix E *Figures 3d, 8a, 13a*). Latinx families in PSSF had negative pre-post gain in nurturing and attachment (Appendix E *Figures 9a*). Thus, an opposite trend was observed from Safe CMP, in that indices of parent well-being (concrete support), child well-being (child education), and

the parenting relationships (nurturing and attachment, losing control when disciplining) were adversely affected by greater levels of stress in PSSF. As one explanation, it may be that there are key differences in the underlying characteristics of the population served in PSSF, as discussed earlier. It may also be that families with high levels of stress exhibit sleeper effects in some programs, with benefits emerging after longer lengths of time in treatment. For example, analyses of EHS and HS revealed that in families with higher levels of stress, positive impacts were observed at age 5 which had not been evident at age 3 (Raikes et al., 2013). Further, children in center-based EHS programs experienced greater benefits at age 3 than those in home-visiting programs, but this effect reversed by age 5 (OPRE, 2006). Therefore, future work should include follow-up assessments to understand downstream impacts of service delivery systems.

Enrollment in PSSF (mixed-delivery system) was also linked with differential benefit for a variety of subgroups, with a notable impact on substance use improvement in the context of elevated stress (see *Table 12* and Appendix E *Figure 5a*). Increasing levels of stress were linked with increasingly adverse change scores in substance abuse across programs, with the exception of PSSF. There is no clear explanation as to why PSSF conferred advantage for substance use in the context of high stress, although it could be that this finding is related to a locally designed treatment approach. Apart from PSSF, however, a pattern of increasing pre-post negative gain was observed across programs. A similar overall pattern of effects has previously been identified in home-visiting models, where small negative effects were identified in substance abuse by the time of treatment

close (e.g. Michalopoulos et al., 2019). The data in these analyses extend on this pattern by suggesting an interaction between stress and apparently worsening substance abuse.

There are several explanations for the seemingly worsening trajectory of substance abuse in families with high stress. Substance abuse may be an issue that caregivers facing already high stress hesitate to disclose to providers (Dauber et al., 2017), especially earlier in treatment when concerns about child welfare involvement may be high and rapport between parent and provider still developing. In support of this idea, parents with greater levels of stress in this sample were less likely to report substance abuse needs at baseline than families with less stress (Appendix D *Table* 3). Because substance abuse has a long-standing relationship with exposure to violence (Chermack & Blow, 2002) and this relationship holds with caregivers (Conners-Burrow et al., 2009), data on domestic violence disclosure in home-visiting may also be instructive. Home visitors have previously reported that caregivers required additional time in the context of home-visiting to gain trust and open up about sensitive challenges, such as domestic violence (Michalopoulos et al., 2019). A statewide evaluation of Health Families America demonstrated rising rates of psychological and physical abuse by treatment end in particular subgroups of women, a phenomenon postulated to reflect increased willingness to disclose these events as opposed to increases in the experiences themselves (McFarlane et al., 2013).

Alternately, decreasing scores may reflect limitations in provider preparation to effectively address and link families with substance abuse needs to services. In one evaluation of 88 home-visiting programs, providers were even less likely to screen, refer,

link, and follow-up with families on substance abuse concerns than caregiver domestic violence (West et al., 2021), making substance use one of the least investigated but also most relevant concerns for child maltreatment prevention in home-visiting. Of all documented service activities, providers were least likely to coordinate warm service hand-offs connected to substance abuse (West et al., 2021). It is also possible that declining scores reflect limited provider understanding of how to identify substance abuse in the home, resulting in delayed positive screenings. One statewide evaluation of the Healthy Start Program identified that providers had significant challenges identifying substance abuse risk in households (Duggan et al., 2004), recognizing a risk that existed only 14% of the time during the first year of treatment. By the third consecutive year of treatment, this rate increased only to 19% (Duggan et al., 2004).

As an alternative explanation, it is possible that declining scores in this sample reflect both limited early recognition of substance abuse and limited knowledge of how to discuss, link, and refer families. In multiple home-visiting evaluations, providers have been identified as less likely to discuss substance abuse or domestic violence concerns with parents after identifying risk than mental health (Dauber et al., 2017) or smoking (Kanda et al., 2022). In a statewide evaluation of Health Families America, providers who knew of substance abuse concerns never referred mothers to community supports; rather, providers attempted to address the concerns directly (Duggan et al., 2004). It is unclear if this was due to limited community resources or the provider's lack of preparation. However, providers reported limited confidence addressing substance abuse due to limited training (Duggan et al., 2018) and may harbor concerns about judicial

system involvement or discomfort with the topic (Kanda et al., 2022). Thus, it is possible that data in this analysis reflect provider need for additional training and preparation to support the needs of families with co-occurring stress (e.g. domestic violence) and substance abuse. Given long-standing knowledge that substance abuse often co-occurs with child maltreatment (Kepple, 2017), that approximately 33% of caregivers enrolled nationally in home-visiting interventions experienced substance abuse needs prior to pregnancy (Michalopoulos et al., 2015), and increasingly worsening national statistics on substance abuse and overdose (Center for Disease Control, 2022), these data may reveal a critical gap between family need and the preparation of family support providers broadly. Failure to address this gap may be consequential, as these experiences are tied to one of the target goals of these programs: the reduction of child maltreatment. Finally, in the absence of a comparison group, it may be that this pattern of results has no connection to programming. Participation in FSPs may have led to less negative gain than would have occurred otherwise. Future studies should examine this issue using experimental study design to better understand the interplay between stress, substance use, and FSP treatment effects.

The conceptualization of family stress in this study in comparison to other studies is an important consideration, as efforts to target and support families facing adversity may benefit from a more unified approach. Previous efforts to catalogue risk commonly index demographic risks (e.g. parent education level, welfare receipt, single-parent) and psychological risks (e.g. caregiver depression), leaving other serious forms of adversity such as domestic violence, substance abuse, and homelessness less commonly examined

(Azzi-Lessing, 2011). This investigation conceptualizes risk as a series of high stress and destabilizing events or experiences in the family (e.g. domestic violence, homelessness). While this approach to measuring and conceptualizing stress differs from prior literature examining program outcomes (e.g. Berlin et al., 2018; McFarlane et al., 2013), the cumulative impacts of these high stress experiences may be especially deleterious and important to capture, both to support targeted enrollment and to guide treatment decisions. Other evaluations (e.g. Green et al., 2014) demonstrate that cumulative models are an effective and straight-forward method to identify families facing high adversity. Furthermore, Dodge (2021) demonstrates that clinical risks including substance abuse, food and housing insecurity, and domestic violence, are better predictors of later maltreatment and, as a result, better indicators of who should be targeted for interventions. Therefore, examination of the role of highly stressful events moves beyond proxies for adversity, increasing understanding of "what works" for families who are experiencing the highest levels of documented adversity.

Future work should isolate the high stress events investigated here, to identify if particular stress experiences are more meaningful than others in identifying patterns of benefit in FSPs. Future work should also incorporate indices of caregiver mental health into considerations of "what works", notably measures of depression and anxiety. A variety of studies have demonstrated that caregivers with higher levels of anxiety benefit more from parent-focused programs (Berlin et al., 2018; McFarlane et al., 2013), although other programs uniquely benefited caregivers with higher levels of avoidance

(Cassidy et al., 2017). Other forms of risk should also be considered in future evaluations that are not captured in this analysis, such as neighborhood safety and cohesion.

Finally, an accurate assessment of risks and their impact on outcomes must occur within the ecological context of communities (Azzi-Lessing, 2011). Communities with high concentration of poverty, fewer jobs and training opportunities, elevated rates of violence or substance use, and low access to well-coordinated service networks may face greater challenges supporting the needs of families and this information would offer critical context into the circumstances in which family support programs "work". The "ecological validity" of these findings may be questioned because there are constraints on a community-level that were not accounted for. Because a key function of FSPs is linkages with other community resources, communities with fewer resources or resources of poorer quality may experience diminished outcomes (Azzi-Lessing, 2011). Resources like medical care, food assistance, and child care may be less accessible in rural regions; whereas, urban regions may face higher levels of crime and poverty. Future work must look closely at how context additionally contributes to interactions between subgroups and programs.

Strengths and limitations

While the very large sample size via integrated data sets is a notable methodological strength of these analyses, smaller subgroups analyses lead to less reliable estimates of effects sizes. On the one hand, many subgroup analyses increase Type 1 error. This risk was reduced by the use of Bonferroni corrections. An alternative issue, however, is that Bonferroni corrections reduce power to detect effects (Nakagawa, 2004). Because

subgroup analyses are not as well-powered as the overall sample, estimates have greater standard errors and may in fact be more conservative and prone to Type II errors. This issue may be especially pronounced in smaller program samples (e.g. Safe CMP). However, given the overall large sample, inclusion of marginal interaction effects, and adjustment of family-wise error, Type I and II error are likely in balance for this exploratory analysis. These analyses are an important step forward, both to inform evaluations of programs implemented in community-based settings and to extract information on areas where inequitable effects are observed. These data are meaningful for programmatic improvement and equity for the families with the greatest need. However, given smaller subgroups and greater standard errors, interpretation should be cautious and reserved for clear patterns of effect as opposed to any single outcome area.

It is also notable that program and subgroup by program effect sizes in these analyses are consistently small in this sample, typically under 0.01. Even while statistically significant, it is unclear that these effects translate to clinically meaningful differences. However, these effects sizes also reflect the extant literature, where effects across family-focused programs are consistently small in community-settings (Azzi-Lessing, 2011; Gubbels et al., 2021; Michalopoulos et al., 2019). Moreover, as study sample sizes in parenting programs increase, effect sizes are noted to systematically decrease (Chen & Chan, 2016; Supplee et al., 2021), a finding also supported here. Yet, modest effect sizes in proximal outcomes are in some cases linked with meaningful outcomes differences for distal targets (e.g. child maltreatment; discussed in Green et al., 2014). Therefore, these effect sizes are not altogether surprising and may be interpreted as relatively normative

and potentially linked with meaningful downstream impacts. As a related issue, the use of a multivariate approach to data analyses may have led to a loss of power to identify effects (Farrell et al., 2013), but this approach also reduced the likelihood that observed relationships with moderators were an artifact of correlation between the moderator and other covariates.

Several other limitations are present in these data. Inverse Probability Weights may be at the highest risk of inflation in PSSF due to extremely high rates of dropout and Safe CMP due to low sample variance. Thus, the greatest confidence in results may be assigned to data from FDS, CCR, and HS, where sample sizes and retention were greater. Several participant profiles in PSSF and Safe CMP had no chance of participation by timepoint 2, which led to extreme weight values. This problem was corrected by removing two predictors with high missingness, entering interaction terms, and trimming weights. Given low levels of engagement in PSSF at timepoint 2, these weights may accurately reflect the uncertainty of participation and may be instructive for future evaluations examining groups with inequitable outcomes. Second, the structure of PSSF constrains the degree to which interpretation can be assigned to findings. While certain types of services are consistently funded by PSSF (e.g. family support, family preservation, adoption and foster-care support) and the bulk of services are clustered under the umbrella of family support programs, chosen curriculum and models vary at the state or county level (Casey Families Programs, 2011). This structure is similarly seen in another widely implemented FSP, Healthy Families America. While this approach is seen as a strength of the program, namely that programs may be responsive and flexible in

response to the needs of their communities, it also presents a major challenge for evaluators (Green et al., 2014) and increases the risk of inconsistent or even negative impacts on families. More prescriptive program models allow for cleaner examination of the interplay between program components and outcomes (e.g. Gubbels et al., 2021), while the variation inherent to PSSF in this state (like Healthy Families America) make interpretation and application of findings challenging. Therefore, additional caution should be used when interpreting the effects of PSSF.

This evaluation does not attempt to isolate or examine moderation by core components or active ingredients across programs. In some home-visiting evaluations, efforts have been made to isolate core program components, the interventions specific to each model. For example, Beatson and colleagues (2021) attempted to isolate active ingredients among seven home-visiting programs which all targeted postpartum women. Ultimately, it was challenging to tie specific content across programs to variation in child-level outcomes (Beatson et al., 2021), leaving the prevailing 'black box' view of home visiting largely intact. An important next step following this evaluation is consideration of "why" some programs worked better than others.

Some work finds that dosage and provider-effects moderate outcomes, predictors that were not measured in this analysis. Many families do not receive the prescribed dosage, the length or intensity of services, which have previously been tied to model efficacy (Ammerman et al., 2006; Goyal et al., 2013; Holland et al., 2018; Kumpfer et al., 2002). While this study does not control for total dosage tied to improvements, length of time between intervention assessments was included as a covariate when meaningful. This

analysis is a short-term longitudinal design examining the baseline and first follow-up assessment received by families, which sometimes occurred within longer treatment periods. While most families did complete their baseline and final CFSA measure within two timepoints, others stayed engaged for longer periods of time, especially in FDS. Future analyses should examine how length of total treatment time impacts improvements. Factors in the participant-provider relationships, such as perceived cultural competence, rapport, and provider experience with children, may impact outcomes (Beasley et al., 2018; Damashek et al., 2011) and should be examined in future subgroup analyses.

Finally, the interventions assessed here do not serve tightly aligned populations. While all evaluated FSPs support families with children experiencing elevated need, children were in varied developmental stages (e.g. Safe CMP only served families with children ages 0-5, HS 3-5, and PSSF/FDS served all ages). This may be problematic for an evaluation of "what works", as the parenting needs of infants appear to be vastly divergent from those of adolescents. The impact of this limitation on the current analyses is mitigated by the fact that neither child specific outcomes nor child-level predictors were included; nevertheless, the experience of parenting of course differs by child age. Subgroups were also not analyzed by differences in child age or other child-level needs. Rather, this analysis identifies the FSPs that effectively stabilized parent and parenting-focused outcomes in family development, utilizing measures that cross-cut ages and developmental stages. It is also notable that all programs in this evaluation are accessible and utilized by children eligible for Head Start (3-5-year-olds).

Implications

The findings in this study reveal that families experiencing elevated stress and disadvantage may benefit from optimized and strategic treatment pathways that draw on the strengths of multiple programs. In turn, prescriptive programs that established evidence on average in single-site studies may be insufficient to move the science of prevention forward (Azzi-Lessing, 2011; Supplee et al., 2021). Increasingly, the field of prevention science seeks to enhance individualized outcomes and is shifting toward strategic testing of subpopulations prior to scaling. This effort is termed Precision Home-Visiting (PHV) and seeks to "match families to the best possible programs and services for their individual needs, interests, and desired outcomes" (The Home Visiting Applied Research Collaborative [HARC], 2018).

First, families could be matched to programs where their benefit is likely highest. While post hoc strategies to link families within existing programs may be helpful, these data may also inform future efforts to optimize interventions prior to scaling. Increasingly, innovative research using rapid-cycle, iterative designs are testing how to best support target populations (Supplee & Duggan, 2019). These methods include Multiphase Optimization Strategies and Sequential Multiple Assignment Randomized Trials to identify subgroup and subcomponent effects. The results described here may inform preliminary hypotheses on subgroup needs for these intervention trial designs. Alternately, programs could also use these results to examine their practices with groups experiencing disparate engagement and improvement. These data may spur identification of alternative programs for families facing adversity who are not benefiting as expected

(August & Gewirtz, 2019). Finally, a third less common strategy could involve modifying existing systems to train home visitors on how to identify subgroups failing to benefit within the model they are delivering and to identify avenues to work specifically with them more effectively (McFarlane et al., 2013). Alternately, as successfully demonstrated by Ondersma and colleagues (2017), efforts may be made to supplement and add on to interventions utilizing e-modules that don't require home visitor retraining or preparation. These modules may effectively address areas of significant need (e.g. substance abuse and domestic violence) that don't appear to be effectively addressed currently by home visitors.

Conclusion

These findings echo those of Layzer and colleagues more than 20 years ago in their national evaluation of family support programs (Layzer et al., 2001). Still, today, small significant effects were identified on average and no evidence supported that any single program or service strategy worked for all populations. Still, as in 20 years ago, program effect sizes appeared to be even smaller in the largest multi-site interventions evaluated (FDS). However, this work pushes forward the precision with which programs effectively supporting families with high levels of stress may be identified. No "one size" fits all programming for families experiencing elevated stress and structural adversities were identified. This study, instead, identifies a series of themes, including that families in HS and CCR models of family support appeared to experience greater gains in a variety of outcome areas. Additionally, families with higher levels of stress and adversity experienced a pattern of differential improvement when enrolled in Safe CMP. These

findings may generate hypotheses regarding how the combination of parenting education and access to concrete supports such as cash assistance paired with home-visiting model may be protective. Center-based services filled a critical role, supporting differential benefits across many groups, but without a sufficiently clear pattern on which to draw conclusions. These data also reveal high levels of needed attention in the areas of substance abuse across programs, especially for families facing high levels of stress. Together, these findings may serve to accelerate impact in family support programs, as future studies continue to identify "what works for whom" and innovate the tailoring of interventions for those that most need them.

References

Administration for Children and Families. (2002). Making a difference in the lives of children and families: The impacts of Early Head Start programs on infants and toddlers and their families. Washington, DC: U. S. Department of Health and Human Services. [*PDF File*] Retrieved from: https://www.acf.hhs.gov/sites/default/files/opre/impacts_vol1.pdf

Allan, H., Currie, D., Drury, I., Merkel-Holguin, L., Fluke, J., & Winokur, M. (2018). *Colorado Department of Human Services Colorado Community Response Final Evaluation Report 2014-2018*. 80.

Ammerman, R. T., Stevens, J., Putnam, F. W., Altaye, M., Hulsmann, J. E., Lehmkuhl, H. D., Monroe, J. C., Gannon, T. A., & Van Ginkel, J. B. (2006). Predictors of Early Engagement in Home Visitation. *Journal of Family Violence*, *21*(2), 105–115. https://doi.org/10.1007/s10896-005-9009-8

Arnold, L. E., Elliott, M., Sachs, L., Bird, H., Kraemer, H. C., Wells, K. C., . . . Wigal, T. (2003). Effects of ethnicity on treatment attendance, stimulant response/dose, and 14-month outcome in ADHD. *Journal of Consulting and Clinical Psychology, 71*(4), 713-727. http://dx.doi.org/10.1037/0022-006X.71.4.713

August, G. J., & Gewirtz, A. (2019). Moving Toward a Precision-Based, Personalized Framework for Prevention Science: Introduction to the Special Issue. *Prevention Science*, *20*(1), 1–9. https://doi.org/10.1007/s11121-018-0955-9

Axford, N., Lehtonen, M., Kaoukji, D., Tobin, K., & Berry, V. (2012). Engaging parents in parenting programs: Lessons from research and practice. *Children and Youth Services Review*, *34*(10), 2061–2071. https://doi.org/10.1016/j.childyouth.2012.06.011

Azzi-Lessing, L. (2011). Home visitation programs: Critical Issues and Future Directions. *Early Childhood Research Quarterly*, *26*(4), 387–398. https://doi.org/10.1016/j.ecresq.2011.03.005

Baker, C. N., Arnold, D. H., & Meagher, S. (2011). Enrollment and Attendance in a Parent Training Prevention Program for Conduct Problems. *Prevention Science*, *12*(2), 126–138. https://doi.org/10.1007/s11121-010-0187-0

Barrera, M., Biglan, A., Taylor, T. K., Gunn, B. K., Smolkowski, K., Black, C., ... & Fowler, R. C. (2002). Early elementary school intervention to reduce conduct problems: A randomized trial with Hispanic and non-Hispanic children. *Prevention Science*, *3*(2), 83-94.

Barnes, J., MacPherson, K., & Senior, R. (2005). Factors influencing the acceptance of volunteer home-visiting support offered to families with new babies. *Child & Family Social Work, 11*(2), 107-117. https://doi.org/10.1111/j.1365-2206.2006.00401.x

Bauer, L., & Schanzenbach, D. W. (2016). The long-term impact of the Head Start program. *The Hamilton Project.* [*PDF File*] Retrieved from: https://pdfs.semanticscholar.org/89fe/d0ffa9b17eab87cd45acba8720a7ebc16a18.pdf

Beachy-Quick, K., Holmquist-Johnson, H., Lee, C., McConnell, L., Orsi, R., Timpe, Z., & Winokur, M. (2017). *Colorado Department of Human Services SafeCare Colorado Pilot Project Evaluation Report*. [*PDF File*] Retrieved from: https://earlychildhoodframework.org/wp-content/uploads/2018/10/Reports_SafeCareColoradoPilotProjectFinalEvaluationReport.pdf

Beasley, L. O., Ridings, L. E., Smith, T. J., Shields, J. D., Silovsky, J. F., Beasley, W., & Bard, D. (2018). A Qualitative Evaluation of Engagement and Attrition in a Nurse Home Visiting Program: From the Participant and Provider Perspective. *Prevention Science*, *19*(4), 528–537. https://doi.org/10.1007/s11121-017-0846-5

Beatson, R., Molloy, C., Perini, N., Harrop, C., & Goldfeld, S. (2021). Systematic review: An exploration of core componentry characterizing effective sustained nurse home visiting programs. *Journal of Advanced Nursing*, *77*(6), 2581–2594. https://doi.org/10.1111/jan.14755

Begle, A. M., Lopez, C., Cappa, K., Dumas, J. E., & de Arellano, M. A. (2012). Ethnicity differences in child and parental outcomes following involvement the PACE program. *Behaviour Research and Therapy, 50*(1), 56-64. https://doi.org/10.1016/j.brat.2011.09.009

Belsky, J. (1984). The determinants of parenting: A process model. *Child Development*, 55(1), 83-96. DOI: 10.2307/1129836

Berlin, L. J., Martoccio, T. L., & Jones Harden, B. (2018). Improving early head start's impacts on parenting through attachment-based intervention: A randomized

controlled trial. *Developmental Psychology*, *54*(12), 2316–2327. https://doi.org/10.1037/dev0000592

Bloom, H. S., & Michalopoulos, C. (2013). When is the Story in the Subgroups?: Strategies for Interpreting and Reporting Intervention Effects for Subgroups. *Prevention Science*, *14*(2), 179–188. https://doi.org/10.1007/s11121-010-0198-x

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology, 22*, 723–742.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of Child Psychology* (pp. 793–828). Hoboken, NJ: Wiley.

Brand, T., & Jungmann, T. (2014). Participant characteristics and process variables predict attrition from a home-based early intervention program. *Early Childhood Research Quarterly*, *29*(2), 155–167. https://doi.org/10.1016/j.ecresq.2013.12.001

Brodsky, A. E., & Marx, C. M. (2001). Layers of identity: Multiple psychological senses of community within a community setting. *Journal of Community Psychology*, *29*(2), 161–178. https://doi.org/10.1002/1520-6629(200103)29:2<161::AID-JCOP1011>3.0.CO;2-1

Brooks-Gunn, J., & Duncan, G. J. (1997). The Effects of Poverty on Children. *The Future of Children*, *7*(2), 55–71. https://doi.org/10.2307/1602387

Brown, C. H., Sloboda, Z., Faggiano, F., Teasdale, B., Keller, F., Burkhart, G., Vigna-Taglianti, F., Howe, G., Masyn, K., Wang, W., Muthén, B., Stephens, P., Grey, S., & Perrino, T. (2013). Methods for Synthesizing Findings on Moderation

Effects Across Multiple Randomized Trials. *Prevention Science*, *14*(2), 144–156. https://doi.org/10.1007/s11121-011-0207-8

Buhi, E. R., Goodson, P., & Neilands, T. B. (2008). Out of sight, not out of mind: Strategies for handling missing data. *American Journal of Health Behavior, 32*, 83–92.

Casey Family Programs (2011). Promoting safe and stable families: Background and context. [*PDF File*] Retrieved from: https://www.casey.org/media/PromotingSafeandStableFamilies.pdf

Casillas, K. L., Fauchier, A., Derkash, B. T., & Garrido, E. F. (2016). Implementation of evidence-based home visiting programs aimed at reducing child maltreatment: A meta-analytic review. *Child Abuse & Neglect*, *53*, 64–80. https://doi.org/10.1016/j.chiabu.2015.10.009

Cassidy, J., Brett, B. E., Gross, J. T., Stern, J. A., Martin, D. R., Mohr, J. J., & Woodhouse, S. S. (2017). Circle of Security–Parenting: A randomized controlled trial in Head Start. *Development and Psychopathology*, *29*(2), 651–673. https://doi.org/10.1017/S0954579417000244

Center for Disease Control. (2022, April 5). Vital Statistics Rapid Release - Provisional Drug Overdose Data. https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm

Chacko, A., Jensen, S. A., Lowry, L. S., Cornwell, M., Chimklis, A., Chan, E., Lee, D., & Pulgarin, B. (2016). Engagement in Behavioral Parent Training: Review of the Literature and Implications for Practice. *Clinical Child and Family Psychology Review*, *19*(3), 204–215. https://doi.org/10.1007/s10567-016-0205-2

Chaffin, M., Bonner, B. L., & Hill, R. F. (2001). Family preservation and family support programs: Child maltreatment outcomes across client risk levels and program types. *Child Abuse & Neglect*, *25*(10), 1269–1289. https://doi.org/10.1016/S0145-2134(01)00275-7

Chaffin, M., Valle, L. A., Funderburk, B., Gurwitch, R., Silvosky, J., Bard, D., … Kees, M.(2009). A motivational intervention can improve retention in PCIT for low-motivation child welfare clients. *Child Maltreatment*, *14*, 356–368. doi:10.1177/1077559509332263

Chase-Lansdale, P. L., & Brooks-Gunn, J. (2014). Two-Generation Programs in the Twenty-First Century. *The Future of Children*, *24*(1), 13–39. https://doi.org/10/f5zvxx

Chen, M., & Chan, K. L. (2016). Effects of Parenting Programs on Child Maltreatment Prevention: A Meta-Analysis. *Trauma, Violence, & Abuse*, *17*(1), 88–104. https://doi.org/10.1177/1524838014566718

Chermack, S. T., & Blow, F. C. (2002). Violence among individuals in substance abuse treatment: The role of alcohol and cocaine consumption. *Drug and Alcohol Dependence*, *66*(1), 29–37. https://doi.org/10.1016/S0376-8716(01)00180-6

Child Care Technical Assistance Network. (n.d.). *Data Explorer and State Profiles | ECTTAS* | Retrieved April 28, 2022, from https://childcareta.acf.hhs.gov/data

Colorado Department of Human Services (2020). Colorado Family First Prevention Plan. Retrieved from: https://co4kids.org/sites/default/files/Family%20First%20Prevention%20Plan.pdf

Counts, J. M., Buffington, E. S., Chang-Rios, K., Rasmussen, H. N., & Preacher, K. J. (2010). The development and validation of the protective factors survey: A self-report measure of protective factors against child maltreatment. *Child Abuse & Neglect, 34*(10), 762-772. https://doi.org/10.1016/j.chiabu.2010.03.003

Colorado Family Resource Center Association, & OMNI. (2019). *Colorado Family Resource Center Association 2018-2019 Evaluation Report Executive Summary*. https://cofamilycenters.org/wp-content/uploads/2019/08/2019-FRCA-Aggregate-Evaluation-Report-Executive-Summary.pdf

Colorado Family Resource Center Association, & OMNI. (2020). *Colorado Family Resource Center Association 2019-2020 Evaluation Report Executive Summary*. https://www.cofamilycenters.org/wp-content/uploads/2020/09/FRCA-2019-20-Aggregate-Executive-Summary-Final.pdf

Colorado Office of Early Childhood (2019) Annual Progress and Services Report: Promoting Safe and Stable Families. Prepared for the Administration for Children and Families. Unpublished (draft reporting shared with the author).

Conger, R. D., Wallace, L. E., Sun, Y., Simons, R. L., McLoyd, V. C., & Brody, G. H. (2002). Economic pressure in African American families: A replication and extension of the family stress model. *Developmental Psychology, 38*(2), 179-193. http://dx.doi.org/10.1037/0012-1649.38.2.179

Conger, R., & Donnellan, M. (2007). An interactionist perspective on the socioeconomic context of human development. *Annual Review of Psychology, 58*, 175-99. https://doi.org/10.1146/annurev.psych.58.110405.085551

Conners-Burrow, N. A., Johnson, B., & Whiteside-Mansell, L. (2009). Maternal Substance Abuse and Children's Exposure to Violence. *Journal of Pediatric Nursing*, *24*(5), 360–368. https://doi.org/10.1016/j.pedn.2008.03.006

Damashek, A., Doughty, D., Ware, L., & Silovsky, J. (2011). Predictors of Client Engagement and Attrition in Home-Based Child Maltreatment Prevention Services. *Child Maltreatment*, *16*(1), 9–20. https://doi.org/10.1177/1077559510388507

Daro, D., McCurdy, K., Falconnier, L., & Stojanovic, D. (2003). Sustaining new parents in home visitation services: Key participant and program factors. *Child Abuse & Neglect*, *27*(10), 1101–1125. https://doi.org/10.1016/j.chiabu.2003.09.007

Data Explorer and State Profiles | ECTTAS | Child Care Technical Assistance Network. (n.d.). Retrieved April 28, 2022, from https://childcareta.acf.hhs.gov/data

Dauber, S., Ferayorni, F., Henderson, C., Hogue, A., Nugent, J., & Alcantara, J. (2017). Substance use and depression in home visiting clients: Home visitor perspectives on addressing clients' needs. *Journal of Community Psychology*, *45*(3), 396–412. https://doi.org/10.1002/jcop.21855

de Haymes, M. V., & Simon, S. (2003). Transracial Adoption: Families Identify Issues and Needed Support Services. *Child Welfare*, *82*(2), 251–272.

DiClemente, C. C. (2018). Addiction and change: How addictions develop and addicted people recover *(2nd ed.). Guilford Press.*

Dodge, K. A., Benjamin Goodman, W., Bai, Y., Murphy, R. A., & O'Donnell, K. (2021). Maximizing the return on investment in Early Childhood Home Visiting through

enhanced eligibility screening. *Child Abuse & Neglect*, *122*, 105339. https://doi.org/10.1016/j.chiabu.2021.105339

Doshi, M., Lopez, W. D., Mesa, H., Bryce, R., Rabinowitz, E., Rion, R., & Fleming, P. J. (2020). Barriers & facilitators to healthcare and social services among undocumented Latino(a)/Latinx immigrant clients: Perspectives from frontline service providers in Southeast Michigan. *PLOS ONE*, *15*(6), e0233839. https://doi.org/10.1371/journal.pone.0233839

Duggan, A., Fuddy, L., Burrell, L., Higman, S. M., McFarlane, E., Windham, A., & Sia, C. (2004). Randomized trial of a statewide home visiting program to prevent child abuse: Impact in reducing parental risk factors. *Child Abuse & Neglect*, *28*(6), 623–643. https://doi.org/10.1016/j.chiabu.2003.08.008

Duggan, A., Portilla, X. A., Filene, J. H., Crowne, S. S., Hill, C. J., Lee, H., & Knox, V. (2018). *Implementation of Evidence-Based Early Childhood Home Visiting: Results from the Mother and Infant Home Visiting Program Evaluation*. 411.

Eckenrode, J., Campa, M., Luckey, D. W., Henderson, C. R., Cole, R., Kitzman, H., Anson, E., Sidora-Arcoleo, K., Powers, J., & Olds, D. (2010). Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial. *Archives of Pediatrics & Adolescent Medicine*, *164*(1), 9–15. https://doi.org/10.1001/archpediatrics.2009.240

Eisner, M., & Meidert, U. (2011). Stages of Parental Engagement in a Universal Parent Training Program. *The Journal of Primary Prevention*, *32*(2), 83–93. https://doi.org/10.1007/s10935-011-0238-8

Euser, S., Alink, L. R., Stoltenborgh, M., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2015). A gloomy picture: a meta-analysis of randomized controlled trials reveals disappointing effectiveness of programs questioning at preventing child maltreatment. *BMC Public Health*, *15*(1), 1068. https://doi.org/10.1186/s12889-015-2387-9

Everson, C., Klopfenstein, K., & Prendergast, S. (2021). *A Cost Analysis of Colorado Community Response*. https://coloradolab.org/wp-content/uploads/2021/07/CCR_CostStudy_PolicyBrief.pdf

Family Resource Center Association (2014). *Colorado Family Support Assessment 2.0.* [*PDF File*] Retrieved from: http://outcomes.ascend.aspeninstitute.org/wp-content/uploads/gravity_forms/1-71975598de9338f71e49a1d13937b9e0/2016/04/Colorado-Family-Support-Assessment_12-10-14.pdf

Farrell, A. D., Henry, D. B., & Bettencourt, A. (2013). Methodological Challenges Examining Subgroup Differences: Examples from Universal School-based Youth Violence Prevention Trials. *Prevention Science*, *14*(2), 121–133. https://doi.org/10.1007/s11121-011-0200-2

Fernandez, M. A., & Eyberg, S. M. (2009). Predicting Treatment and Follow-up Attrition in Parent–Child Interaction Therapy. *Journal of Abnormal Child Psychology, 37*(3), 431–441. https://doi.org/10.1007/s10802-008-9281-1

Friars, P., & Mellor, D. (2009). Drop-out from parenting training programmes: A retrospective study. *Journal of Child & Adolescent Mental Health*, *21*(1), 29–38. https://doi.org/10.2989/JCAMH.2009.21.1.5.807

FRIENDS National Center for Community Based Child Abuse Prevention (2020). *The Protective Factors Survey User's Manual*. [*PDF File*] Retrieved from *https://friendsnrc.org*

Garson, G. D. (2015). *Missing Values Analysis & Data Imputation*. 26. North Carolina: Statistical Associates Publishers: Statistical Associates Publishers;

Goyal, N. K., Hall, E. S., Meinzen-Derr, J. K., Kahn, R. S., Short, J. A., Van Ginkel, J. B., & Ammerman, R. T. (2013). Dosage Effect of Prenatal Home Visiting on Pregnancy Outcomes in At-Risk, First-Time Mothers. *Pediatrics*, *132*(Suppl 2), S118–S125. https://doi.org/10.1542/peds.2013-1021J

Green, B. L., Tarte, J. M., Harrison, P. M., Nygren, M., & Sanders, M. B. (2014). Results from a randomized trial of the Healthy Families Oregon accredited statewide program: Early program impacts on parenting. *Children and Youth Services Review*, *44*, 288–298. https://doi.org/10/f6gbbk

Gross, D., Julion, W., & Fogg, L. (2001). What Motivates Participation and Dropout Among Low-Income Urban Families of Color in a Prevention Intervention?*. *Family Relations*, *50*(3), 246–254. https://doi.org/10.1111/j.1741-3729.2001.00246.x

Gubbels, J., van der Put, C. E., Stams, G.-J. J. M., Prinzie, P. J., & Assink, M. (2021). Components associated with the effect of home visiting programs on child

maltreatment: A meta-analytic review. *Child Abuse & Neglect*, *114*, 104981. https://doi.org/10.1016/j.chiabu.2021.104981

Hackworth, N. J., Matthews, J., Westrupp, E. M., Nguyen, C., Phan, T., Scicluna, A., Cann, W., Bethelsen, D., Bennetts, S. K., & Nicholson, J. M. (2018). What Influences Parental Engagement in Early Intervention? Parent, Program and Community Predictors of Enrolment, Retention and Involvement. *Prevention Science*, *19*(7), 880–893. https://doi.org/10.1007/s11121-018-0897-2

Hair, J., Black, W. C., Babin, B. J. & Anderson, R. E. (2010) *Multivariate data analysis (7th ed).* Upper Saddle River, New Jersey: Pearson Educational International.

Head Start Early Childhood Learning & Knowledge Center (2018). *Understanding Family Engagement Outcomes: Research to Practice Series*. 8. [*PDF File*] Retrieved from: https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/rtp-family-well-being.pdf

Head Start Early Childhood Learning & Knowledge Center (2019). *Head Start Program Facts: Fiscal Year 2019*. 8.

Head Start Early Childhood Learning & Knowledge Center (2022). *Poverty Guidelines and Determining Eligibility for Participation in Head Start Programs.* Retrieved from: https://eclkc.ohs.acf.hhs.gov/eligibility-ersea/article/poverty-guidelines-determining-eligibility-participation-head-start-programs

Hebbeler, K. M., & Gerlach-Downie, S. G. (2002). Inside the black box of home visiting: A qualitative analysis of why intended outcomes were not achieved. *Early*

Childhood Research Quarterly, *17*(1), 28–51. https://doi.org/10.1016/S0885-2006(02)00128-X

Henry, J. B., Julion, W. A., Bounds, D. T., & Sumo, J. (2020). Fatherhood Matters: An Integrative Review of Fatherhood Intervention Research. *The Journal of School Nursing*, *36*(1), 19–32. https://doi.org/10.1177/1059840519873380

Holland, M. L., Olds, D. L., Dozier, A. M., & Kitzman, H. J. (2018). Visit Attendance Patterns in Nurse-Family Partnership Community Sites. *Prevention Science*, *19*(4), 516–527. https://doi.org/10.1007/s11121-017-0829-6

Holland, M. L., Yinglin Xia, Kitzman, H. J., Dozier, A. M., & Olds, D. L. (2014). Patterns of Visit Attendance in the Nurse–Family Partnership Program. *American Journal of Public Health*, *104*(10), e58–e65. https://doi.org/10.2105/AJPH.2014.302115

Holmes, E. K., Egginton, B. R., Hawkins, A. J., Robbins, N. L., & Shafer, K. (2020). Do Responsible Fatherhood Programs Work? A Comprehensive Meta-Analytic Study. *Family Relations*, *69*(5), 967–982. https://doi.org/10.1111/fare.12435

Hosmer, D.W. & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.

Huston, A. C., & Bentley, A. C. (2010). Human development in societal context. Annual review of psychology, 61, 411-437. https://doi.org/10.1146/annurev.psych.093008.100442

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical

trials – a practical guide with flowcharts. *BMC Medical Research Methodology*, *17*(1), 162. https://doi.org/10/gcrgxr

Jones, H. A., Epstein, J. N., Hinshaw, S. P., Owens, E. B., Chi, T. C., Arnold, L. E., … Wells, K. C. (2010). Ethnicity as a Moderator of Treatment Effects on Parent—Child Interaction for Children With ADHD. *Journal of Attention Disorders, 13*(6), 592–600. https://doi.org/10.1177/1087054709332158

Kanda, K., Blythe, S., Grace, R., Elcombe, E., & Kemp, L. (2022). Variations in sustained home visiting care for mothers and children experiencing adversity. *Public Health Nursing*, *39*(1), 71–81. https://doi.org/10.1111/phn.13014

Korfmacher, J. (2020). Balancing Rigor with Complexity in Understanding the Impacts of Child Maltreatment Prevention Programs. *Prevention Science*, *21*(1), 47–52. https://doi.org/10.1007/s11121-019-01079-1

Korfmacher, J., Green, B., Staerkel, F., Peterson, C., Cook, G., Roggman, L., Faldowski, R. A., & Schiffman, R. (2008). Parent Involvement in Early Childhood Home Visiting. *Child & Youth Care Forum*, *37*(4), 171–196. https://doi.org/10.1007/s10566-008-9057-3

Kepple, N. J. (2017). The Complex Nature of Parental Substance Use: Examining Past Year and Prior Use Behaviors as Correlates of Child Maltreatment Frequency. *Substance Use & Misuse*, *52*(6), 811–821. https://doi.org/10.1080/10826084.2016.1253747

Kumpfer, K. L., Alvarado, R., Smith, P., & Bellamy, N. (2002). Cultural Sensitivity and Adaptation in Family-Based Prevention Interventions. *Prevention Science*, 6. https://doi.org/10.1023/A:1019902902119

Layzer, J. I., Goodson, B. D., Bernstein, L., & Price, C. (2001). *National Evaluation of FamilySupport Programs. Final Report Volume A: The Meta-Analysis*. [*PDF File*] Retrieved from https://files.eric.ed.gov/fulltext/ED462186.pdf

Leslie, L. K., Mehus, C. J., Hawkins, J. D., Boat, T., McCabe, M. A., Barkin, S., Perrin, E. C., Metzler, C. W., Prado, G., Tait, V. F., Brown, R., & Beardslee, W. (2016). Primary Health Care: Potential Home for Family-Focused Preventive Interventions. *American Journal of Preventive Medicine*, *51*(4 Suppl 2), S106-118. https://doi.org/10.1016/j.amepre.2016.05.014

Lightburn, A., & Warren-Adamson, C. (2015). Family Poverty: Reviewing the Evidence for an Integrated Community-Based Practice. In E. Fernandez, A. Zeira, T. Vecchiato, & C. Canali (Eds.), *Theoretical and Empirical Insights into Child and Family Poverty* (Vol. 10, pp. 285 301). Springer International Publishing. https://doi.org/10.1007/978-3-319-17506 5_18

Little, R. J. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review/Revue Internationale de Statistique*, 139-157.

Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., … Vogel, C.

(2005). The Effectiveness of Early Head Start for 3-Year-Old Children and Their Parents: Lessons for Policy and Programs. *Developmental Psychology*, *41*(6), 885–901. https://doi.org/10.1037/0012-1649.41.6.885

Lundahl, B. W., Nimer, J., & Parsons, B. (2006). Preventing child abuse: A meta-analysis of parent training programs. *Research on Social Work Practice, 1*6(3), 251-262. https://doi.org/10.1177/1049731505284391

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews. Neuroscience*, *10*(6), 434–445. http://dx.doi.org.du.idm.oclc.org/10.1038/nrn2639

Marcal, K. E. (2018). The Impact of Housing Instability on Child Maltreatment: A Causal Investigation. *Journal of Family Social Work*, *21*(4–5), 331–347. https://doi.org/10.1080/10522158.2018.1469563

Marshall, K. J., Urrutia-Rojas, X., Mas, F. S., & Coggin, C. (2005). Health Status and Access to Health Care of Documented and Undocumented Immigrant Latino Women. *Health Care for Women International*, *26*(10), 916–936. https://doi.org/10.1080/07399330500301846

Mauricio, A. M., Mazza, G. L., Berkel, C., Tein, J.-Y., Sandler, I. N., Wolchik, S. A., & Winslow, E. (2018). Attendance Trajectory Classes Among Divorced and Separated Mothers and Fathers in the New Beginnings Program. *Prevention Science*, *19*(5), 620–629. https://doi.org/10.1007/s11121-017-0783-3

McConnell, D., Breitkreuz, R., & Savage, A. (2013). Parent needs and family support service outcomes in a Canadian sample. *Journal of Social Work*, *13*(5), 447–470. https://doi.org/10.1177/1468017311434819

McCurdy, K., & Daro, D. (2001). Parent Involvement in Family Support Programs: An Integrated Theory*. *Family Relations*, *50*(2), 113–121. https://doi.org/10.1111/j.1741-3729.2001.00113.x

McCurdy, K., Daro, D., Anisfeld, E., Katzev, A., Keim, A., LeCroy, C., McAfee, C., Nelson, C., Falconnier, L., McGuigan, W. M., Park, J. K., Sandy, J., & Winje, C. (2006). Understanding maternal intentions to engage in home visiting programs. *Children and Youth Services Review*, *28*(10), 1195–1212. https://doi.org/10.1016/j.childyouth.2005.11.010

McDaniel, M., Woods, T., Pratt, E., & Simms, M. C. (2017). *Identifying Racial and Ethnic Disparities in Human Services*. [*PDF File*] Retrieved from: https://www.acf.hhs.gov/sites/default/files/documents/opre/identifying_racial_and_ethnic_disparities_b508.pdf

McEwen, C. A., & McEwen, B. S. (2017). Social structure, adversity, toxic stress, and intergenerational poverty: An early childhood model. *Annual Review of Sociology, 43*, 445-472. https://doi.org/10.1146/annurev-soc-060116-053252

McFarlane, E., Burrell, L., Crowne, S., Cluxton-Keller, F., Fuddy, L., Leaf, P. J., & Duggan, A. (2013). Maternal Relationship Security as a Moderator of Home Visiting Impacts on Maternal Psychosocial Functioning. *Prevention Science*, *14*(1), 25–39. https://doi.org/10.1007/s11121-012-0297-y

McKee, K., Cabrera, N., Alonso, A., Turcios, M., & Reich, S. (2021). Determinants of fathers' and mothers' involvement in a parenting intervention. *Psychology of Men & Masculinities*, *22*(3), 521–537. https://doi.org/10.1037/men0000320

McWayne, C. M., Melzi, G., Schick, A. R., Kennedy, J. L., & Mundt, K. (2013). Defining family engagement among Latino Head Start parents: A mixed-methods measurement development study. *Early Childhood Research Quarterly*, *28*(3), 593–607. https://doi.org/10.1016/j.ecresq.2013.03.008

Mendez, J. L., Carpenter, J. L., LaForett, D. R., & Cohen, J. S. (2009). Parental Engagement and Barriers to Participation in a Community-Based Preventive Intervention. *American Journal of Community Psychology*, *44*(1–2), 1–14. https://doi.org/10.1007/s10464-009-9252-x

Michalopoulos, C., Crowne, S. S., Portilla, X. A., Lee, H., Filene, J. H., Duggan, A., & Knox, V. (2019). *A Summary of Results from the MIHOPE and MIHOPE-Strong Start Studies Of Evidence-Based Home Visiting*. OPRE Report 2019-09. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Miller, E. B., Farkas, G., & Duncan, G. J. (2016). Does Head Start differentially benefit children with risks targeted by the program's service model? *Early Childhood Research Quarterly, 34*, 1-12. https://doi.org/10.1016/j.ecresq.2015.08.001

Miller WR, Rollnick S. Motivational interviewing, preparing people to change addictive behavior. New York: The Guildford Press; 1991.

Moran, P., & Van der Merwe, A. (2004). *What works in parenting support?: A review of the international evidence.* Dept. for Education and Skills.

Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral ecology*, *15*(6), 1044-1045. https://doi.org/10.1093/beheco/arh107

National Committee to Prevent Child Abuse. (1995). *Final evaluation of nine child abuse prevention programs*. Chicago: Author.

National Family Support Network (n.d.) What is a family resource center? National Family Support Network. Retrieved from: https://www.nationalfamilysupportnetwork.org/_files/ugd/ec0538_c9a795f747a34530a0ebce21b11c1c59.pdf

Nix, R. L., Bierman, K. L., & McMahon, R. J. (2009). How attendance and quality of participation affect treatment response to parent management training. *Journal of Consulting and Clinical Psychology*, *77*(3), 429–438. https://doi.org/10.1037/a0015028

Office of Planning, Research and Evaluation, Administration for Children and Families, & Department of Health and Human Services. (2006). *Findings from the Survey of Early Head Start Programs: Communities, Programs, and Families*. 277. Retrieved from: https://www.acf.hhs.gov/sites/default/files/documents/opre/findings_ehs_survey.pdf

Olds, D. L., Sadler, L., & Kitzman, H. (2007). Programs for parents of infants and toddlers: recent evidence from randomized trials. *Journal of Child Psychology and Psychiatry, 48*(3-4), 355-391. https://doi.org/10.1111/j.1469-7610.2006.01702.x

OMNI Institute & Family Resource Center Association (2018). Family Pathways & CFSA 2.0 Evaluation Report. [PDF File] Retrieved from: http://www.cofamilycenters.org/wp-content/uploads/2018/10/2018-FRCA-Family-Development-Report_FINAL_Updated-Sept-2018_Executive-Summary.pdf

Ondersma, S. J., Martin, J., Fortson, B., Whitaker, D. J., Self-Brown, S., Beatty, J., Loree, A., Bard, D., & Chaffin, M. (2017). Technology to Augment Early Home Visitation for Child Maltreatment Prevention: A Pragmatic Randomized Trial. *Child Maltreatment*, *22*(4), 334–343. https://doi.org/10.1177/1077559517729890

Osman, F., Flacking, R., Klingberg Allvin, M., & Schön, U.-K. (2019). Qualitative study showed that a culturally tailored parenting programme improved the confidence and skills of Somali immigrants. *Acta Paediatrica*, *108*(8), 1482–1490. https://doi.org/10.1111/apa.14788

Pampaka, M., Hutcheson, G., & Williams, J. (2016). Handling missing data: Analysis of a challenging data set using multiple imputation. *International Journal of Research & Method in Education*, *39*(1), 19–37. https://doi.org/10/gmxdvm

Patterson, J. M. (2002). Understanding family resilience. *Journal of Clinical Psychology*, *58*(3), 233–246. https://doi.org/10.1002/jclp.10019

Paul, G. (1967). Strategy of outcome research in psychotherapy. Journal of Consulting Psychology, 31,109–118. doi:10.1037/h0024436.

Pelham, W. E., Dishion, T. J., Tein, J.-Y., Shaw, D. S., & Wilson, M. N. (2017). What Doesn't Work for Whom? Exploring Heterogeneity in Responsiveness to the Family Check-Up in Early Childhood Using a Mixture Model Approach. *Prevention Science*, *18*(8), 911–922. https://doi.org/10.1007/s11121-017-0805-1

Pellerin, K. A., Costa, N. M., Weems, C. F., & Dalton, R. F. (2010). An Examination of Treatment Completers and Non-Completers at a Child and Adolescent Community Mental Health Clinic. *Community Mental Health Journal*, *46*(3), 273–281. https://doi.org/10.1007/s10597-009-9285-5

Peters, S., Calam, R., & Harrington, R. (2005). Maternal attributions and expressed emotion as predictors of attendance at parent management training. *Journal of Child Psychology and Psychiatry, 46*(4), 436-448. https://doi.org/10.1111/j.1469-7610.2004.00365.x

Pinto, R. J., Correia-Santos, P., Levendosky, A., & Jongenelen, I. (2019). Psychological Distress and Posttraumatic Stress Symptoms: The Role of Maternal Satisfaction, Parenting Stress, and Social Support Among Mothers and Children Exposed to Intimate Partner Violence. *Journal of Interpersonal Violence*, *34*(19), 4114–4136. https://doi.org/10.1177/0886260516674199

Placa, V. L., & Corlyon, J. (2014). Barriers to inclusion and successful engagement of parents in mainstream services: Evidence and research. *Journal of Children's Services*, *9*(3), 220–234. https://doi.org/10.1108/jcs-05-2014-0027

Pote, I., Doubell, L., Brims, L., Larbie, J., Stock, L., & Lewing, B. (2019). *Engaging disadvantaged and vulnerable parents An evidence review*. Early Intervention Foundation. [PDF File] Retrieved from: https://www.drugsandalcohol.ie/34423/1/engaging-disadvantaged-vulnerable-parents.pdf.

Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In *Proceedings of the American Statistical Association, Section on Survey Research Methods* (Vol. 225230). Washington, DC: American Statistical Association.

Prochaska, J. O., & DiClemente, C. C. (1994). *The transtheoretical approach: Crossing traditional boundaries of therapy.* Krieger Pub Co.

Raikes, H. H., Vogel, C., & Love, J. M. (2013). Iv. Family Subgroups and Impacts at Ages 2, 3, and 5: Variability by Race/Ethnicity and Demographic Risk. *Monographs of the Society for Research in Child Development*, *78*(1), 64–92. https://doi.org/10.1111/j.1540-5834.2012.00703.x

Redmond, C., Spoth, R., & Trudeau, L. (2002). Family-and community-level predictors of parent support seeking. *Journal of Community Psychology*, *30*(2), 153-171.

Rees, C. A., & Selwyn, J. (2009). Non-infant adoption from care: Lessons for safeguarding children. *Child: Care, Health and Development*, *35*(4), 561–567. https://doi.org/10.1111/j.1365-2214.2009.00978.x

Reid, M. J., Webster-Stratton, C., & Beauchaine, T. P. (2001). Parent Training in Head Start: A Comparison of Program Response Among African American, Asian

American, Caucasian, and Hispanic Mothers. *Prevention Science*, *2*(4), 209–227. https://doi.org/10.1023/A:1013618309070

Reyno, S. M., & McGrath, P. J. (2006). Predictors of parent training efficacy for child externalizing behavior problems—A meta-analytic review. *Journal of Child Psychology and Psychiatry*, *47*(1), 99–111. https://doi.org/10.1111/j.1469-7610.2005.01544.x

Richmond, M. K., Pampel, F. C., Zarcula, F., Howey, V., & McChesney, B. (2017). Reliability of the Colorado family support assessment: A self-sufficiency matrix for families. *Research on Social Work Practice, 27*(6), 695-703. https://doi.org/10.1177/1049731515596072

Rutherford, H. J., & Mayes, L. C. (2019). Parenting stress: A novel mechanism of addiction vulnerability. *Neurobiology of Stress*, *11*, 100172. https://doi.org/10.1016/j.ynstr.2019.100172

Roggman, L., Cook, G., Peterson, C., & Raikes, H. (2008). Who Drops Out of Early Head Start Home Visiting Programs? *EARLY EDUCATION AND DEVELOPMENT*, *19*, 574–599. https://doi.org/10/d7hqh9

Sama-Miller, E., & Baumgartner, S. (2017). *Features of Programs Designed to Help Families Achieve Economic Security and Promote Child Well-being*. 14.

Santiago, C. D., Kaltman, S., & Miranda, J. (2013). Poverty and Mental Health: How Do Low-Income Adults and Children Fare in Psychotherapy? *Journal of Clinical Psychology*, *69*(2), 115–126. https://doi.org/10.1002/jclp.21951

Schafer, J. L., & Olsen, M. K. (1998). Multiple Imputation for Multivariate Missing-Data

Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research*, *33*(4), 545–571. https://doi.org/10/b852p3

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, *57*(1), 1–10. https://doi.org/10/dn8qzd

Schmidt, S. C., & Woll, A. (2017). Longitudinal drop-out and weighting against its bias. *BMC medical research methodology*, *17*(1), 164.

Schneider, B., Gerdes, A., Haack, L., & Lawton, K. (2013). Predicting Treatment Dropout in Parent Training Interventions for Families of School-Aged Children With ADHD. *Child & Family Behavior Therapy, 35*(2), 144-169. https://doi.org/10.1080/07317107.2013.789365

Seaman, S. R., White, I. R., Copas, A. J., & Li, L. (2012). Combining Multiple Imputation and Inverse-Probability Weighting. *Biometrics*, *68*(1), 129–137. https://doi.org/10/dsg9h2

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, *22*(3), 278–295. https://doi.org/10/dm22q8

Seaman, S., & White, I. (2014). Inverse Probability Weighting with Missing Predictors of Treatment Assignment or Missingness. *Communications in Statistics - Theory and Methods*, *43*(16), 3499–3515. https://doi.org/10/gmtk4d

Seltman, H. J. (2015). Experimental design and analysis. [*PDF File*] Retrieved from http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf

Shonkoff, J. P., Boyce, W. T., & McEwen, B. S. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities: building a new framework for health promotion and disease prevention. *JAMA, 301*(21), 2252-2259. doi:10.1001/jama.2009.754

Shonkoff, J. P., Garner, A. S., Siegel, B. S., Dobbins, M. I., Earls, M. F., McGuinn, L., ... & Committee on Early Childhood, Adoption, and Dependent Care. (2012). The l ifelong effects of early childhood adversity and toxic stress. *Pediatrics, 129*(1), e232-e246. doi:10.1542/peds.2011-2663

Sidebotham, P., Heron, J., & ALSPAC Study Team. (2006). Child maltreatment in the "children of the nineties": A cohort study of risk factors. *Child Abuse & Neglect, 30*(5), 497-522. https://doi.org/10.1016/j.chiabu.2005.11.005

Simon, J. D., Gandarilla Ocampo, M., Drake, B., & Jonson-Reid, M. (2021). A Review of Screened-Out Families and Child Protective Services Involvement: A Missed Opportunity to Prevent Future Maltreatment With Community-Based Services. *Child Maltreatment*, 10775595211033596. https://doi.org/10.1177/10775595211033597

Smith, J. D., Berkel, C., Hails, K. A., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2018). Predictors of Participation in the Family Check-Up Program: A Randomized Trial of Yearly Services from Age 2 to 10 Years. *Prevention Science : The Official Journal of the Society for Prevention Research*, *19*(5), 652–662. https://doi.org/10.1007/s11121-016-0679-7

Snyder, M. A. (2019). *Colorado Office of Early Childhood*. Powerpoint Presentation.

Retrieved from: https://leg.colorado.gov/sites/default/files/images/office_of_early_childhood_0.pdf.

So, M., Almeida Rojo, A. L., Robinson, L. R., Hartwig, S. A., Heggs Lee, A. R., Beasley, L. O., Silovsky, J. F., Morris, A. S., Stiller Titchener, K., & Zapata, M. I. (2020). Parent engagement in an original and culturally adapted evidence-based parenting program, Legacy for Children™. *Infant Mental Health Journal*, *41*(3), 356–377. https://doi.org/10.1002/imhj.21853

Stewart, M. J., Makwarimba, E., Reutter, L. I., Veenstra, G., Raphael, D., & Love, R. (2009). Poverty, Sense of Belonging and Experiences of Social Isolation. *Journal of Poverty*, *13*(2), 173–195. https://doi.org/10.1080/10875540902841762

Stith, S. M., Liu, T., Davies, L. C., Boykin, E. L., Alder, M. C., Harris, J. M., ... & Dees, J. E. M. E. G. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. *Aggression and Violent Behavior, 14*(1), 13-29. https://doi.org/10.1016/j.avb.2006.03.006

Supplee, L. H., Ammerman, R. T., Duggan, A. K., List, J. A., & Suskind, D. (2021). The Role of Open Science Practices in Scaling Evidence-Based Prevention Programs. *Prevention Science*. https://doi.org/10.1007/s11121-021-01322-8

Supplee, L. H., & Duggan, A. (2019). Innovative Research Methods to Advance Precision in Home Visiting for More Efficient and Effective Programs. *Child Development Perspectives*, *13*(3), 173–179. https://doi.org/10.1111/cdep.12334

Sweet, M. A., & Appelbaum, M. I. (2004). Is Home Visiting an Effective Strategy? A Meta-Analytic Review of Home Visiting Programs for Families With Young Children. *Child Development*, *75*(5), 1435–1456. https://doi.org/10.1111/j.1467-8624.2004.00750.x

Tarren-Sweeney, M., & Goemans, A. (2019). A narrative review of stability and change in the mental health of children who grow up in family-based out-of-home care. *Developmental Child Welfare*, *1*(3), 273–294. https://doi.org/10.1177/2516103219874810

Taylor, C. A., Guterman, N. B., Lee, S. J., & Rathouz, P. J. (2009). Intimate partner violence, maternal stress, nativity, and risk for maternal maltreatment of young children. *American Journal of Public Health, 99*(1), 175-183. DOI: 10.2105/AJPH.2007.126722

The Family Center / La Familia. (n.d.). The Family Center/La Familia. Retrieved April 25, 2022, from https://thefamilycenterfc.org/

The Home Visiting Applied Research Collaborative. (2018). *HARC-2018_Introduction-to-Precision-Home-Visiting.pdf*. http://www.hvresearch.org/wp-content/uploads/2019/09/HARC-2018_Introduction-to-Precision-Home-Visiting.pdf

Thomas, C., Hypponen, E., Power, C. (2007). Prenatal exposures and glucose metabolism in adulthood. Diabetes Care 30:918–924.

Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing

data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, *110*, 63–73. https://doi.org/10/gfx5hq

U.S. DHHS (1990) *Head Start Research and Evaluation: a Blueprint for the Future. Recommendations of the Advisory Panel for the Head Start Evaluation Design Project.* Overall Strategy and General Principles. Publication no. ACY 91-31195.

U.S. Department of Health & Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau. (2019). *Child Maltreatment 2017.* Available from https://www.acf.hhs.gov/cb/research-data-technology/statistics-research/child-maltreatment.

U.S. Department of Health & Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau. (2022). *Child Maltreatment 2020*. Available from https://www.acf.hhs.gov/cb/data-research/child-maltreatment.

Victor, B. G., Grogan-Kaylor, A., Ryan, J. P., Perron, B. E., & Gilbert, T. T. (2018). Domestic violence, parental substance misuse and the decision to substantiate child maltreatment. *Child Abuse & Neglect*, *79*, 31–41. https://doi.org/10.1016/j.chiabu.2018.01.030

Wagner, M. M., & Clayton, S. L. (1999). The Parents as Teachers program: Results from two demonstrations. *Future of Children, 9*, 91-115. [*PDF File*] Retrieved from: https://pdfs.semanticscholar.org/fb6f/3b65b43281f557909ecfe44b761ab52ef35d.pdf

Wagner, M., Spiker, D., Inman Linn, M., & Hernandez, F. (2003). Dimensions of Parental Engagement in Home Visiting Programs: Exploratory Study. *Topics in Early Childhood Special Education, 23*(4), 171–187. https://doi.org/10.1177/02711214030230040101

Whittaker, K. A., & Cowley, S. (2012). An effective programme is not enough: A review of factors associated with poor attendance and engagement with parenting support programmes. *Children & Society*, *26*(2), 138–149. https://doi.org/10.1111/j.1099-0860.2010.00333.x

West, A., Duggan, A., Gruss, K., & Minkovitz, C. (2021). Service Coordination to Address Maternal Mental Health, Partner Violence, and Substance Use: Findings from a National Survey of Home Visiting Programs. *Prevention Science*, *22*(5), 633–644. https://doi.org/10.1007/s11121-021-01232-9

Zhang, Z. (2016). Model building strategy for logistic regression: purposeful selection. *Annals of translational medicine*, *4*(6). doi: 10.21037/atm.2016.02.15

Zill, N., Resnick, G., Kim, K., O'Donnell, K., Sorongon, A., McKey, R. H., ... & D'Elio, M. A. (2003). Head Start FACES 2000: A Whole-Child Perspective on Program Performance. Fourth Progress Report

Appendix A. Data Preparation

Duplicate records. Data were exported directly from administrative record-keeping systems (e.g. Salesforce) and FDS and OEC had limited capacity to address data quality questions. As a result, files were prepared for initial data cleaning through a series of logic steps. Both FDS and the OEC had a significant number of duplicate records, defined as any participant record that shared a date of assessment with another record or any participant record in which more than one record were undated and unlabeled or more than one record was undated with duplicate labels (e.g. undated and baseline-baseline or post-post). When choosing the record to retain as primary, data completeness on the CFSA and PFS were first individually examined. In the most common scenario, duplicates had identical data to each other, in which case either record were retained and the other removed. Prior to removing duplicates with identical data, variables means were compared to ensure data were identical between cases listed as primary and duplicate. In the next common scenario, duplicates were entirely blank or missing significantly more data outside of the shared assessment date. Records with more complete data were labeled as primary and duplicates removed. If data were mismatched with no difference in level of completeness, both records were discarded or partial records were discarded (the CFSA and the PFS were considered separately), as there was no way to determine which record accurately belonged to the primary caregiver.

When removing duplicate records, the vast majority were captured within the logic rules explained above. A few exceptional situations with duplicates occurred. In 22 cases (10 from FDS and 12 from the OEC), the first two timepoints between CFSAs ranged between 1 to 6 days. Given that none of the investigated programs have treatment lengths that are designed to be implemented under one week, these cases were treated as duplicates. Each record was examined for completeness; records with more complete data were retained as timepoint 1 (in 7 cases, the second timepoint had more complete data). When records had similar levels of completeness, the earlier date's record was retained. When a third timepoint was available, it was moved to the position of timepoint 2 (10 cases).

In the OEC file, two cases were entirely removed due to duplicates: first, baseline duplicates that were equally complete and mismatched were removed, per the logic rules outlined above. These same cases each had two sets of duplicate (identical) post records that were undated. Given the inability to order the subsequent two timepoints accurately, all 'post' records were removed on these two cases. Thirty-four records in the OEC file were removed due to a missing participant ID (could not link the CFSA record to any demographics). Five cases were removed that were missing both demographics, CFSA, and PSF records. In FDS file of CFSA record entries, total records (long-form of each CFSA) were reduced from 34,081 records to 23,198 usable records (note: the number of CFSA records attached to a single ID ranged up to 160 records). In 16 cases (11 FRCA, 5 HS) no CFSA or PSF data was provided on dates of record and were removed. In the OEC file of CFSA record entries, total records were reduced from 9,908 to 9,503. HS had

devoted capacity to address data quality questions, eliminating the need to make independent determinations around which records to retain ahead of data cleaning.

Ordering of timepoints. Cases with label and date discrepancies were examined individually for the OEC and FDS to determine which record constituted timepoint 1 vs. later timepoints. In the FDS files, CFSA labeling (e.g. 'baseline' vs. 'post') were normally aligned with dates. Cases labeled as 'baseline' were, as a rule, moved to the position of timepoint 1 and any misaligned dates on subsequent timepoints were examined by hand. In instances where dates and labels were misaligned, dates were examined individually for evidence of data entry mistakes (e.g. date of administration listed after the date the researcher received the data file or more than a year prior to the listed baseline and intake date). When data entry appeared to be the cause of a discrepancy on a date (as opposed to the ordering of timepoints), dates were removed (n=36). This means that on a subset of cases, there is no determination of length of time between timepoints 1 and 2. When dates appeared to be accurate, they took precedent over the use of the 'baseline' label.

In the OEC file, there was significant inconsistency between the use of labeling and ordering of dates. As a result, dates were used as a primary mechanism to order timepoints and labels (e.g. 'baseline' vs. 'post') were used secondarily when dates were unavailable. In 12 cases, more than two records were available with dates along with a stand-alone undated record. Each of these cases were examined by hand alongside accompanying available information on the case's intake date (last date of face-to-face contact, and completion date), in order to determine the most likely timepoint position for

the undated CFSA. In cases where a CFSA record was both undated and unlabeled, information from accompanying records for that participant ID were used to position the record. For example, in nine cases, an undated and unlabeled record accompanied a dated record labeled as 'follow-up'. Given the decision to utilize labeling when dates were unavailable, the unlabeled/undated record was positioned as timepoint 1. In two cases, accompanying available information on case dates (intake, last face-to-face contact, and completion date) were used to determine a data entry error specific to the year and the year was updated by hand. In one case, a participant had two unlabeled records and only one of which had a date. In this case, accompanying available information on case dates was used to position the dated record as a baseline assessment.

Demographics. In the OEC cases, participant demographics were generally identified through contact designations as 'primary caregiver' or 'primary child'. However, in many instances, contact designations were missing, but identification as a 'primary contact' was available. Primary contacts were not designated as adults or children; however, when their age was available, it was used to generate parent or child age. When primary contacts were between the ages of 14 and 19 (with no other designation as to being primary caregiver vs. child), participants aged 14 and over were coded as caregivers when no other adult contacts in the household could be identified (n=22, OEC cases) and as children when any adult contacts in the household over the age of 21 could be identified (n=100). If multiple adult contacts were in the home, the adolescent primary contact was coded as a child, but no primary adult caregiver was designated (n=36). If a single adult caregiver was in the home, the adolescent primary contact was coded as a

child and the adult caregiver was re-coded as primary and use for demographic characteristics when available (n =64). When coding demographics outside of age (family language, ethnicity, race), primary caregiver information was used when available. If there was no primary caregiver designation or if the demographic was missing or 'NA', data were pulled from the primary contact's record when available. If the primary contact's record was missing or 'NA', data were lastly pulled from 'primary child' records, except in the case of gender (child gender was not used as a proxy for caregiver gender). In total, the OEC's race variable (n=1866) was comprised of 921 contacts designated as primary caregiver, 927 contacts designated as 'primary' contacts, and finally, on 18 cases designated as primary child. The OEC's ethnicity variable (n=1736) was comprised of 874 contacts designated as primary caregivers, 845 contacts designated as 'primary', and 17 cases designated as primary child. The language variable (n=3782) was comprised of 2,147 contacts designated as primary caregiver, 1,607 contacts designated as 'primary' contacts, and 28 cases designated as primary child. The gender variable (n= 3,356) was comprised of 2,247 contacts designated as primary caregivers and 1,109 contacts designated as 'primary contacts' when parent age was over 14. This approach may pose some limitation as it's possible that the primary child's race, ethnicity, and language would have differed from the primary caregiver. However, the needs of the family may still be impacted by the child's race, language, or ethnicity designation and this accounted for a very small proportion of available data (less than 1%).

Appendix B. Missing Data

The 14 domains of the CFSA were analyzed for missingness prior to creating subscales. In many cases, data were classified as missing due to valid response options embedded in the CFSA. Three domains (employment, child care, and child education) had a response option of 'not applicable' which indicated that 'all adults in the family are not employable' (employment domain), 'family does not have children under 12 years old, children are in someone else's care (e.g. foster care), or the family is adequately able to care for children and does not need child care' (child care domain), and 'all children in the family are not school-aged or they have earned a GED' (child education domain). On these three domains, a significant proportion of missing data across the full sample was tied to the 'not applicable' response

Missingness related to the "not applicable" response option or to the inability to collect sufficient information from the participant on the Child Care Domain accounted for 19.7% of missing data at timepoints 1 and 2; on the Child Education Domain accounted for 8.3% of missing data at timepoint 1 and 7.9% of missing data at timepoint 2. However, the levels of missingness connected to these valid response options would have likely been much higher if FDS had differentiated 'not applicable' response options from other forms of missing data (FDS has missingness on the Child Care domain of 47.7% and 53.6% and on the Child Education domain of 33.7% and 38.5% at timepoint

1 and 2, respectively). Without the ability to separate types of missingness from FDS, total missing data in the sample for Child Care was 29.8% and 36.2% and for Child Education was 20.7% and 24.1% at timepoints 1 and 2, respectively. However, in FSPs outside of FDS who had reported "not applicable" response options, rates of Child Care domain missingness ranged from 0% (Safe CMP) to 2.9% (CCR) at timepoint 1 and 0% (PSSF, Safe CMP) to 9.8% (CCR) at timepoint 2. Rates of Child Education domain missingness varied from 0% (Safe CMP) to 1.4% (Head Start) at timepoint 1 and 0% (PSSF, Safe CMP) to 3.9% (Head Start) at timepoint 2. Data were determined to be NMAR in the employment domain, due to a valid 'not applicable' response option (indicating 'all adults in the house are not employable').

A small proportion of missing data on all 14 domains was connected to the response option of 'not enough information', which was used when interviewers were not able to obtain sufficient information during a family interview to provide a score. Given the applied context of data collection, difficulty assessing domains were expected and typically accounted for a low proportion of missing data. An additional small proportion of missing data resulted from provider data entry errors (use of scores that fell outside of the measure instructions, e.g. '8' on a scale of 1-5). While some programs had no data entry errors, other programs (the OEC) occasionally reported scores of 7, 8, and 9 (invalid on the scale of 1-5). Finally, data which were missing beyond these categories on the CFSA timepoint 1 for the full sample ranged from a low of 2.5% (Food and Transportation domains) to a high of 29.8% on the Child Care domain. On the CFSA 2, missing data ranged from a low of 8.8% on the Transportation Domain to a high of

36.2% on the Child Care domain. Missing data on the PFS in the full sample at timepoint 1 ranged from a low of 19.2% missing on the Social Support subscale to a high of 24.8% missing on item 15 ('I praise my child when he/she behaves well'). At timepoint 2, missing data ranged from a low of 27.5% on the first three subscales (family functioning/resiliency, social support, and concrete support) to a high of 30.8% missing on item 15.

When examined on a program-level and including all forms of missingness (e.g. 'not applicable' and 'not enough information' responses), missing CFSA data in the FDS at timepoint 1 ranged from 3.2% on Transportation and Food Domains to 47.7% on the Child Care Domain. At timepoint 2, missing data ranged from a low of 12.4% on Food and Transportation domains to a high of 53.6% on Child Care. Missing data on the PFS at timepoint 1 ranged from a low of 12% on subscales 2 and 3 (social support and concrete support) to a high of 18.6% missing on item 15 ('I praise my child when he/she behaves well'). Missing data at timepoint 2 ranged from a low of 26.5% on subscales 1, 2, and 3 (family functioning/resiliency, social support, and concrete support) to a high of 31.4% missing on item 15.

In HS at timepoint 1 missing data was 1.4% across all domains and at timepoint 2 was 3.9% across all domains. Missing data on the PFS at timepoint 1 ranged from a low of 0.4% on all items other than items 13, 14, and 16 with 0.6% missing. Missing data at timepoint 2 ranged from a low of 6.7% on subscales 1, 2, and 3 to a high of 7.6% missing on item 12 ('There are many times when I don't know what to do as a parent').

In CCR at timepoint 1 missing data ranged from 0.4% on the Employment Domain to 4.6% on Family Income. At timepoint 2, missing data ranged from 0% on Employment to 9.8% on Child Care. Missing data on the PFS at timepoint 1 ranged from a low of 32.1% missing subscale 2 to a high of 32.8% missing on items 13 and 14. Missing data at timepoint 2 ranged from a low of 35% on subscales 1 and 2 to a high of 35.7% missing on items 14 and 16.

In PSSF at timepoint 1 missing data ranged from 0.1% on Employment, Child Care, and Child Education Domains to a high of 7% on Substance Use. At timepoint 2, missing data ranged from no missing data on Employment, Child Care, and Child Education to a high of 7.2% on Substance Use. Missing data on the PFS at timepoint 1 ranged from a low of 13.5% missing on subscales 1, 2, and 3 to a high of 22.5% missing on item 12. Missing data at timepoint 2 ranged from a low of 24% on subscale 2 to a high of 29.3% on item 12.

In Safe CMP at timepoint 1, missing data ranged from 0% in seven domains to a high of 10.8% on Housing. At timepoint 2, missing data ranged from 0% in nine domains to a high of 11% on Housing. Missing data on the PFS at timepoint 1 ranged from a low of 1.8% missing on subscales 1, 2, and 3 to a high of 3.1% missing on item 15. Missing data at timepoint 2 ranged from a low of 4.3% on subscales 1, 2, 3, and 4 and item 13 to a high of 5.5% on item 16.

Missing data on the CFSA was tested using Little's MCAR test separately within each FSP if missing data exceeded 5% on any of the domains. The test was conducted on each timepoint separately and on both timepoints together, when available. Missing data

in FDS, CCR, PSSF, and Safe CMP were identified as MAR in all instances. Missing data on the PFS was tested using Little's MCAR test separately within each FSP if missing data exceeded 5% on any of the domains. The test was conducted on each timepoint separately and on both timepoints, when available. Missing data in FDS, CCR, PSSF, and Safe CMP were identified as MAR in all instances. HS was MCAR. Missingness on the PFS protective factor subscales at timepoint 2 was associated with the predictors in Table 1.

Table 1. Associations between missing data on the PFS and predictors

	Timepoint 1	Timepoint 2
PFS Subscale 1: Family Functioning and Resilience		
English as primary language	$\chi 2$ (1, N=7100)=70.7, p=<.001	$\chi 2$ (1,5031)=59.1, p=<.001
White race family	$\chi 2$ (1, N=7378)=64.7, p=<.001	$\chi 2$ (1,4907)=8.7, p=<.01
Non-Latinx ethnicity	$\chi 2$ (1, N=10242)=154.2, p=<.001	$\chi 2$ (1,6880)=52.7, p=<.001
High school/college degree	$\chi 2$ (3, N=14335)=27.8, p=<.001	--
Fathers	$\chi 2$ (1, N=11655)=21.6, p=<.001	$\chi 2$ (1,7676)=19.7, p=<.001
Single-parent household	--	$\chi 2$ (1,3540)=4.5, p=<.05
Lower income	t(4363)=3.3, p<.01	t(5232)=2.9, p<.01
Older parent age	t(2287)=-12.1, p<.001	t(7291)=-4.3, p<.001
Smaller household size	t(14680)=6.5, p<.001	t(4630)=4.7, p<.001
PFS Subscale 2: Social Support		
English as primary language	$\chi 2$ (1, N=7100)=71.2, p=<.001	$\chi 2$ (1,5031)=60.4, p=<.001
White race family	$\chi 2$ (1, N=7378)=62.3, p=<.001	$\chi 2$ (1,4907)=8.5, p=<.01
Non-Latinx ethnicity	$\chi 2$ (1, N=10242)=156.6, p=<.001	$\chi 2$ (1,6880)=52.4, p=<.001
High school/college degree	$\chi 2$ (3, N=14335)=30.9, p=<.001	--

Fathers	$\chi 2$ (1, N=11655)=22.4, p=<.001	$\chi 2$ (1,7676)=19.3, p=<.001
Single-parent household	--	$\chi 2$ (1,3540)=4.4, p=<.05
Lower income	*t*(4319)=3.2, p<.01	*t*(5217)=2.7, p<.01
Older parent age	*t*(2264)=-11.5, p<.001	*t*(7291)=-4.2, p<.001
Smaller household size	*t*(14680)=6.4, p<.001	*t*(4629)=4.7, p<.001
PFS Subscale 3: Concrete Support		
English as primary language	$\chi 2$ (1, N=7100)=71.7, p=<.001	$\chi 2$ (1,5031)=61.8, p=<.001
White race family	$\chi 2$ (1, N=7378)=62.3, p=<.001	$\chi 2$ (1,4907)=8.9, p=<.01
Non-Latinx ethnicity	$\chi 2$ (1, N=10242)=155.7, p=<.001	$\chi 2$ (1,6880)=52.5, p=<.001
High school/college degree	$\chi 2$ (3, N=14335)=30.7, p=<.001	--
Fathers	$\chi 2$ (1, N=11655)=22.2, p=<.001	$\chi 2$ (1,7676)=19.5, p=<.001
Single-parent household	--	$\chi 2$ (1,3540)=4.2, p=<.05
Lower income	*t*(4326)=3.3, p<.01	*t*(5208)=2.8, p<.01
Older parent age	*t*(2267)=-11.4, p<.001	*t*(7291)=-4.1, p<.001
Smaller household size	*t*(14680)=6.3, p<.001	*t*(4615)=4.6, p<.001
PFS Subscale 4: Nurturing and Attachment		
English as primary language	$\chi 2$ (1, N=7100)=33.9, p=<.001	$\chi 2$ (1,5031)=30.8, p=<.001
White race family	$\chi 2$ (1, N=7378)=16.1, p=<.001	$\chi 2$ (1,4907)=7.4, p=<.01
Non-Latinx ethnicity	$\chi 2$ (1, N=10242)=29.8, p=<.001	$\chi 2$ (1,6880)=10.4, p=<.01
High school/college degree	--	--
Single-parent household	--	$\chi 2$ (1,7676)=19.5, p=<.001
Fathers	$\chi 2$ (1, N=11655)=25.6, p=<.001	$\chi 2$ (1,7676)=20.9, p=<.001
Lower income	*t*(6303)=5.0, p<.001	*t*(6196)=2.9, p<.01
Older parent age	*t*(3420)=-8.4, p<.001	*t*(7291)=-4.0, p<.001
Smaller household size	*t*(5646)=11.2, p<.001	*t*(5319)=5.5, p<.001

Appendix C. Marginal Main Effects.

Self-Sufficiency

A marginal effect of readiness to change was found, *F* (1, 6967)=[3.6-7], [*p*<.10-*p*<.01], ηp2=.001, such that higher readiness was linked with greater improvements.

Child Education

A marginal effect of parent education was found, F(1,2775)=3.8, p=.05, ηp2=.001, such that families with less education made greater improvements.

Substance Abuse

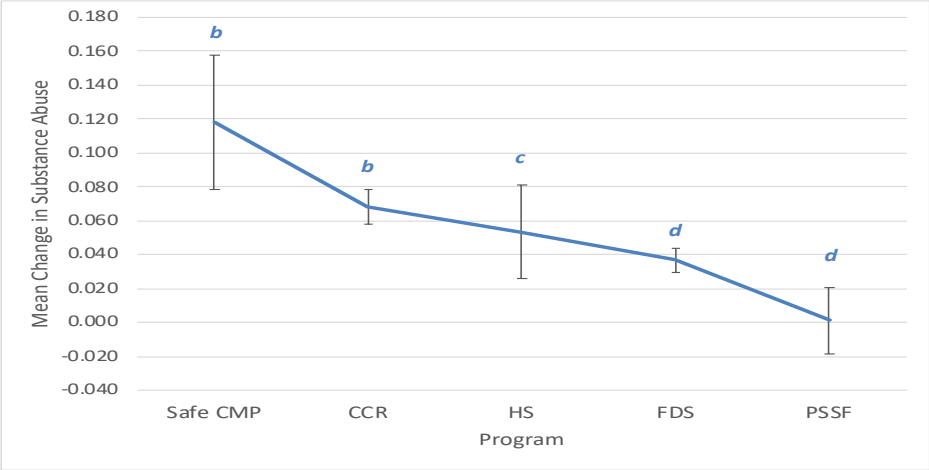


Figure 1. Mean Substance Abuse Improvements by Program

Note: a=significantly higher (p<.05) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three or more programs.

In a recalculated model that included Safe CMP and dropped readiness to change, a main effect of program was identified (*Figure 1*), *F*(4,12279)=[Range=3-5.9], [*p*<.05-*p*<.001], [ηp2=.001-.002], such that enrollment in Safe CMP (M=.12, SE=.04) and CCR (M=.07, SE=.01) were linked

with greater improvements on average than enrollment in FDS (M=.04, SE=.01) or PSSF (M=.00, SE=.02). Main effects of stress, $F(4,12279)$=[Range=47.4-73.5], p<.001, [ηp2=.004-.006], and starting score, $F(4,12279)$=[Range=4601.7-4792.5], p<.001, [ηp2=.26-.27], were identified such that greater levels of each were linked with fewer gains.

Family Functioning and Resilience

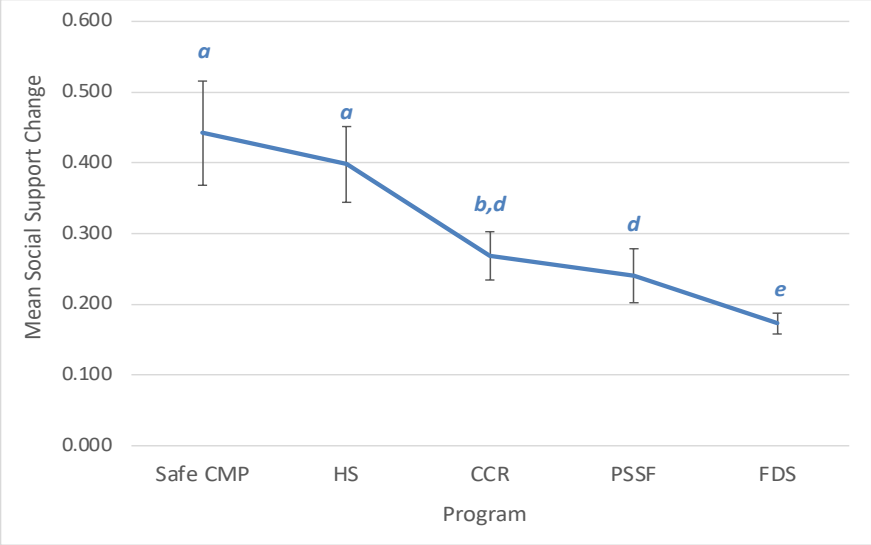


Figure 2. Mean Social Support Change by Program

Note: a=significantly higher (p<.05) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three or more programs.

A marginal main effect of household structure was found, $F(1,7311)$=[1.4-10.7], [p=.23-p<.01], [ηp2=.001-.004], such that single-parent households had greater improvements (M=.18, SE=.06) than two-parent households (M=.06, SE=.05).

Social Support

In a secondary model with Safe CMP (dropping the readiness to change covariate), a marginal main effect of program was observed (*Figure 2*), $F(1, 12804)$=[1.4-6.6], [p=.25-

p<.001], [ηp2=.000-.002], such that enrollment in Safe CMP (M=.44, SE=.07) and HS (M=.40, SE=.05) were linked with greater improvements than CCR (M=.27, SE=.03), PSSF (M=.24, SE=.04), and FDS (M=.17, SE=.17).

A marginal main effect of parent education was observed, *F*(1, 12804)=[2.4-10.4], [*p*=.12-*p*<.01], [ηp2=.000-.001], such that families with greater education made greater improvements. Starting score was also significant in the updated model,

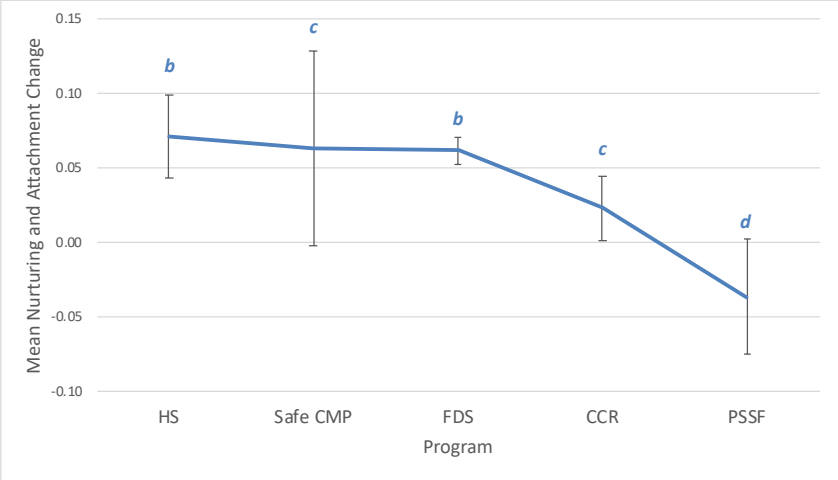


Figure 3. Mean Nurturing and Attachment Improvements by Program

Note: a=significantly higher (p<.05) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three or more programs.

F(1,12804)=[3934.6-5594.4], *p*<.001, [ηp2=.29-.30], such that greater starting scores were linked with fewer improvements.

Nurturing and Attachment

A marginal main effect of program was identified (see *Figure 3*), F(1,9383)=[1.9-9.2], [*p*=.11-*p*<.001], [ηp2=.001-.004], such that families enrolled in HS (M=.07, SE=.03) and

FDS (M=.06, SE=.01), had greater improvements on average than those enrolled in PSSF (M=.-.04, SE=.04).

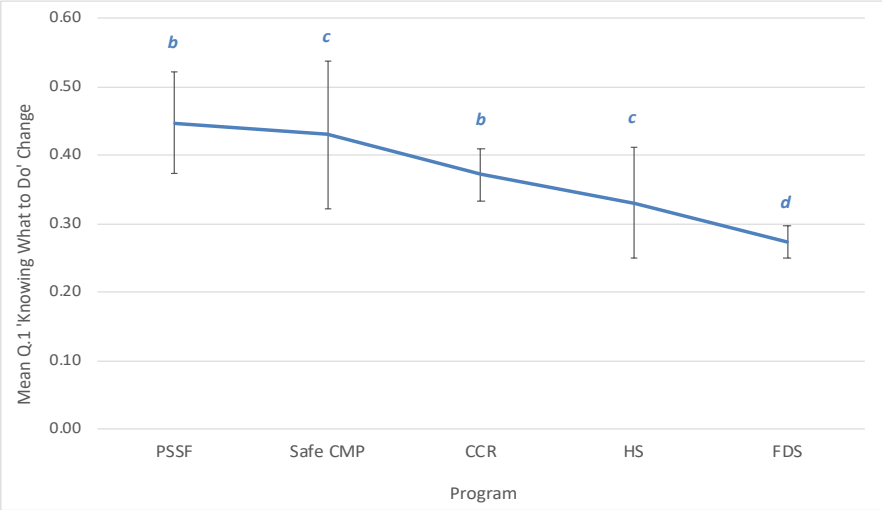


Figure 4. Mean Q1. 'Knowing What to Do' Change by Program

Note: a=significantly higher (p<.05) than three or more programs, b=significantly higher than one or more programs, c=no differences with any programs, d=significantly lower than one or more programs, e=significantly lower than three or more programs.

Knowledge of Child Development and Parenting Skills

Question 1. I know what to do as a parent. A marginal main effect of program was found (*Figure 4*), *F*(4,13094)=[1.0-6.4], [*p*=.39-*p*<.001], [ηp2=.000-.002], such that families enrolled in PSSF (M=.45, SE=.07) and CCR (M=.37, SE=.04) had significantly greater improvements than families enrolled in FDS (M=.27, SE=.02).

Question 3. My child does not misbehave to upset me. A marginal main effect of stress was found, *F*(4,12985)=[2.0-5.5], [*p*=.15-*p*<.05], ηp2=.000, such that greater stress was linked with fewer gains.

Question 5. I do not lose control when I discipline my child. A marginal main effect of stress was identified, *F*(1,13094)=[2.2-8.2], [*p*=.14- *p*<.01], [ηp2=.000-.001], such that greater stress was linked with greater gains.

Appendix D. Additional Results Tables

Table 1. Baseline Differences by Caregiver Gender, Language, and Race

	Gender			Family Language			Race		
Baseline Scores	**Female** M (SD)	**Male** M (SD)	*t*(df)	**EPL** M (SD)	**ESL** M (SD)	*t*(df)	**White** M (SD)	**BIPOC** M (SD)	*t*(df)
Economic SS	2.88 (0.75)	2.94 (0.78)	-3.58 (2897)***	2.92 (0.77)	2.94 (0.63)	-0.95 (1698)	2.96 (0.79)	2.81 (0.71)	7.54 (2854)***
M/P Health	4.03 (1.04)	3.99 (1.07)	1.56 (13107)	3.91 (1.06)	4.36 (0.88)	-15.02 (1685)***	3.96 (1.07)	4.13 (1.01)	-6.00 (2753)***
Debt M	2.84 (1.54)	2.79 (1.5)	1.37 (2817)	2.67 (1.5)	3.6 (1.39)	-19.68 (1522)***	2.74 (1.51)	2.73 (1.58)	0.25 (2249)
Child Care	3.56 (1.44)	3.65 (1.39)	-2.08 (1581)*	3.56 (1.42)	3.73 (1.37)	-2.92 (824)**	3.59 (1.42)	3.41 (1.42)	3.50 (4895)***
Child Ed	3.79 (1.23)	3.81 (1.22)	-0.45 (9506)	3.69 (1.24)	4.05 (1.09)	-8.90 (1391)***	3.78 (1.23)	3.87 (1.25)	-2.32 (6079)*
Sub Abuse	4.72 (0.82)	4.64 (0.92)	3.92 (2662)***	4.66 (0.89)	4.84 (0.66)	-7.67 (1784)***	4.74 (0.79)	4.63 (0.97)	4.16 (2170)***
FF/Res	4.47 (1.21)	4.42 (1.13)	1.78 (2582)	4.4 (1.18)	4.68 (1.28)	-6.38 (1347)***	4.45 (1.16)	4.42 (1.21)	1.08 (2457)
Soc Sup	4.56 (1.49)	4.41 (1.48)	3.82 (11138)***	4.56 (1.46)	4.57 (1.6)	-0.11 (1346)	4.6 (1.45)	4.41 (1.49)	4.78 (7528)***
Conc Sup	4.12 (1.68)	4.18 (1.59)	-1.36 (2564)	4.23 (1.6)	3.6 (1.99)	9.49 (1266)***	4.27 (1.62)	4.15 (1.6)	2.72 (7530)**
Nur/AA	5.31 (0.83)	5.23 (0.88)	3.23 (2194)**	5.25 (0.87)	5.42 (0.75)	-6.27 (1454)***	5.26 (0.84)	5.37 (0.81)	-4.75 (2378)***
Knowledge of Parenting/Child Development									
Question 1:	3.63 (1.97)	3.73 (1.91)	-1.85 (2297)	3.71 (1.89)	3.0 (2.18)	9.49 (1264)***	3.64 (1.91)	3.85 (1.94)	-3.80 (7138)***
Question 2:	4.68 (1.52)	4.63 (1.58)	1.24 (2220)	4.68 (1.52)	4.69 (1.58)	-0.20 (1343)	4.69 (1.47)	4.69 (1.67)	0.09 (2106)
Question 3:	4.33 (1.89)	4.38 (1.82)	-0.98 (2310)	4.25 (1.9)	4.39 (1.94)	-2.16 (6349)*	4.28 (1.88)	4.44 (1.82)	-3.04 (2350)**
Question 4:	5.22 (1.11)	5.16 (1.07)	1.98 (10303)*	5.29 (1.03)	4.86 (1.45)	8.68 (1130)***	5.27 (1.03)	5.27 (1.12)	-0.06 (2114)
Question 5:	5.1 (1.33)	5.14 (1.28)	-1.14 (2318)	5.15 (1.25)	4.73 (1.59)	7.67 (1222)***	5.12 (1.27)	5.28 (1.23)	-4.23 (7133)***

Note: Lower baseline scores indicate greater need. Economic SS=Economic Self-Sufficiency, M/P Health=Mental and Physical Health, Debt M=Debt Management, Child Ed=Child Education, Sub Abuse=Substance Abuse, FF/Res=Family Functioning and Resilience, Soc Sup=Social Support, Conc Sup=Concrete Support, Nur/Att=Nurturing and Attachment**p<.05, **p<.01, ***p<.001*

Table 2. Baseline Differences by Caregiver Ethnicity and Household Structure

	Ethnicity			Household Structure		
Family Development Baseline Scores	**Non-Latinx** M (SD)	**Latinx** M (SD)	*t*(df)	**Two-parent** M (SD)	**Single-parent** M (SD)	*t*(df)
Economic Self-Sufficiency	2.92 (0.8)	2.83 (0.68)	6.42 (11166)***	3.11 (0.8)	2.74 (0.63)	17.86(4437)***
Mental/Physical Health	3.93 (1.07)	4.22 (0.96)	-15.07 (11059)***	4.12 (1.02)	3.97 (1.02)	4.83(4769)***
Debt Management	2.64 (1.5)	3.11 (1.55)	-15.80 (10725)***	2.92 (1.45)	2.6 (1.56)	7.15(4551)***
Child Care	3.51 (1.43)	3.67 (1.41)	-4.59 (6448)***	3.77 (1.37)	3.47 (1.44)	5.69(2809)***
Child Education	3.74 (1.25)	3.94 (1.18)	-7.32 (8073)***	3.9 (1.19)	3.82 (1.19)	2.09(3484)*
Substance Abuse	4.71 (0.83)	4.77 (0.75)	-4.06 (10609)***	4.77 (0.76)	4.75 (0.75)	.82(4581)
Family Functioning/Resilience	4.42 (1.16)	4.56 (1.23)	-5.57 (9347)***	4.63 (1.08)	4.42 (1.23)	5.79(4033)***
Social Support	4.56 (1.47)	4.53 (1.51)	0.87 (9489)	4.71 (1.39)	4.49 (1.55)	4.81(4045)***
Concrete Support	4.3 (1.56)	3.92 (1.79)	10.96 (8968)***	4.16 (1.7)	4.07 (1.67)	1.66(4059)
Nurturing/Attachment	5.25 (0.85)	5.41 (0.77)	-9.14 (9128)***	5.36 (0.76)	5.35 (0.81)	.48(3968)
Knowledge of Parenting/Child Development						
Question 1:	3.71 (1.87)	3.64 (2.04)	1.62 (8804)	3.69 (1.92)	3.74 (1.9)	-.77(3975)
Question 2:	4.69 (1.52)	4.69 (1.53)	-0.20 (9149)	4.75 (1.48)	4.7 (1.52)	1.07(3986)
Question 3:	4.28 (1.87)	4.51 (1.84)	-6.03 (9223)***	4.44 (1.85)	4.38 (1.81)	1.06(3983)
Question 4:	5.29 (1)	5.13 (1.23)	6.63 (7722)***	5.22 (1.14)	5.31 (1.06)	-2.33(3869)*
Question 5:	5.17 (1.24)	5.03 (1.43)	4.91 (8615)***	5.04 (1.4)	5.16 (1.28)	-2.93(3888)**

Note: Lower baseline scores indicate greater need. **p<.05, **p<.01, ***p<.001*

Table 3. Baseline Differences by Caregiver Stress, Income, Age, and Household Size

Family Development Baseline Scores	Parent Stress	Family Income	Parent Age	Household Size
Economic Self-Sufficiency	-.351**	.607**	.082**	.046**
Mental/Physical Health	-.169**	.093**	-.225**	0.008
Debt Management	-.204**	.170**	-.019*	-0.007
Child Care	-.162**	.152**	.055**	.039**
Child Education	-.060**	0.012	-.075**	.019*
Substance Abuse	-.082**	.020*	-.031**	0.012
Family Functioning/Resilience	-.164**	.082**	-.049**	.076**
Social Support	-.182**	.112**	-.043**	.018*
Concrete Support	-.128**	.093**	-0.01	-0.013
Nurturing/Attachment	-.342**	-.089**	-.177**	-.066**
Knowledge of Parenting/Child Development				
Question 1:	-.047**	-.045**	-.089**	-.034**
Question 2:	-.047**	-.031**	-.062**	0.01
Question 3:	-.052**	0.007	-.111**	-.058**
Question 4:	-.019*	-.030**	-.045**	-.050**
Question 5:	0.006	-.055**	-.102**	-.048**

Note: Lower baseline scores indicate greater need. **p<.05, **p<.01, ***p<.001*

Table 4. Highest Average Improvements by Service Delivery System

Family Development Outcome	Top Program Overall		
Economic Self-Sufficiency	Center	Home	
Mental/Physical Health	--		
Debt Management	Home	Center	Mix
Substance Abuse	Home		
Child Care	Home		
Child Education	Center		
Family Functioning/Resilience			
Social Support	Home*	Center*	
Concrete Support	Home	Center	
Nurturing/Attachment	Center*	Mix*	
Q1. What to do as a parent	Mix*	Home*	
Q2. Help my child learn	Home	Center	
Q3. Does not misbehave to upset me	Mix	Mix	
Q4. Praise my child when they behave	Home	Center	
Q5. Do not lose control	Center		

Note: Programs are listed based on indication of greatest advantage (e.g. 'a'(s) listed; if no 'a'(s), 'b'(s) greater than 2; if no 'b'(s) greater than 2, 'b'(s) greater than 1).=HS, =CCR, =PSSF, =FDS, =Safe CMP. *Differences only reached marginal significance.

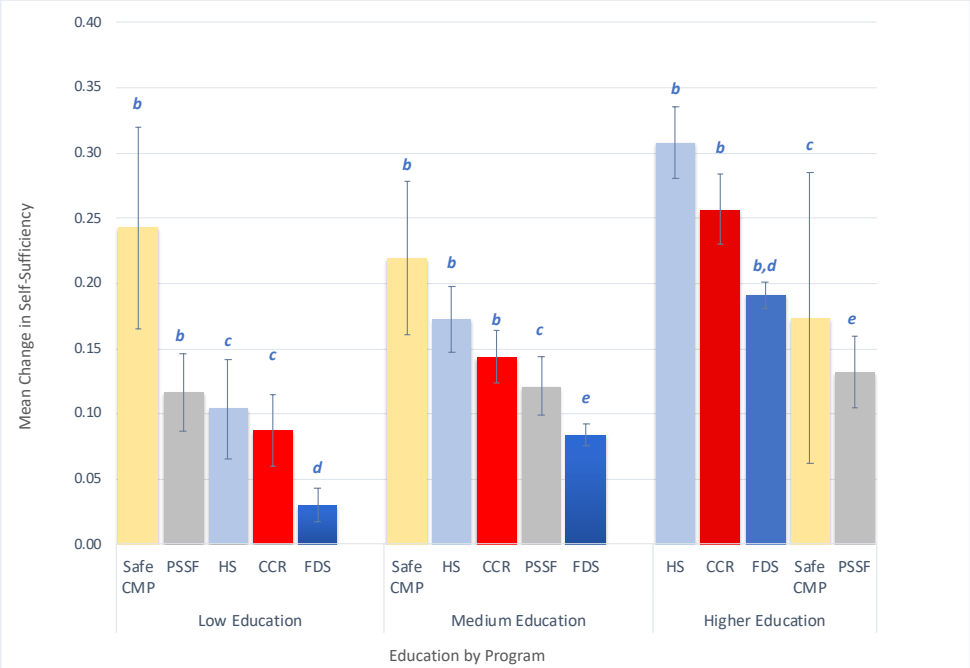
Table 5. Greatest Improvements by Service System Type and Participant Subgroup

Family Development Outcome	Low Income ($0)	Low Education (<HS)	ESL	BIPOC	Latinx	Single-parent household	High Stress (n=2)
Economic Self-Sufficiency		Hm / Mix		Ctr			
Mental/Physical Health							
Debt Management			Hm / Ctr / Mix				
Substance Abuse							Mix
Child Care	Hm						Hm
Child Education		Ctr / Mix	Hm	Hm / Ctr			Ctr
Family Functioning/Resilience	Ctr					Hm / Hm / Ctr	
Social Support		Hm					
Concrete Support				Hm		Ctr	Ctr
Nurturing/Attachment					Hm / Mix Ctr / Hm		
Q1. What to do as a parent							
Q2. Help my child learn				Ctr / Hm			
Q3. Does not misbehave to upset me	Mix						Mix
Q4. Praise my child when they behave							
Q5. Do not lose control							Hm / Ctr Mix / Hm

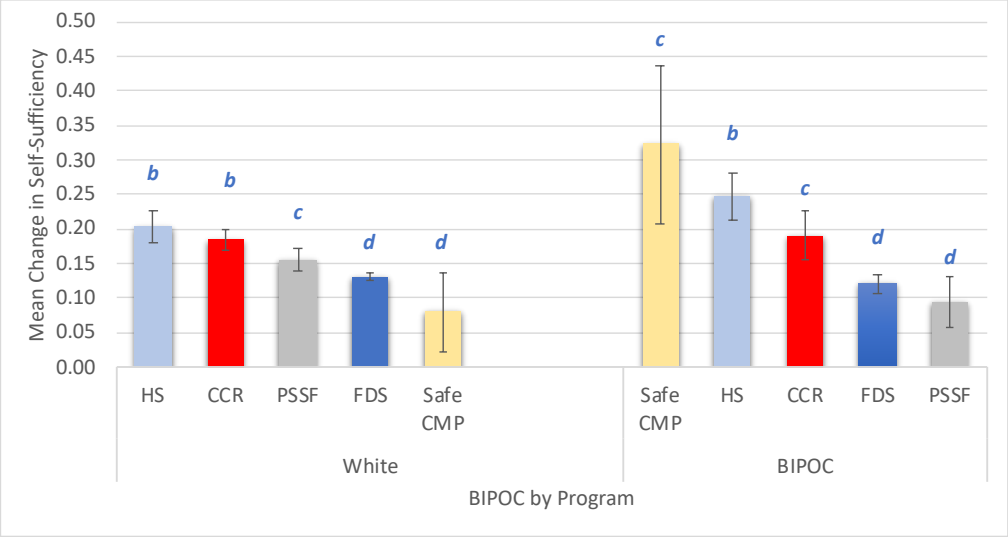
Note: Ctr=center-based, Hm=Home or community-based, Mix=Mixed delivery system. Programs are listed based on indication of greatest advantage (e.g. 'a'(s) listed; if no 'a'(s), 'b'(s) greater than 2; if no 'b'(s) greater than 2, 'b'(s) greater than 1). =HS, =CCR, =PSSF, =FDS, =Safe CMP

Appendix E. Interaction Graphs.

1. Interactions between program and identified predictors of self-sufficiency change and program

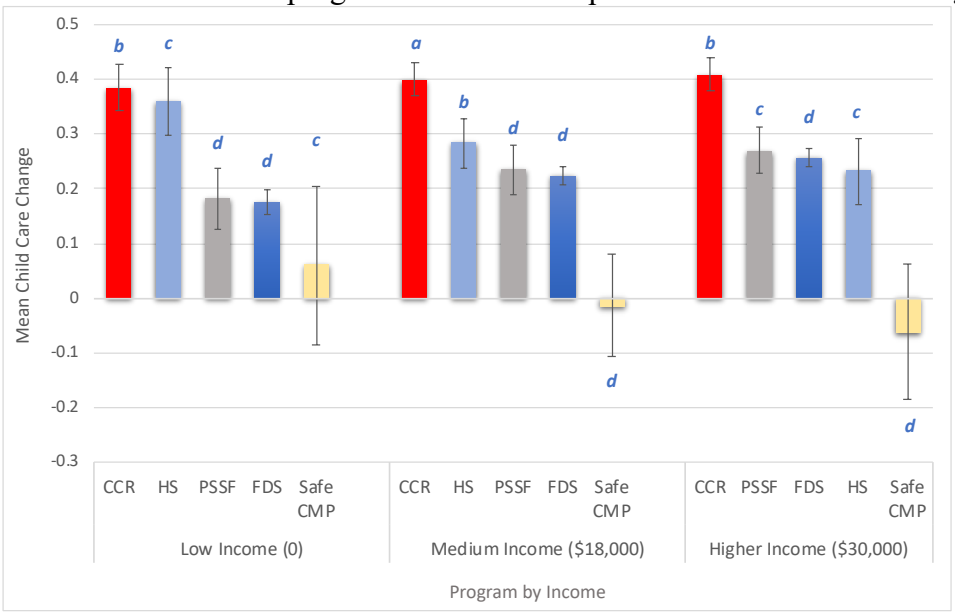


a. Program by Parent Education

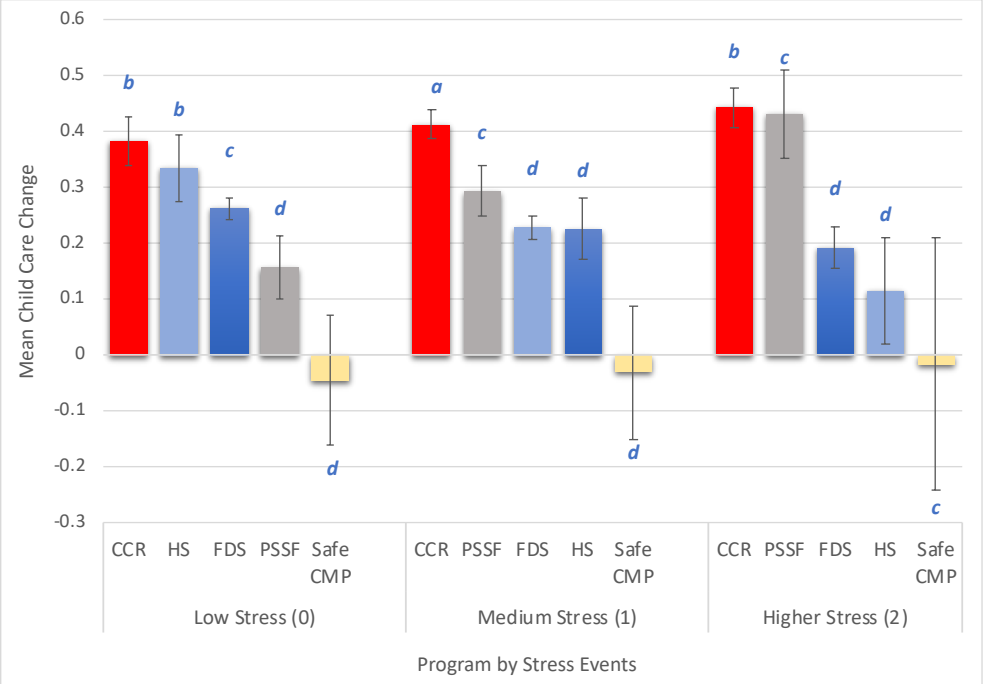


b. Program by BIPOC

2. Interactions between program and identified predictors of child care change

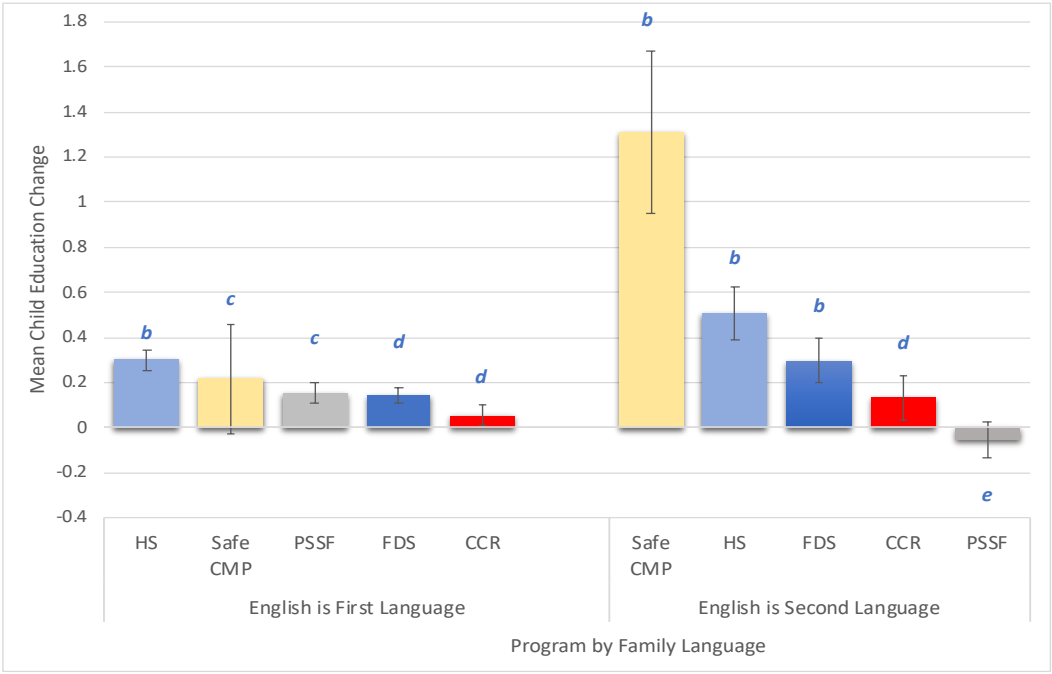


a. Program by Income

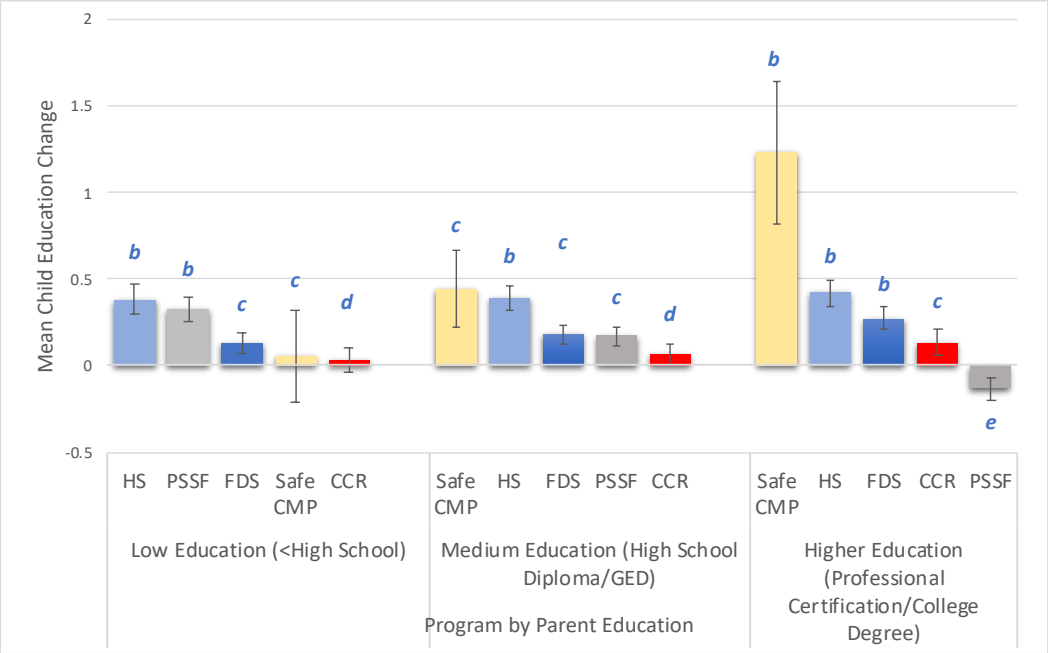


b. Program by Stress

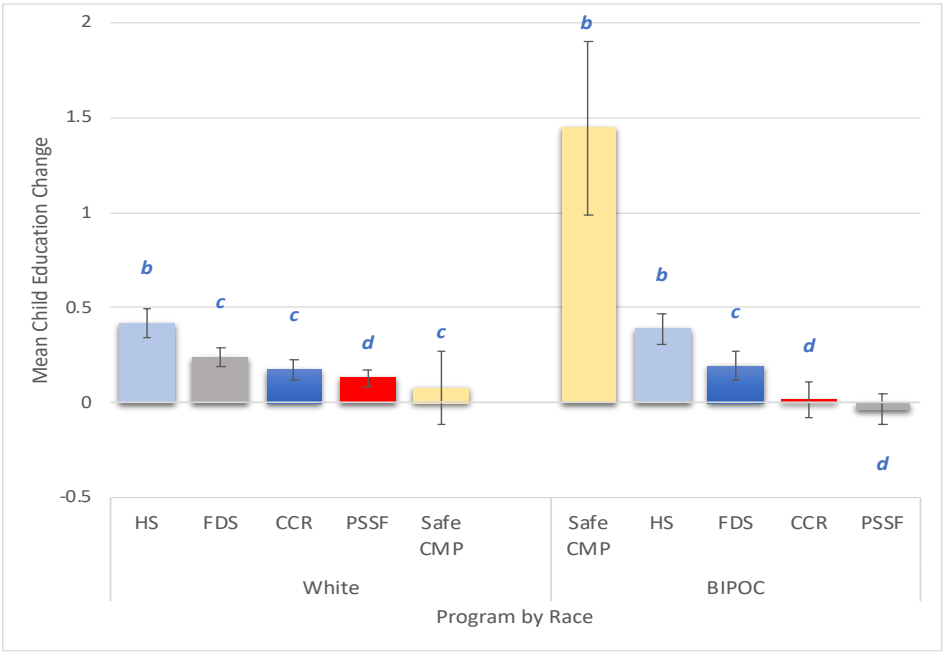
3. Interactions between program and identified predictors of child education improvements



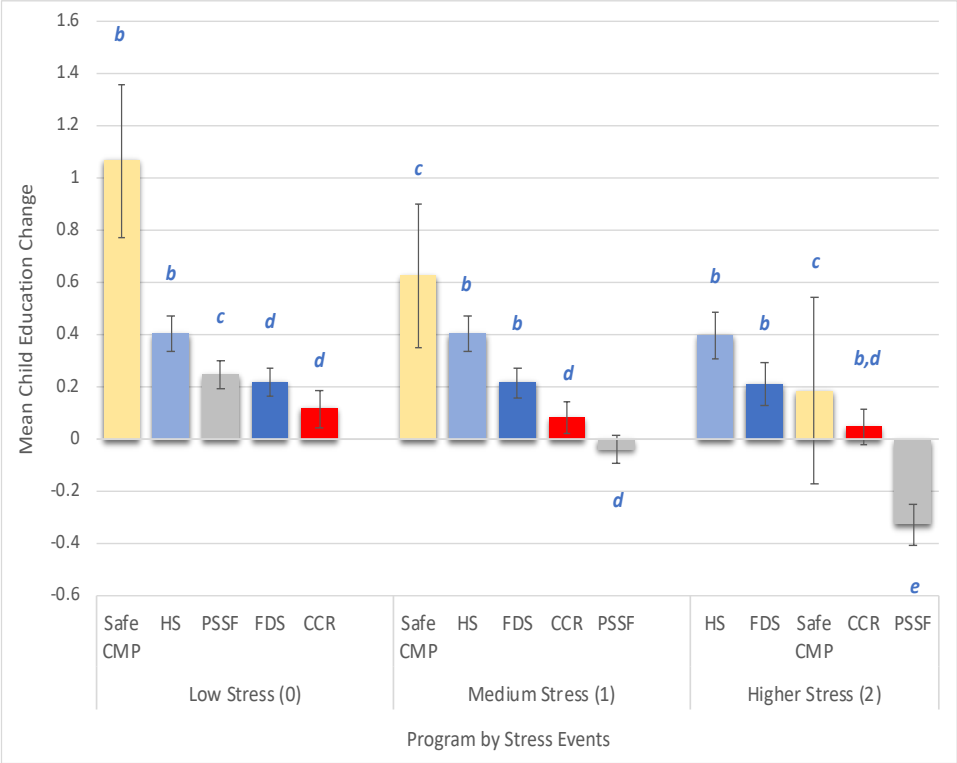
a. Program by ESL



b. Program by Parent Education

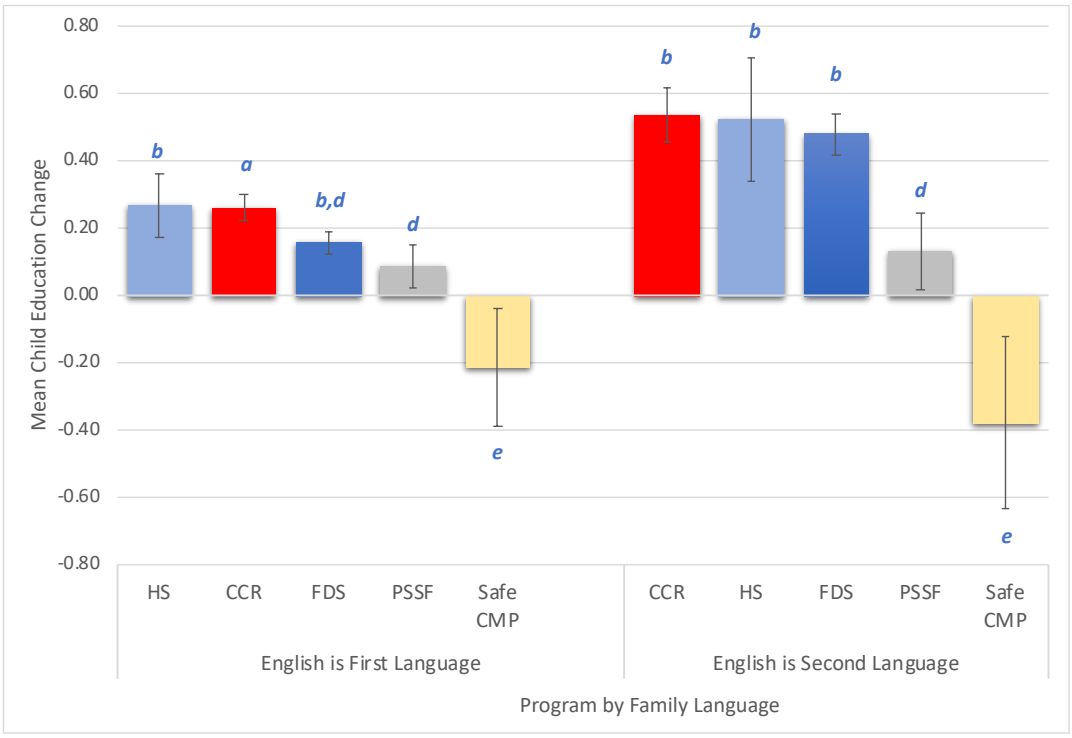


c. Program by BIPOC



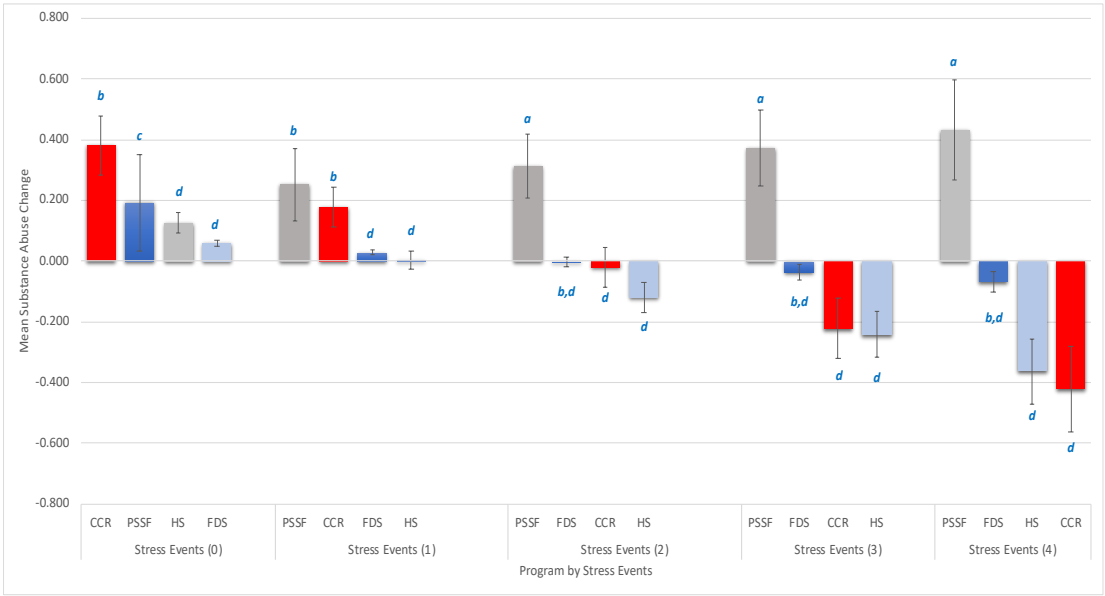
d. Program by Stress

4. Interactions between program and identified predictors of Debt Improvements



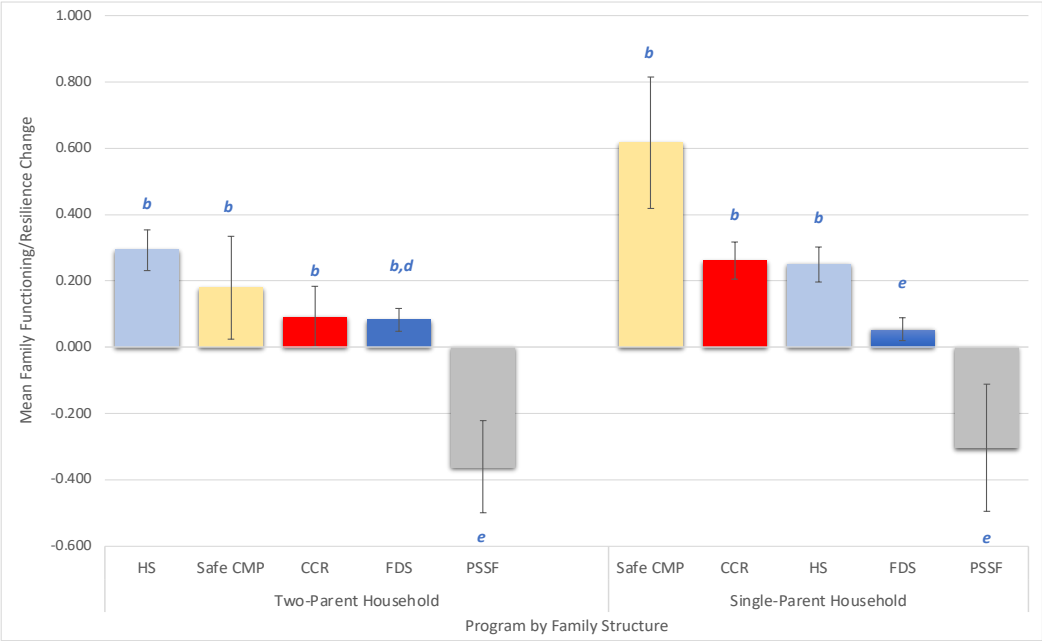
a. Program by ESL

5. Interactions between program and identified predictors of Substance Abuse Improvements



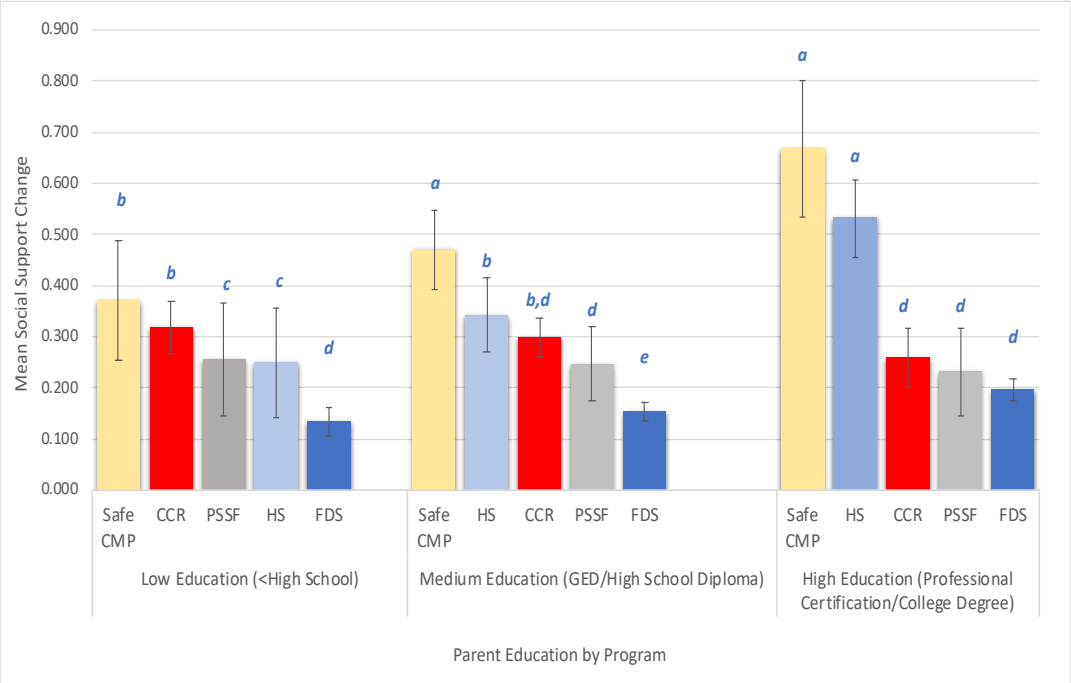
a. Program by Stress

6. Interactions between program and identified predictors of family functioning and resilience



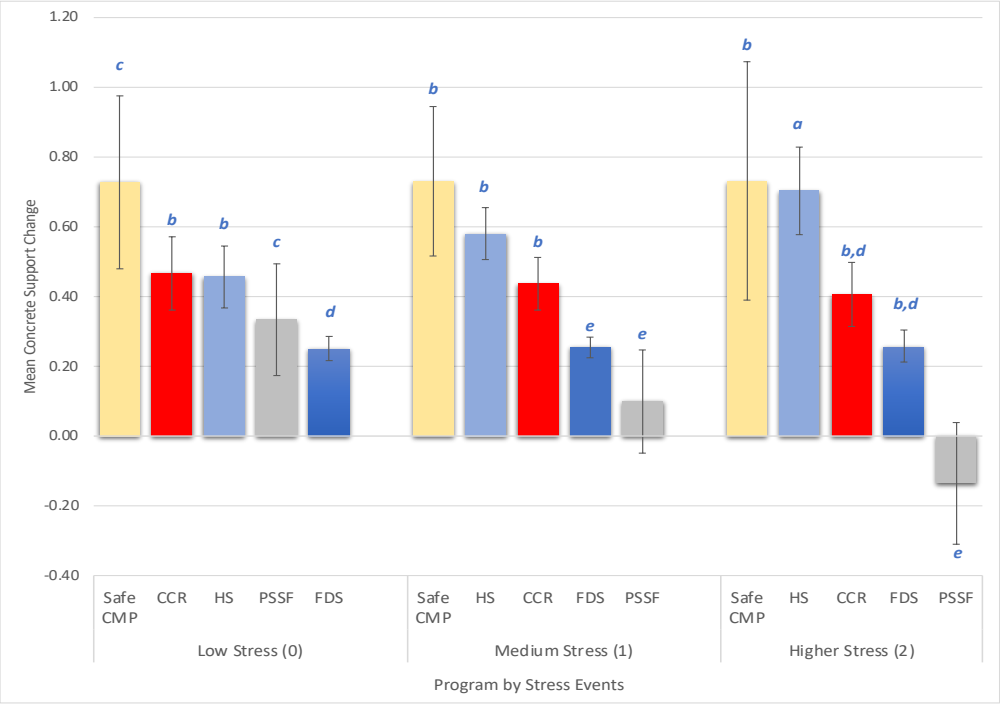
a. Program by Household Structure

7. Interactions between program and identified predictors of social support

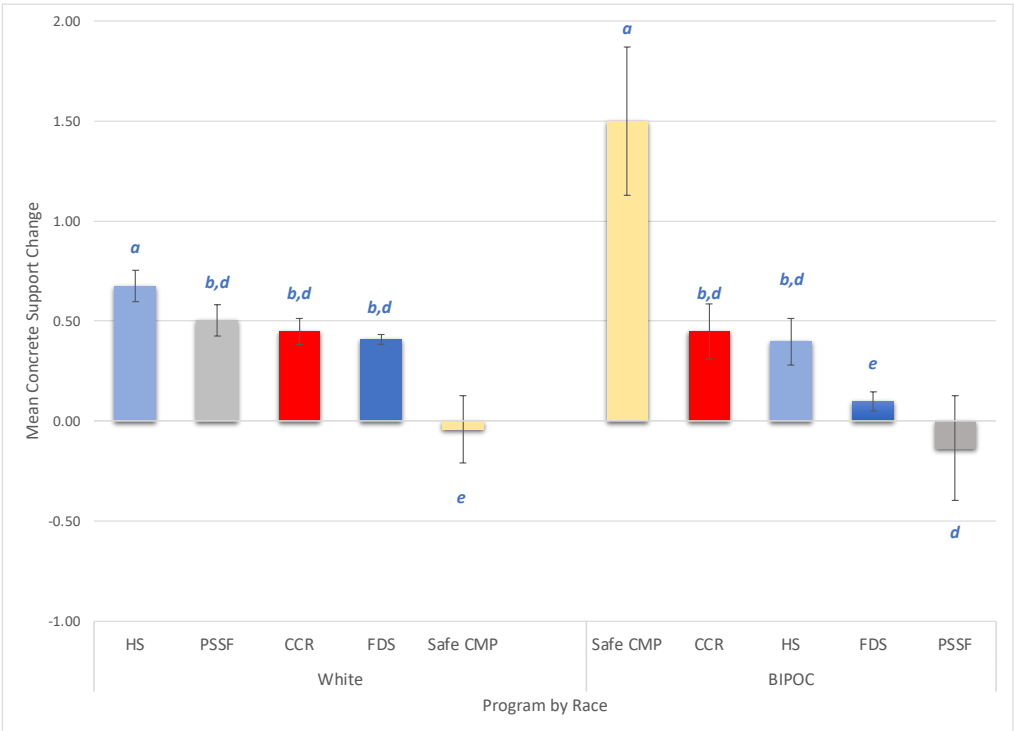


a. Program by Education

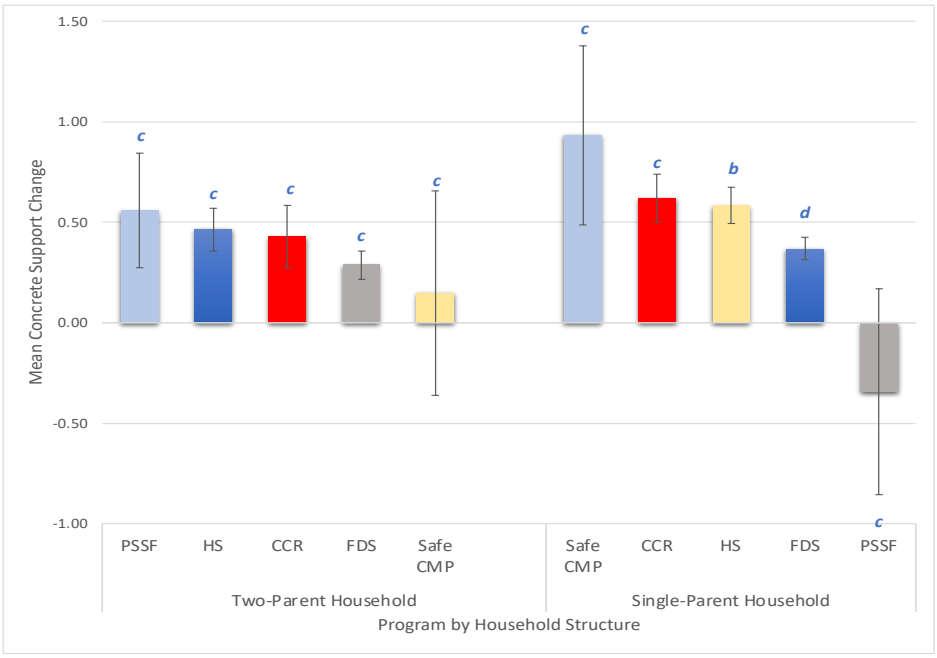
8. Interactions between program and identified predictors of Concrete Support



a. Program by Stress

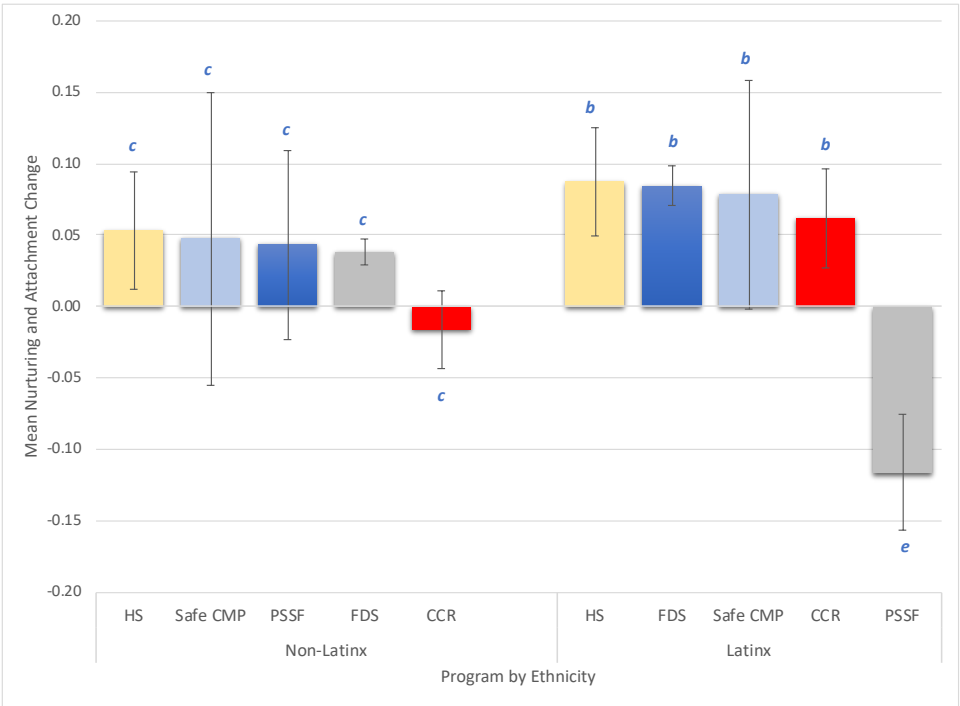


b. Program by Race



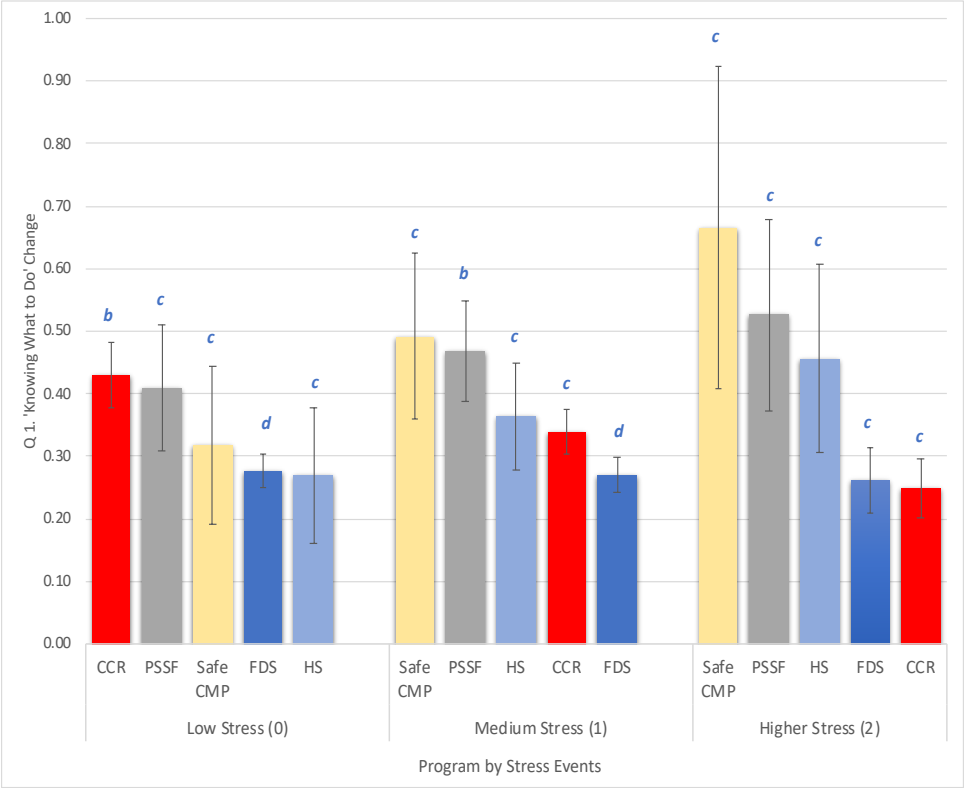
c. Program by Household Structure

9. Interactions between program and identified predictors of Nurturing and Attachment



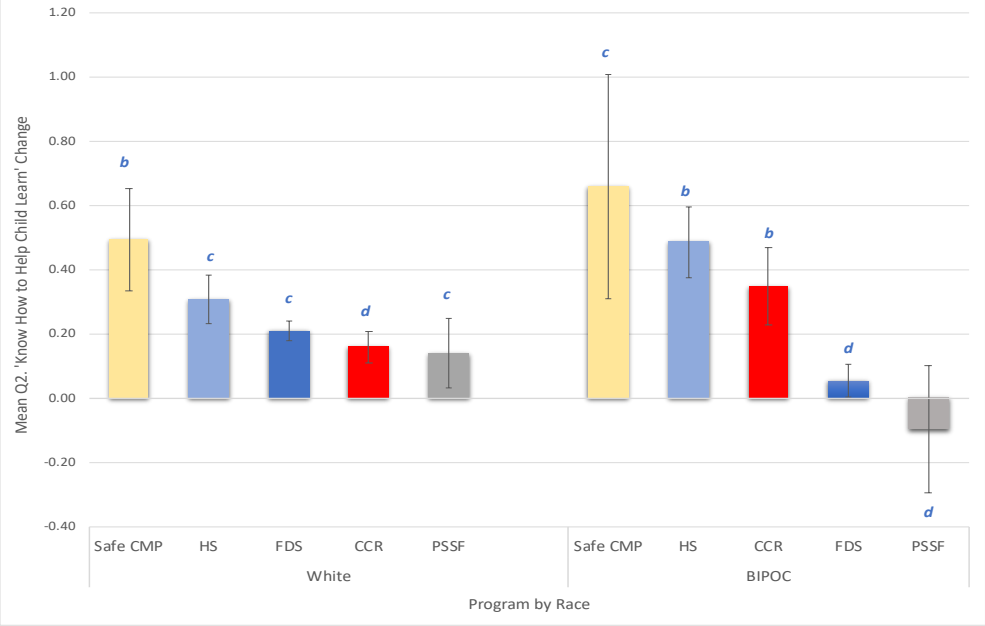
a. Program by Ethnicity

10. Interactions between program and identified predictors of Question 1: ‘I know what to do as a parent’.



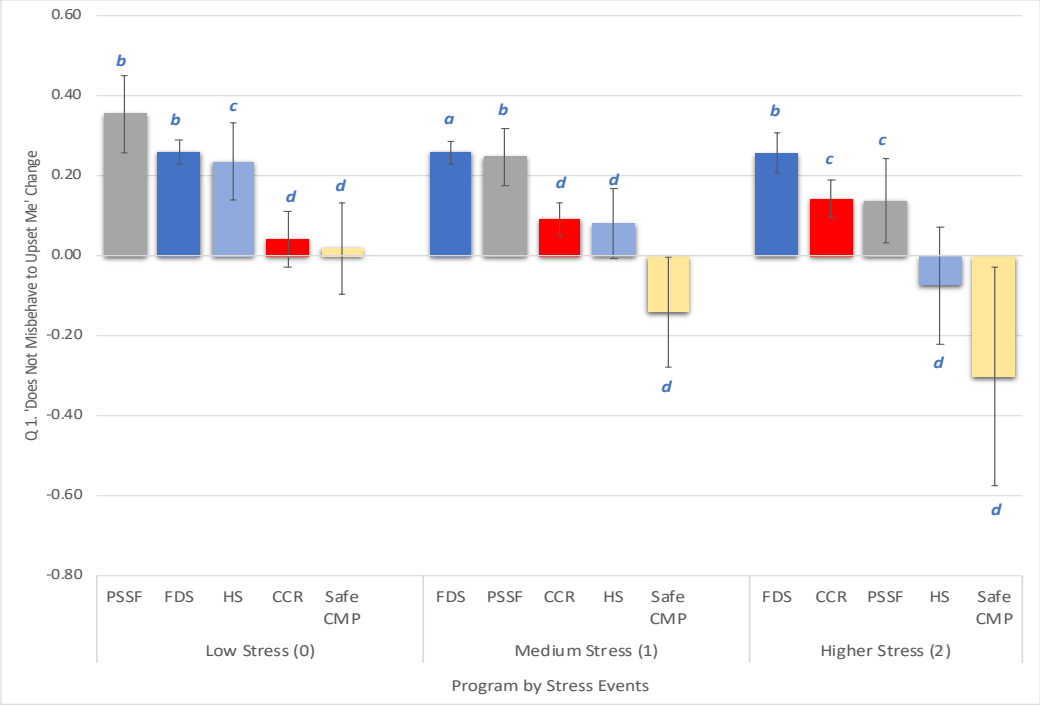
a. Program by Stress

11. Interactions between program and identified predictors of Question 2:'I know how to help my child learn'.

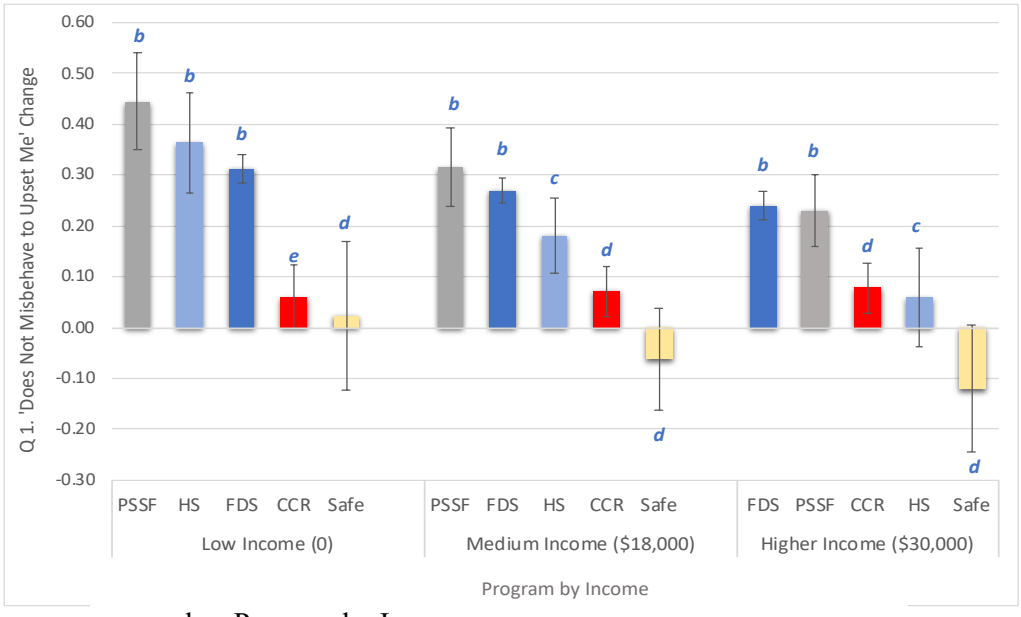


a. Program by Race

12. Interactions between program and identified predictors of Question 3: ‘My child does not misbehave to upset me’.

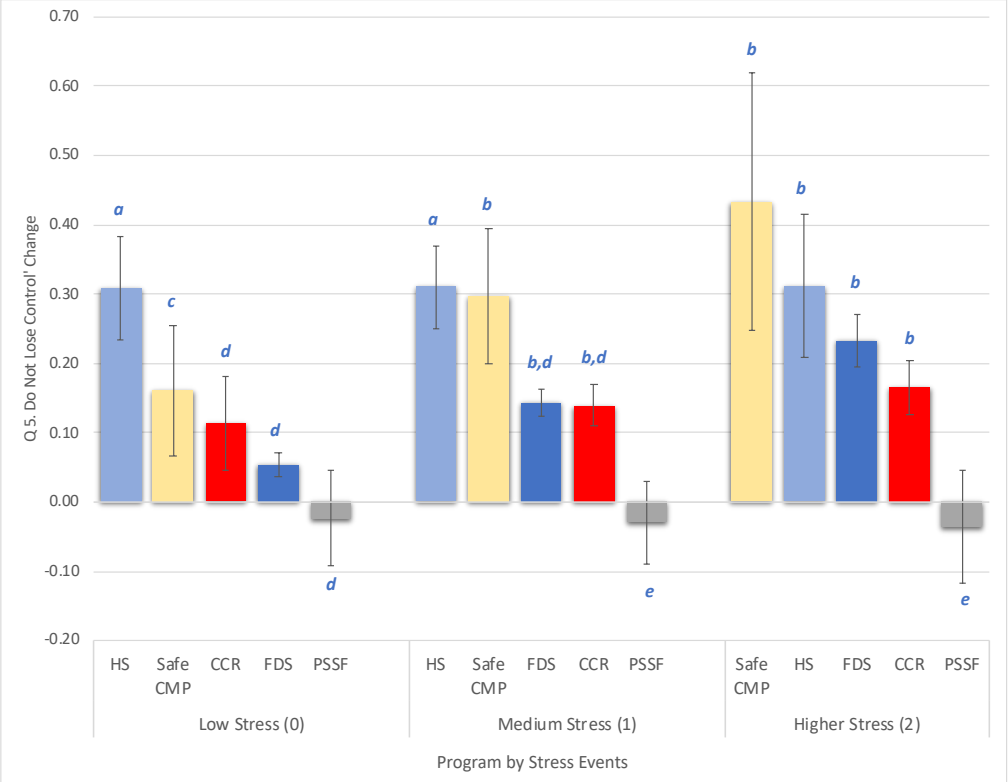


a. Program by Stress



b. Program by Income

13. Interactions between program and identified predictors of Question 5: ‘I do not lose control when I discipline my child’



a. Program by Stress