WORKING PAPER · NO. 2022-58

Investing in Early Childhood Development in Preschool and at Home

Greg Duncan, Ariel Kalil, Magne Mogstad, and Mari Rege

MAY 2022

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

INVESTING IN EARLY CHILDHOOD DEVELOPMENT IN PRESCHOOL AND AT HOME

Greg Duncan
Ariel Kalil
Magne Mogstad
Mari Rege

April 2022

This paper was prepared as a chapter for the Handbook of the Economics of Education Volume 6. We are grateful to the editors, Ludger Woessmann, Eric Hanushek and Stephen Machin, for helpful comments and suggestions.

Investing in Early Childhood Development in Preschool and at Home
Greg Duncan, Ariel Kalil, Magne Mogstad, and Mari Rege
April 2022
JEL No. J13

ABSTRACT

The foundations for successful child development are established in early childhood. Two main policy approaches for strengthening these foundations have been subsidized preschool programs and programs targeting the home environment. Our chapter reviews a large body of empirical work investigating whether these programs make a difference for children's development, and if so, how and under what conditions do they help, how cost-effective are they, and which programs are scalable. We start by reviewing studies that estimate how much of the variation in child outcomes can be explained by genetics versus environmental factors. These studies demonstrate that variation in environmental factors plays a key role in explaining individual life outcomes. This suggests that early childhood programs might play a significant role in helping children realize their potential in life. Nevertheless, our review of early childhood programs demonstrates that the evidence is mixed – some programs are successful in fostering lasting skill development, but many are not. We conclude that existing research on early childhood education falls short of sufficiently answering fundamental questions about what works for whom and why. A tighter link between theory, econometric methods and data is essential to compare and reconcile the mixed and sometimes conflicting empirical results across studies, and to understand when and why the impacts of home environment and pre-school interventions fade out.

Greg Duncan
University of California, Irvine
School of Education
2056 Education Building
Mail Code 5500
Irvine, CA 92697
gduncan@uci.edu

Magne Mogstad
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
magne.mogstad@gmail.com

Ariel Kalil
Harris School of Public Policy
University of Chicago
1307 E. 60th Street
Chicago, IL 60637
a-kalil@uchicago.edu

Mari Rege
Department of Economics and Finance
University of Stavanger
4036 Stavanger
Norway
mari.rege@uis.no

1. Introduction

Governments around the world are investing heavily in early childhood development (Engel et al. 2015). An important goal of these investments is to ensure that children from all kinds of family backgrounds are provided with warm and stimulating care environments at home and in other childcare settings. These investments are typically based on two premises: First, hundreds of studies document large gaps between disadvantaged and more advantaged children in socioemotional and academic development (Duncan & Murnane 2011). Second, the returns to skill investments in early childhood may be particularly high because this is a time when children's brain development is particularly rapid and malleable. Children's experiences early in life shape their developing brain architecture and can affect their skill and socioemotional development for life (National Scientific Council on the Developing Child 2010). These insights are incorporated into models of skill development positing processes by which "skills beget skills" (Cunha & Heckman 2007).

Despite a consensus that early childhood is a critical period, a time in which the foundation for successful child development is laid, translating that insight into policy is not an easy task. The two main policy approaches to promoting early childhood development, aside from providing health care and parental leave[1], have been subsidized preschool programs and programs targeting the home environment. Do these programs make a difference for children's development, and if so, how do they help, under which conditions, how cost-effective are they, and which programs are scalable? In this chapter we review the literature on the impacts of these programs, discuss limitations of this research and outline promising avenues for future research. We emphasize studies from the field of economics, but also review major insights from the fields of developmental psychology and early childhood education. We build on the earlier chapters in the Handbook of the Economics of Education by Blau and Currie (2006) and Cunha et al. (2006).

At the outset, a caveat: Our examination of the returns of early childhood investments is focused on skill development. Childcare and, especially, home environments can be critical for children's well-being and happiness, even if they fail to generate lasting impacts on skill development. Happy childhoods are important in and of themselves. Unfortunately, little research has focused on the returns to these kinds of outcomes.

[1] There is a substantial literature investigating how programs that allow parents to stay home from work to care for their children affect child development (Bettinger et al. 2014, Carneiro et al. 2015, Dahl et al. 2016, Danzer & Lavy 2018, Dustmann & Schönberg 2012). This literature is not reviewed in this chapter.

We begin by taking a broader look at the role that environmental factors play in affecting the lifelong developmental trajectories of individual children. We do this by reviewing studies that have sought to determine how much of the variation in child outcomes can be explained by genetics and how much by environmental factors. Recent methodological advances have improved our understanding of the importance of the environment for child development (Mogstad & Torsvik 2021). Our review in Section 2 concludes that genes account for less than half of the population variance in education, earnings, wealth and other key outcomes. This suggests that variation in environmental factors plays a key role in explaining individual observed differences in life outcomes. Moreover, and perhaps more importantly, the degree of heritability does not, in itself, tell us anything about the potential effectiveness of policy interventions.

What skills and capacities should early intervention policies target? Adopting a broad definition of skills, in Section 3 we describe the nature of skill development from birth to age 6 and identify the skills and capacities that are most important at the point of school entry. These are skills and capacities that must be fundamental for later success but, from an intervention perspective, they must also be malleable and unlikely to develop in the absence of early home and childcare interventions. Alternatively, they must be central for children's capacity to navigate the transition to formal schooling.

A key purpose of public preschool is to provide a safe and stimulating care environment for children. In Section 4 we review studies investigating how preschool enrollment affects a child's skill development. Preschool programs can be either targeted or universal. Enrollment in targeted programs, reviewed in section 4.1, is almost always restricted to children from disadvantaged families. Universal programs, reviewed in section 4.2, are open for all, but children from disadvantaged families are typically prioritized or more heavily subsidized relative to advantaged families. Targeted programs are common in the US, whereas most programs in Europe are universal. Our review shows mixed results for many targeted programs. Several studies of targeted programs show that initial learning gains quickly fade out. In the case of older preschoolers from disadvantaged families, however, preschool participation in targeted program can be beneficial for child development and produce lasting effects, especially if the counterfactual is informal care. For universal programs, the evidence is more consistent. For older preschoolers from disadvantaged families, this literature documents positive effects of

preschool participation that persist into adulthood. Some studies also suggest positive effects for younger preschoolers, but the evidence is mixed.

The literature on the impacts of preschool enrollment offers few insights into *why* the effects of preschool differ across studies because they provide few details on the features of the programs, the alternative care arrangements for children who are unable to enroll in them and the context in which the preschool services are offered. In Section 5 we review studies that have looked at how child outcomes are affected by various features of preschool programs – teacher education, classroom size, quality of teacher-student interactions, preschool curriculum, and the like. Not surprisingly, these studies are plagued by the challenge of manipulating one factor in the production technology while holding all others constant. When one factor is changed, it is common for other investments to change as well; for example, an improved preschool curriculum may lead parents to invest less time and money in their children at home, believing that the children are learning enough in preschool. Accordingly, results from the studies discussed in this section should be understood as policy effects that include the adjustment effects of other investments rather than as estimates of production technologies. Our review suggests that investments in quality features, such as teacher education or class size, have a modest or no effect on children's development, while investing in skill-specific curricula is a promising approach that warrants more research. Specifically, some studies suggest fadeout of preschool curricula investment, and we need more research to better understand what skills curricula should target to have persisting effects.

In early childhood, parents play important roles in cultivating children's skills and interests (Vygotsky 1967). They invest in their children by creating a nurturing and stimulating home environment. This involves structural factors such as books and toys, routines to promote sleep, physical activity and healthy eating, and the safety of the home environment, as well as process factors, such as the amount of time the parents spend interacting with the child and the quality of these interactions, i.e. parental behaviors. There is a large gap between advantaged and disadvantaged families in how parents invest in their children's home environment (Duncan et al. 2010, Waldfogel & Washbrook 2011). In Section 6 we review studies of public policies that seek to close these gaps by supporting disadvantaged families with young children through income support and interventions targeting parent behaviors. We conclude that programs targeting family income appear to benefit children in very poor households. Moreover, precise and tailored

approaches to changing parents' behavior in a way that does not place excessive demands on parents' time and attention seem promising, but more research is needed.

In Section 7, finally, we conclude that although much progress has been made, existing research on early childhood education fails to answer fundamental questions about what works for whom and why. A tighter link between theory, econometric methods and data is essential to compare and reconcile the studies' mixed and sometimes conflicting empirical results, and to understand when and why the positive impacts of home-environment and pre-school interventions fade out. Such an understanding is needed not only to explain how current policies affect children in the short and longer run, but also to prescribe new policies that can help children realize their potential in life.

2. The importance of environmental factors and genetics

It is well documented that economic prosperity tends to persist across generations. Children born to parents with high levels of education or income can expect to do better than children born into less favorable conditions. These differences can be attributed in part to the fact that children are born with different cognitive and non-cognitive capacities to acquire the knowledge, skills and attitudes—the human capital—that the labor market values. The genes that are passed on from parent to child may limit the potential of an individual to achieve in the labor market, while the environment that person grows up in influences the extent to which that potential is reached.

In this section, we review studies that seek to determine how much of the variation in child outcomes can be explained by genetics and how much by environmental factors.[2] We begin by discussing how sibling correlations in outcomes have been used to obtain an omnibus measure of the role that family background plays. These correlations reflect not only the impact of shared genes, but also any shared family environment. The correlation in outcomes between siblings can be expected to be low if family background plays only a minor role in individuals' life outcomes.

Next, we discuss two strands of the literature that try to separate genes from family environment. One uses methods from behavioral genetics to isolate the role of heritability. The other uses quasi-random assignment of adopted children to estimate the causal impact of growing

[2] This section draws on the review article by Mogstad and Torsvik (2021).

up in one family environment rather than another. We describe the pros and cons of these two approaches, and then discuss some influential findings and their policy implications. An important conclusion is that less than half of the population variance in outcomes such as education, earnings and wealth is explained by genes. This finding suggests that environmental factors play a key role in explaining the observed differences in individual life outcomes.

As discussed in greater detail later, the policy implications of the heritability estimates are the subject of much controversy. Part of this controversy may be due to confusion about what these estimates do and do not capture. A heritability estimate measures the fraction of the population variation in individual outcomes that can be explained by genetic variation in the population. However, the degree of heritability of an outcome cannot tell us how important genes are for shaping the outcome, nor how easy it is to change that outcome. Furthermore, estimates of the role of environmental factors do not tell us which features of the environment are particularly important. Instead, these estimates show how a specific combination of environmental factors – including home environment and preschool programs – lead to differences in child outcomes.

2.1 Sibling correlations

Sibling correlations are frequently used to construct an omnibus measure of how family background affects children's income or education. It is useful to express the earnings (or any other life outcome) for individual i who was raised in family j as $Y_{ij} = a_j + b_{ij}$, where a captures the environmental component shared by all siblings and b is the sibling-specific component. Since these components are constructed to be independent, the share of the variance in earnings that is explained by the shared environmental component is given by $\rho_{\text{sib}}^{Y} = \frac{\sigma^a}{\sigma^a+\sigma^b}$.

Sibling correlations necessarily explain more of the variation in offspring earnings (or in any other life outcome) than do parents' earnings alone. This is simply because one of the shared environmental components in a is parents' earnings, and a also contains other family and neighborhood factors that are relevant for offspring's earnings, but independent of parents' earnings. Solon (1999) shows that the sibling correlations and intergenerational correlations (IGC) can be linked in the following way:

$$\rho_{\text{sib}}^{Y} = (\text{IGC})^2 + \text{ all sibling shared factors not correlated with parent Y.} \quad (1)$$

where IGC is defined as the correlation in the outcome of interest (e.g. earnings) between the child and her parents. [3] This expression is useful for two reasons: It helps interpret and compare sibling correlations and intergenerational correlations, and it allows for the construction of bounds on quantities of interest, as discussed in Björklund and Jäntti (2020).[4]

One can derive a lower bound on environmental influence from the correlation in outcomes between siblings. It is a lower bound because a given environment may affect children differently. For example, siblings may receive different genetic endowments from their parents, the birth order may matter for outcomes, and there could be temporal changes in the environment that will create differences between siblings. Sibling correlations will erroneously assign all these factors---non-shared genes and family factors that affect siblings differently---to the individual component b, not to the family component a.

One can construct an upper bound on family influence from the correlation in outcomes among monozygotic twins. The argument is that monozygotic twins receive the same genetic endowment from their parents and, since they are born at the same time, they also share the entire family environment. However, to the extent that monozygotic twins are treated more equally by the environment than ordinary siblings, and influence each other more than ordinary siblings do, their resemblance in outcomes might overstate the importance of environmental factors for the population at large.

In Table 1, we report sibling and twin correlations from a set of empirical studies. The sibling correlations are much higher for monozygotic twins than for ordinary siblings. It is also interesting to observe that sibling correlations suggest that environmental factors are more important for educational attainment and earnings (or income) in the U.S. than in the Nordic countries. By comparing the IGC of an outcome Y, for example educational attainment or earnings, with sibling correlations in Y one can use the expression in equation (1) to calculate how much of the variation in children's outcomes that is explained by the variation in parents' outcomes. Björklund and Jäntti (2020) make this comparison and conclude that the IGC in education and income explains relatively little, roughly ten percent, of the sibling correlations in

[3] Formally, the ICG is the Pearson correlation in outcomes across the generations. The ICG is closely related to the much used intergenerational elasticity. Specifically, the ICT can be transformed to the intergenerational elasticity by multiplying by the ratio of the standard deviation in child outcome relative to parent outcome.

[4] In an early study of sibling correlations, (Solon et al. 1991) show that transitory shocks to earnings will attenuate the degree of sibling correlation in permanent income, just as for the estimation of intergenerational correlation.

these outcomes. This suggests that factors other than parental income and education are likely to be important for the observed differences across children in income and education.

Table 1. Sibling and twin correlations from a set of empirical studies

			Ordinary Siblings			Twins	
			Brothers	Sisters	Mixed	MZ-male	MZ-female
Country	Study	Outcome					
Denmark	(Schnitzlein 2014)	long run earnings	.20	.19			
Germany	(Schnitzlein 2014)	long run earnings	.43	.39			
USA	(Schnitzlein 2014)	long run earnings	.45	.29			
USA	(Mazumder 2008)	long run earnings	.49				
Norway	(Pekkarinen et al. 2017)	long run earnings	.32				
Sweden	(Björklund et al. 2009)	long run earnings	.35				
Norway	(Björklund & Jäntti 2009)	years of schooling			.41		
Denmark	(Bredtmann & Smith 2018)	years of schooling	.31	.39	.33		
Sweden	(Björklund & Jäntti 2012)	years of schooling	.43	.40			
USA	(Mazumder 2008)	years of schooling			.60		
Finland	(Hyytinen et al. 2019)	long run earnings				.54	.41
Sweden	(Benjamin et al. 2012)	long run income				.62	.48
Sweden	(Björklund & Jäntti 2012)	long run earnings				.73	
Sweden	(Björklund & Jäntti 2012)	years of schooling				.75	.73
Australia	(Miller et al. 2001)	years of schooling				.65	.70
USA	(Behrman & Taubman 1989)	years of schooling				.75	

Note: This table is based on Table 1 and Table 2 in (Björklund & Jäntti 2020).

2.2 Heritability and the ACE model

The canonical model for inferring the importance of genes relative to environment is the so-called ACE model. The basic version of the ACE model assumes that the phenotype, educational attainment or earnings of individual i in family j (Y_{ij}) can be represented by an additive function of genes (A_i), shared family environment (C_j) and idiosyncratic influences (E_i):

$$Y_{ij} = hA_i + bC_j + dE_i. \quad (2)$$

If the genetic component is independent of environment, the degree of heritability, which is defined as the fraction of the overall variance in the phenotype that can be attributed to the genetic component, is given by $h^2 = \frac{VAR(A_i)}{VAR(Y_{ij})}$. By comparison, the contribution of family environment is given by $c^2 = \frac{VAR(C_i)}{VAR(Y_{ij})}$.

Heritability can be estimated by comparing the correlation in outcomes for sibling pairs who differ in the degree to which they are genetically related. If we normalize Y with a standard deviation of 1, the correlation in outcomes for a sibling pair of type k is given by

$$\rho^k_{YY'} = \rho^k_{AA'}h^2 + \rho^k_{CC'}c^2.$$

The difference in correlation between monozygotic (m) and dizygotic (d) twins is then given by

$$\rho^m_{YY'} - \rho^d_{YY'} = \left(\rho^m_{AA'} - \rho^d_{AA'}\right)h^2 + \left(\rho^m_{CC'} - \rho^d_{CC'}\right)c^2.$$

If monozygotic twins share 100% of their genes, while dizygotic twins share 50% of their genes, but both types of twins have the same degree of shared environment, we get

$$h^2 = 2\left(\rho^m_{YY'} - \rho^d_{YY'}\right)$$

Comparing ordinary siblings and adopted siblings yields similar expressions (see e.g. the discussion in Sacerdote (2007)).

Findings from the ACE model. According to a recent meta-study by Polderman et al. (2015), over the last fifty years 2,748 publications have used nearly 15 million pairs of twins to estimate the heritability of 17,804 human traits. The average heritability for all traits tested is around 0.5, and physical traits such as height are more heritable (around 0.8) than more complex behavioral outcomes.

Taubman (1976) is an early study that uses twins to infer the genetic component of earnings. Since this study was published, this framework has been used extensively by social scientists to estimate the heritability of many socio-economic outcomes, most often educational attainment, earnings, and income, but more recently also wealth and other aspects of the economic phenotype, for example risk preferences and entrepreneurship (Nicolaou et al. 2008).

Using long panels of earnings for monozygotic and dizygotic twins, Hyytinen et al. (2019) estimate that the heritability of lifetime earnings is 53% for men and 39% for women (see Table 1). Their findings are broadly in line with other studies of the heritability of earnings and income, such as Sacerdote (2011) and Björklund and Jäntti (2020). In a meta study, Branigan et al. (2013) find that, on average, around 40% of the variance in educational attainment can be attributed to variation in genetic components. The heritability of educational attainment will naturally vary across countries, depending on environmental factors such as access to and quality of educational institutions.

Engzell and Tropf (2019) combine cross country data on intergenerational mobility in education and data from twin studies of the heritability of education. They find a positive

association between heritability and intergenerational schooling mobility. A possible explanation is that in a society with equal opportunities and universal access to higher education, mobility will be relatively high. Since everyone in such a society has equal opportunities to choose higher education, ability will explain much of the variation in education and, as a result, heritability will be high (Trzaskowski et al. 2014).

Limitations and critiques of the ACE model. The basic ACE model rests on several restrictive assumptions. One is that ordinary siblings share, on average, 50% of their genes. However, that number is probably too low because individuals tend to mate and have children with persons who resemble themselves (Kalmijn 1998). With assortative mating, dizygotic twins or siblings more generally can be expected to share more than 50% of their genes and, as a result, the ACE model underestimates the heritability of traits.

Another restrictive assumption of the ACE model is that it assumes independence between genes and the environment and no interaction effects between them. With respect to the environmental factors and genes that produce the economic phenotype, it seems likely that good genes are correlated with a favorable environment, partly because the genes "choose" and shape their environment (Plomin et al. 1977). There is also an increasing body of evidence, discussed below, suggesting that the impact of genes -- that is, the way they express themselves in a phenotype -- depends on the environment.[5]

The ACE model further assumes that all sibling types, irrespective of the degree to which they are genetically related, share family environment to the same degree. This is an untenable assumption. It is likely that monozygotic twins are treated more equally by parents and peers than dizygotic twins, and, as a consequence, monozygotic twins may share more environmental factors than do dizygotic twins. Some of the additional correlation between monozygotic twins might therefore be attributed to the fact that their environment is more similar.

When assessing the restrictions of the basic ACE model, it is useful to observe that additional data may make it possible to relax some (but not all) of its strong assumptions. This is the case if the analyst has access to data on a number of different types of sibling pairs, such as monozygotic twins, dizygotic twins, regular siblings, half siblings and adoptees. Such data have

[5] Epigenetics is the study of how the environment impacts how genes express themselves in phenotypes. This could make the strict dichotomy between nurture versus nature misleading. It is therefore arguably more appropriate to talk about nature via nurture; for an interesting layperson's introduction to these ideas, see Ridley and Pierpoint (2003).

been used to allow for gene environment correlations and differences in the degree to which siblings share the same environment (see e.g. Bjorklund et al. 2005, Fagereng et al. 2021). Even in these cases, however, caution is warranted when interpreting results from the ACE model.

On the policy (ir)relevance of heritability estimates. In the social sciences, the heritability of human traits and achievements has long been a highly debated and controversial topic. Part of the controversy stems from the fact that the concept is often misunderstood. A heritability estimate measures the fraction of the population variation in individual outcomes, such as education or income, that is explained by genetic variation in the population. The degree of heritability of an outcome cannot tell us how important genes are for shaping the outcome, nor how easy it is to change the outcome.

A natural and relevant question is whether, and in what situations, heritability estimates can be useful for economic policy. Manski (2011) discusses several objections to the assumption that heritability estimates from the ACE model are relevant for policy. One objection is that the nature of policy interventions is to change the environment, while heritability is calculated from data collected in the environment that prevailed before the intervention. For example, the heritability of educational attainment will depend on the amount of heterogeneity in the quality of primary schools in the population. If we changed the environment and made schools more unequal, this would reduce the heritability of educational attainment.

While Manski (2011) raises important concerns, these problems are not specific to empirical research on heritability. It is a general concern in empirical analysis that a parameter estimated on data from one population or in one environment may not generalize to other populations or to other environments.[6] To gauge the degree of invariance in the heritability of educational attainment, it is possible to empirically examine or model how the parameter would differ in an alternative or counterfactual environment.

A second objection is that a high degree of heritability does not, in itself, tell us anything about the effectiveness of policy interventions or how costly it is to alter outcomes. As Goldberger (1979) pointed out, bad eyesight is highly heritable, but can readily be fixed by good opticians. Again, this critique is not specific to heritability estimates. The observation that the root cause of a problem may not be relevant for determining the best solution applies to a wide

[6] See, for example, Heckman (2005) for a broad and insightful discussion of structural models, treatment effects, and invariance assumptions in econometrics.

variety of situations. For example, the best way to avoid getting wet if it rains may be to stay indoors or bring an umbrella, not to change the rainy weather. And the most effective way to reduce labor market inequalities could be to change the tax-transfer system, even if one thinks that globalization and technological changes are the root causes of increased labor market inequalities.

Finally, it is important to note that even if most of the variation in an outcome can be explained by nature, this does not imply that this is the way it should be. The fact that a given outcome is highly heritable tells us nothing about whether we ought to reduce differences in that outcome across individuals. To what extent a society should strive for more or less equal outcomes in a particular dimension is a separate, normative question -- and one that science has little, if anything, to say about.

2.3 The impact of environment

An alternative approach to the ACE model for gauging the impact of the environment is to vary the environment while holding constant children's genetic relatedness to parents. The ideal experiment would be to randomly assign newborn children to parents who have different levels of education, income and wealth and who live in different neighborhoods. To see what is and is not possible to identify with such an experiment (this is discussed in greater detail in Fagereng et al. 2021, Holmlund et al. 2011), consider the extended intergenerational transmission equation:

$$Y_i = \alpha + \beta Y_{j(i)} + X'_{j(i)}\eta + \gamma\kappa_{j(i)} + X'_i\lambda + \delta\chi_i + \varepsilon_i, \quad (3)$$

where Y_i is the outcome of interest of the child i, say earnings; the characteristics of her family j consist of parental log earnings $Y_{j(i)}$ and a vector of observable (pre-determined) family characteristics other than earnings $X'_{j(i)}$ and an unobservable component $\kappa_{j(i)}$. Similarly, the child has (pre-determined) observable characteristics X'_i (e.g. gender and birth cohort) and unobservable characteristics χ_i such as genes. The scalar error term ε_i is by definition orthogonal to all other variables in the equation. The unobservable variables that may correlate with the explanatory variable of interest, parental income, are κ_{ji} and χ_i.

With random assignment of children to families, the potential outcomes, defined by the genes of a child, are uncorrelated with the family environment the child grew up in; that is, χ_i is independent of the family components $Y_{j(i)}, X'_{j(i)}, \kappa_{j(i)}$. Random assignment of children, therefore, makes it possible to estimate the causal effect of being raised in a high-earning family versus a

low-earning family. Without making further assumptions, however, one cannot use random assignment of children to estimate the ceteris paribus effect that an exogenous increase in the family's earnings would have on the child's outcome. There is likely to be some correlation between $Y_{j(i)}$ and $\kappa_{j(i)}$: higher-earning families may have other unobservable qualities in their family environment that also affect the child's outcome. Drawing causal inference about the ceteris paribus effect of an exogenous increase in parental earnings requires random variation in the earnings of a given family, not random assignment of children to a high-earning family or a low-earning family.

While most children grow up with their biological parents, some do not, and comparing outcomes of adopted children raised in different families is a frequently used empirical strategy to gauge how family environments affect children's outcomes. However, to be able to conclude that family environments are the cause of differences in outcomes among adopted children, the children must be randomly assigned to families. Kinship adoption is relatively common in many countries, and is clearly at odds with the notion of random assignment of adoptees to families. Even in the case of non-relative adoptions, there may well be a genetic association between child and parents, based on requests from adopting parents or matching criteria used by adoption agencies.

Because of these concerns, Sacerdote (2007) and Fagereng et al. (2021) use data from infants adopted from Korea to the United States and Norway, respectively. Both confirm at the assignment of these adoptees to pre-approved adoptive families is quasi-random by providing detailed descriptions of the placement rules and by ensuring that observable features of the adopting family do not predict pre-adoption characteristics of the adoptees. Sacerdote (2007) studies several outcomes, including income and education. He finds no effect of being assigned to a higher-earning family; however, adoptees who were assigned to small families in which the mother was highly educated tend themselves to attain higher levels of education. He also finds strong family environmental effects on children’s smoking and drinking habits in adulthood.

Fagereng et al. (2021) estimate the correlation between parents and their randomly assigned children with regard to wealth and financial risk-taking. They find that children who were assigned to wealthier families are significantly wealthier in adulthood. On average, adoptees accrue an extra US$2,250 of wealth if assigned to an adoptive family with US$10,000 additional wealth. This suggests that adoptees raised by parents with a wealth level that is 10% above the

mean in the parent generation can expect to achieve a wealth level that is almost 3.7% above average for their own generation. They also find that adoptees' stock market participation and portfolio risk increase with the financial risk position of their adoptive parents. To assess the importance of family environment for wealth transmission, they compare the intergenerational correlation in wealth for adoptees with non-adopted children. They find that the influence of parental wealth on children's wealth is twice as large for biological children as for adoptees.

Several other studies use the outcomes of adoptees to examine the relative importance of nature versus nurture for children's outcomes; see Holmlund et al. (2011) for an overview of the literature. None of these studies, however, can argue that adoptees are randomly assigned to parents. In fact, the authors generally acknowledge that adoptions are selective, and with selective placement it is difficult to separate the influence of genes from the influence of the family environment. One indirect solution is to find a proxy to control for the genetic disposition of the adopted child. This is the approach taken by Björklund et al. (2006). Using data from Swedish adoptees, they are able to observe the income and education (at least partially) of both the adoptive and the biological parents.

In their study, Björklund et al. (2006) regress the education of the adopted child on the education of all four of parents. They find the following transmission coefficients for years of schooling: 0.13 for biological mother, 0.11 for biological father, 0.11 for adoptive father, 0.07 for adoptive mother. Interestingly, the sum of the coefficients for the biological and the adoptive mothers resemble the coefficient of the standard educational intergenerational elasticity iIGEn the Swedish data. The same is true for the biological and adoptive fathers. This suggests that the influence of pre-birth (nature) and post-birth factors (family environment) on the outcomes of children in a variety of family types may be additive, a possibility that is explored further in Bjorklund et al. (2005). Black et al. (2020) use the same model to estimate family environment effects on wealth transmission by looking at Swedish adoptions. They also find substantial effects of the family environment on wealth transmission. The critical assumption in this approach is that selection bias in non-random adoptions can be adjusted for by controlling for the observed outcomes of biological parents. A natural concern is that the observed outcomes (e.g. wealth) of parents who give up their child for adoption may, in part, reflect adverse shocks, and as a result they would be poor control variables for the biological parents' potential outcomes (e.g. wealth).

A research design with quasi-random assignment of adoptees to families, as in Sacerdote (2007) and Fagereng et al. (2021), has strong internal validity; random assignment makes it possible to identify the effects of being raised in different families on the sample of adoptees. The question, however, is to what extent these effects generalize to the overall population of children.

There are two key reasons why the external validity of these studies may be limited. Parents who adopt may differ from non-adopting parents in their characteristics and behavior, and thus their influence on their children may be different as well; or the adopted children may differ in some way from non-adopted children. Fagereng et al. (2021) take a careful look at these threats to external validity. For example, they estimate the intergenerational transmission of wealth within the subsample of adoptive parents who have both biological and adopted children. The difference in wealth transmission between biological children and adoptees within this sample turns out to be roughly equivalent to the difference they find when comparing biological children and adoptees who have different parents. This indicates that the parents who adopt are similar to other parents when it comes to intergenerational wealth transmission.

External validity may be more limited in studies that use data on non-random domestic adoptions. Comparing non-adopting parents in Norway with parents who have adopted within the country, Fagereng et al. (2021) find substantial differences between these two groups. Hence, analysis based on non-random domestic adoptions may lack both internal and external validity.

2.4 Discussion

The research reviewed in this section suggests that less than half of the population variance in key outcomes, such education, earnings and wealth, can be explained by genes. This indicates that variation in nurture, i.e. in environmental factors, plays a key role in explaining individual differences in life outcomes.

In the remainder of the paper, we focus on studies of two specific environmental factors: home environment and preschool programs. These studies are more relevant to policy than heritability estimates for two reasons. First, heritability estimates are unable to isolate specific (and potentially offsetting) features of the environment. Second, heritability estimates of the explanatory power of genetic and environmental factors are based on the observed differences in child outcomes in the current environment. The purpose of policy interventions, however, is to

change a specific feature of the environment in order to cause a desired change in children's development.

3. Skill development in early childhood and beyond

All successful societies meet basic needs and provide socialization and educational experiences that prepare children to grow up and assume adult roles and responsibilities. In industrialized countries, the skills and capacities required of adults have become quite sophisticated, and include advanced literacy and communication skills, high-level analytic thinking and, increasingly, the social skills that allow individuals to get along with others and work in teams (Deming 2017). Developing these skills and capacities requires a complementary sequence of investments – by parents in the child's home environment, by educational institutions, and by governments, in supporting the efforts of parents and educational institutions.

But what skills to target with these investments? First off, it is important to note that we use the term "skills" broadly to encompass any skill, behavior, capacity or psychological resource that helps children attain successful outcomes. We follow the practice common in other social and behavioral sciences of referring to collections of "cognitive" skills (principally, IQ and achievement test scores) and a catch-all "socioemotional skills" category (sometimes lumped together as "noncognitive skills").

In the "socio-emotional" skill category, self-regulation and relationship skills appear to be of particular importance in early childhood (Buhs et al. 2006, Hamre & Pianta 2001, McClelland et al. 2007). Self-regulation can be defined as the ability to control thoughts, feelings and behaviors in order to adapt to the demands and social standards and expectations of the environment (Berger 2011), whereas relationship skills are the capacities to form attachments, relate to others and interact with other people.

Language, mathematics and socioemotional skills are often thought of as school readiness skills, as numerous studies have documented that they are strong predictors of success at the start of formal schooling and longer-run academic achievement and social adjustment (e.g. Buhs et al. 2006, Duncan et al. 2007, Hamre & Pianta 2001, McClelland et al. 2007, Romano et al. 2010). Our reviews in Sections 4.6 will demonstrate that even if these school readiness skills are strongly correlated with future skill development, this does not mean that early childhood education and parenting programs targeting these capacities will necessarily produce lasting

effects. Some of these skills are soon mastered by children in the comparison groups when they are taught in the early years of formal schooling. The challenge for preschools is to foster lasting skill development over and above skill-developing activities that most children enjoy in their home and later schooling environments or skills that are central for children's capacity to navigate the transition to formal schooling.

In Section 3.1 we begin with a discussion of skill-building models developed by cognitive psychologists and economists and highlight the importance of complementarity and substitutability across preschool, home environment, and elementary school educational investments. In Section 3.2 we introduce the concept of *trifecta skills* as important targets of early intervention. In Section 3.3 we discuss possible time-sensitive skills that may be foundational for young children to navigate the transition to formal schooling, and provide a discussion of future research directions.

3.1 Skill-building models

Key to the kinds of skill-building processes that economists have modeled is the well-researched idea that simpler skills can support the learning of more sophisticated ones. Sequential skill-building processes can be readily seen in math learning, for example. Counting serves as a cognitive basis for solving addition problems, and addition is often key to solving multiplication problems (Baroody 1987, Lemaire & Siegler 1995). In reading, similarly, children's ability to match letters to sounds is important for learning to recognize written words; this helps to build vocabulary, which in turn supports reading comprehension (LaBerge & Samuels 1974). Cunha and Heckman (2007) use the term "self-productivity" to describe these processes.

More generally, Cunha and Heckman (2007) describe a cumulative model of the production of human capital that allows for different childhood investment stages as well as roles for the past effects and future development of both cognitive and noncognitive skills. Their model highlights the interactive nature of children's skill building and investments by families, preschools and schools, and other agents.

Cunha and Heckman (2007) posit that human capital accumulation, as summarized in their phrase "skill begets skill," results from two distinct processes. As described above, the first is self-productivity, which suggests that the impacts of preschool interventions are particularly likely to persist when those interventions are designed to build skills incrementally within

developmental domains such as mathematics or literacy. Indeed, sequential learning goals embedded in curricular materials and instructional guidelines (e.g., the U.S. Common Core State Standards and the Next Generation Science standards) are designed to drive sequential skill-building in most K-12 schools around the world.

The second set of processes introduced by Cunha and Heckman (2007; also see Ceci and Papierno (2005)) are dynamic complementarity and substitutability. Bailey et al. (2020) explain these concepts in the following way: "Consider two children, Child A and Child B, who differ in their levels of counting knowledge; Child A has higher foundational skills than Child B. If both receive the same amount of teaching time and effort from teachers or parents to learn addition and subtraction, which child will profit the most from the instruction? If Child A will profit most, we say that the skill-building model features dynamic complementarity—the teaching investment complements a child's incoming level of foundational skills and produces a Matthew effect, where the rich get richer. On the other hand, if teaching investments are more productive for Child B, then we say that the model features dynamic substitutability—the teaching investment is compensatory by raising skills already mastered by Child A but not Child B (p. 68)." Education planners obviously want sequential investments that are as complementary to one another as possible.

The hypothesis of dynamic complementarity suggests that children who enter school with the strongest cognitive and noncognitive skills will profit most from K-12 schooling, for example by learning the most from classroom instruction. As reviewed in Section 4.1, Johnson and Jackson (2019) find positive interactions between attending Head Start and K-12 school funding. On the other hand, studies of more recent educational environments have shown that higher-quality kindergarten environments fail to interact in a positive way with the advantages of having been randomly assigned to effective preschool programs (Bailey et al. 2020).

The flipside hypothesis – of dynamic substitutability – is a very real possibility. Rossin-Slater and Wüst (2020) studied a Danish preschool program and a nurse home-visiting program that were both rolled out across communities in a haphazard way during the same period. Although each produced a long-term impact on educational attainment, children participating in both programs showed no larger longer-term improvements than children participating in just one of them. In other words, these two programs appeared to be dynamic substitutes for each other. Jenkins et al. (2016) studied outcomes for children who either transitioned from Head Start at age

3 into a Pre-K program at age 4 or remained in Head Start at age 4. They found much more dynamic substitutability in the Head Start/Head Start children than in the Head Start/pre-K children.

3.2 Trifecta skills

Cunha and Heckman (2007) speak generally of cognitive and noncognitive skills, but do not identify the skills that matter the most. Bailey et al. (2020) posit that to provide persistent intervention-generated benefits for young children, the skills, behaviors, capacities or beliefs targeted by preschool and home environment interventions must share three key features: they are malleable through intervention, they are fundamental for success, and they would not develop eventually in most counterfactual conditions. They characterize these skills as "trifecta" to highlight the importance of meeting all three criteria.

Consider first the dimensions of malleability and fundamentality. Bailey et al. (2020) argue that some of the most fundamental skills are not malleable and therefore not useful targets for preschool interventions. General intelligence is the best example. It is a fundamental capacity because it supports performance across a wide variety of important tasks and is the single strongest predictor of many important life outcomes (Cawley et al. 1997, Gottfredson 1997, Heckman 1995).

Unfortunately, attempts to augment general intelligence experimentally within the commonly observed range of intervention intensity and child characteristics have rarely proved successful (Jensen 1998), (but also see Nisbett et al. (2012) for a more optimistic review). Indeed, only one of the best-known early childhood education interventions – Abecedarian – generated persistent effects on children's general intelligence, perhaps because of the intense nature of the Abecedarian program, combined with the conditions of relative deprivation facing control group children and their families.

Among noncognitive skills, conscientiousness – one of the "Big Five" personality traits – also appears to be fundamental. It is a powerful correlate of both schooling and later life outcomes (Almlund et al. 2011, Bogg & Roberts 2004, Judge et al. 1999, Poropat 2014). And while ample evidence suggests that many socioemotional measures can be altered with interventions (Almlund et al. 2011) and may be important mediators of the impacts of early childhood education programs on adult outcomes (e.g. Heckman et al. 2013), few studies provide

measures of core personality characteristics and preferences that would shed light on whether programs can change them. An interesting exception is the study of Cappelen et al. (2020), who show that a center-based preschool program in Chicago Heights increased egalitarian preferences two years after the end of the program.

Bailey et al. (2020) suggest that the skills most likely to meet the twin malleability and fundamentality criteria include proficiency in math and literacy, an ability to regulate one's attention and emotions, and the social and relationship skills that enable children to work well with peers and teachers. But targeting these kinds of skills during the early childhood period may be insufficient for preschool and home environment interventions to generate persistent impacts, because this neglects the third leg of the trifecta stool: many of these skills are soon mastered by children in the comparison groups. For example, although knowing the sounds corresponding to the letters of the alphabet is essential for learning how to read words, it is a skill that almost all children will eventually learn in elementary school, whether or not they receive targeted intervention in early childhood (Paris 2005). Counting is an analogous universally learned skill from early mathematics.

Many early cognitive and noncognitive skills develop very quickly in counterfactual conditions across the early and middle-childhood periods. For example, nationally normed reading and mathematics tests show that children learn more than a full standard deviation of material between kindergarten and first grade, but considerably less in later grades (Hill et al. 2008). Thus, while these early cognitive skills may be among the most fundamental and malleable, the effects of interventions targeting these skills may also fade out most quickly owing to control group catch-up. A preschool intervention example of fadeout caused by rapid growth among children in counterfactual conditions comes from the Clements et al. (2013) evaluation of the TRIAD/Building Blocks pre-K math intervention. We will discuss this study in Section 4.1.

Also in Section 5, we will argue that interventions that emphasize basic math and literacy skills are probably a poor choice if preschool graduates are likely to transition to kindergarten teachers who focus on children lacking those basic skills. A more promising approach might be to implement a sequence of skill-building interventions across both the early childhood years and the early grades – perhaps ages 3 to 8 – that build on one another and promote dynamic complementarity during an early intervention period in which skill-building is particularly rapid.

Moreover, as we discuss in the next section, it may be important to target skills that are central for children's capacity to navigate the transition to formal schooling.

3.3 Discussion and future research directions

It is relatively easy to identify skills needed to succeed in modern industrial societies (e.g., analytic thinking, communication, working in teams), as well as the foundational capacities that correlate most strongly with higher levels of these skills (IQ, conscientiousness). But it is quite another matter to identify skills that early educational and parenting interventions should be targeting, because productive interventions need to target skills that would not otherwise be developed in families and schools. Alternatively, they must be central for children's capacity to navigate the transition to formal schooling.

For addressing the question of which skills to target, the Bailey et al. (2020) conception of "trifecta" skills could be useful, which are defined to be fundamental for success, malleable through intervention, and unlikely to develop eventually in most counterfactual conditions. And while important foundational elements of literacy and numeracy (e.g., knowing letters and numbers) might be obvious candidates for interventions seeking to promote school readiness, in most K-12 school systems children will quickly develop these skills if they begin their formal schooling without knowing them. If, on the other hand, K-12 schooling is structured to provide dynamic complementarity to skills developed in homes and preschools, then early investments can be quite profitable.

Beyond trifecta skills is the possibility that early time-sensitive skills can open doors for children in the education system by, for example, enabling them to learn more effectively from teachers and peers, or to avoid unproductive elements of school structures such as grade failure and special education tracks. In other words, some skills may be uniquely important for developmental transitions, and are thus time sensitive. Transition from preschool to primary school often entails a substantial shift in pedagogical practices from a play-based to a more instruction-based approach. As a result, children face new demands on their ability to regulate behaviors, such as paying attention and following instructions (Blair & Diamond 2008, DiPrete & Jennings 2012). In addition, they have to form new friendships, often with less support from teachers scaffolding development of good relationships. Among many possible skills, a theoretical possibility is that socioemotional skills and executive functioning help children handle

these challenges, which may give them a better start socially, emotionally and academically (Blair & Diamond 2008). For example, a child who cannot listen and pay attention at school start may alienate her teachers and peers, even if he eventually catches up on attention skills.

To conclude, identifying trifecta skills is an important priority for future research. Additionally, we need more research investigating skills that are critical to navigating the transition to formal schooling. As eloquently formulated by Elder Jr (1998): "Early transitions can have enduring consequences by affecting subsequent transitions, even after many years and decades have passed. They do so, in part, through behavioral consequences that set in motion cumulating advantages and disadvantages."

4. Impacts of preschool participation

Attending preschool can promote a child's development in several ways. First, it can strengthen children's school readiness skills, for example by increasing the amount of time spent in a structured educational setting and reducing the time spent in other modes of care, such as parental care, other family care, informal care, or alternative formal care. However, whether substituting preschool for another type of care is beneficial or detrimental for child development depends on the capacity of preschool to build skills that are important for child development, relative to the development of those skills in other type of care.

Second, by reducing the time some parents spend caring for their children, preschool enrollment may lower parental stress and improve the quality of interactions during the more limited hours parent and child spend together. Third, no- or low-cost preschool in effect transfers resources to parents, which may be spent exclusively on the child or used to meet other family goals. Family resources may also increase if the preschool enrollment enables parents, and especially mothers, to work more, which may benefit child development because it provides income that can be spent on goods and services to enrich children's development. On the other hand, more market work may increase maternal stress, leading to less appropriate parenting.

Fourth, preschool enrollment allows parents to learn from interacting with teachers and observing their child's classroom. Gelber and Isen (2013) found that parents whose children were randomly assigned to a Head Start program showed increased involvement in at-home interactions with their children, such as reading and math activities. They also found that non-resident fathers were more involved with their children.

Below we provide a brief introduction to the literature concerning the causal impact of preschool enrollment on child development. We first review evidence on programs targeting disadvantaged families, then look at universal programs, which are open to all families. It can sometimes be difficult to distinguish between targeted and universal programs. Some universal programs are, in principle, open to all, but access is limited and low-income families are given priority. This is true of the Tennessee Voluntary Prekindergarten Program, which we have classified here as a targeted program. Much of the strongest evidence on targeted preschool programs comes from randomized controlled trials in the United States, whereas most of the strongest evidence on universal programs is derived from natural experiments in Europe and Canada.

Finally, we note that decisions about government funding of preschool enrollment should be based on the value of its benefits to children, their families and society as a whole as well as on the costs to parents and taxpayers. With per-child program costs in today's dollars ranging from a few thousand dollars to more than $100,000, in the case of the Abecedarian program, merely demonstrating statistically significant impacts on some measures of child development and adult attainment is not sufficient. Public subsidies and investments should be based on evidence that benefits exceed costs.

4.1 Participation in targeted preschool programs

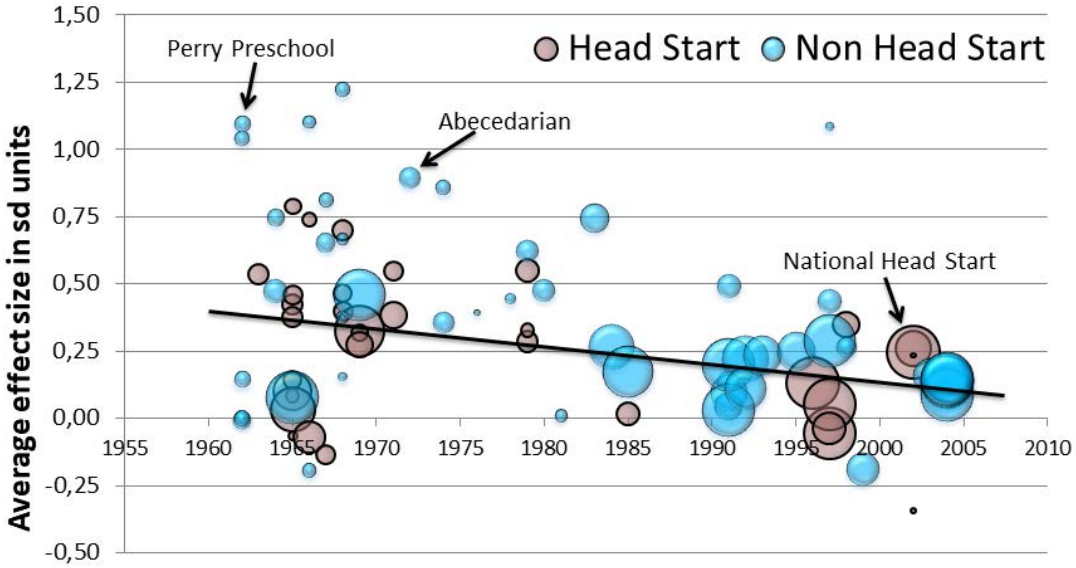
Evaluation studies of preschool programs targeting low-income families date back to the early 1960s – nearly 60 years, with virtually all of the studies based in the U.S. Many of the early studies evaluated Head Start summer programs that were being rolled out at the time; some, like the famous Perry Preschool and Abecedarian programs, were very high-quality programs designed and run by researchers. Even after one restricts the evaluations to those with reasonably strong research designs (e.g., random assignment, quasi-experimental designs such as sibling fixed effects or propensity score matching after demonstrating baseline equivalence), the number of evaluations easily exceeds 100. Our review takes advantage of the fact that Duncan and Magnuson (2013) and Li et al. (2020) provide summaries of U.S. studies published between 1960 and 2007, about one-third of which employed random assignment designs. After reviewing results from this meta-analysis of older programs, we provide a more selective look at highly

effective programs (Perry and Abecedarian), Head Start and several recent targeted pre-K programs.

Adapted from Duncan and Magnuson (2013, Figure 2), Figure 1 shows the distribution of 84 program-average treatment effect sizes for cognitive and achievement outcomes, measured at the end of each program's treatment period, by the calendar year in which the program began. Bubble sizes are proportional to the inverse of the squared standard error of the estimated program impact. Thus, impacts associated with the very small sample sizes in the Perry and Abecedarian evaluations are represented by small circles, while impacts from the much larger National Head Start Impact study are represented by much larger circles. Figure 1 also differentiates between evaluations of Head Start and all other early childhood education programs, and includes a regression line fit to effect sizes by calendar year that is weighted by the inverse of the squared standard errors of the impact estimates.

Across all the studies, the unweighted average effects size for early childhood education on cognitive scores was .35 standard deviations at the end of the programs, an amount equal to about one-third of the kindergarten achievement gap between children from families in the top and bottom income quintiles (Duncan & Magnuson 2013). However, as Figure 1 demonstrates, effect sizes vary substantially; studies with the smallest number of subjects tend to generate the largest effect sizes. When weighted by the inverse of the squared standard errors of the effect-size estimates, the average drops to .21 standard deviations.

Figure 1 Average impact at end of treatment



Some of the programs lasted for only a couple of summer months, while others continued for as long as five years. Some of the evaluations used random assignment, whereas others relied on less rigorous quasi-experimental methods. Almost all focused on children from low-income families, but they varied in the racial and ethnic composition of treatment groups.

One might assume that these differences would account for much of the effect-size variability observed in Figure 1. However, that is not always the case. Duncan and Magnuson (2013) report that weighted average effect sizes were insignificantly larger for evaluations that did and did not use random assignment (.25 and .19 standard deviations, respectively), and for studies that were and were not published in peer-reviewed journals (.31 and .18 standard deviations, respectively). One notable difference is that effect sizes of programs designed by researchers (.39 standard deviations) were significantly larger than in the case of programs not designed by researchers (.18 standard deviations), which may reflect the difficulty of scaling up model programs. The downward slope of the trend line is driven by the fact that programs beginning before 1980 produced significantly larger end-of-program effect sizes (.33 standard deviations) than those that began later (.16 standard deviations).

Secular declines in effect sizes are disappointing, given hopes that lessons from early evaluations and advances in the science of child development would lead to more effective programs and larger program impacts over time. One likely explanation is the substantial

improvement in the care environments of children in the counterfactual comparison groups. Duncan and Magnuson (2013) report that between 1970 and 2000, the education levels of mothers of preschool-age children in the bottom 20 percent of the income distribution rose dramatically, while family sizes fell, which means that the quality of parental care, as an alternative to early childhood education, is likely to have increased as parental time has been spread across fewer children. Expenditure survey data show that despite stagnant family incomes, child enrichment expenditures by U.S. families in the bottom quintile grew by more than 50% between 1972 and 2005 (Duncan & Murnane 2014, Figure 3.2). Similarly, Bassok et al. (2016) report substantial (although by no means complete) convergence in recent years between low- and high-income families in young children's exposure to books and reading in the home, access to educational games on computers, and engagement with parents, both inside and outside the home.

Another important factor is the increasing availability of center-based care. The fraction of low-income three- and four-year-olds in center-based care nearly tripled between 1970 and 2010 in the U.S. Heckman et al. (2000) and Kline and Walters (2016) discuss the challenges of drawing lessons from social experiments when control-group participants are able to access other, close-substitute programs. Although close center-based substitutes may reduce the apparent intent-to-treat impact of a given preschool program, the entire collection of available preschool options may still produce important impacts relative to home and other less formal care arrangements.

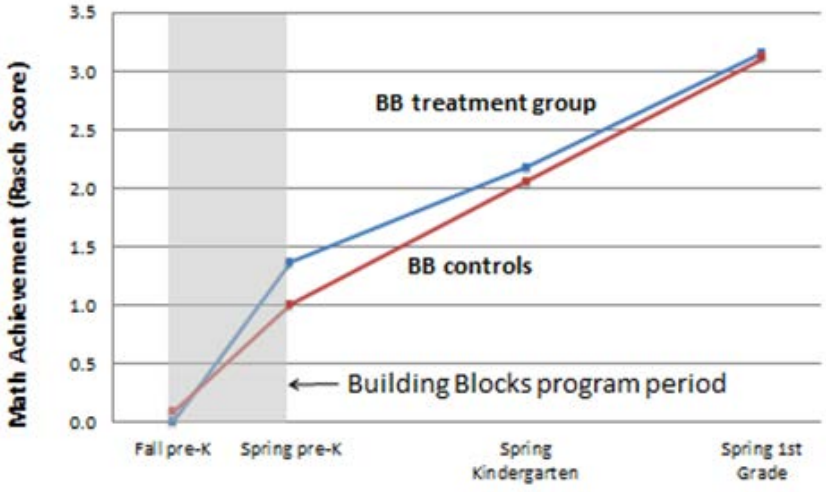
Higher-quality home environments and the increasing availability of close center-based substitutes complicate the task of drawing lessons from evaluation research literature. Secular increases in home and center-based care environments make it difficult for both targeted and universal programs to generate impacts large enough to offset the costs of these programs. Moreover, these trends raise questions about the generalizability of long-run evaluations of programs begun decades ago.

Fadeout. All of the impacts shown in Figure 1 were measured at or near the end of the program. In about half of the studies, children were tracked beyond the end of the program, making it possible to observe patterns of fadeout. Li et al. (2020) find that effect sizes drop roughly in half within 1-2 years, and then again several years later. Reviewing patterns of fadeout in interventions targeting children's academic skills, Bailey et al. (2020) find that rapid fadeout

of impact on these skills is the norm rather than the exception, according to all research syntheses.

A preschool intervention example of fadeout comes from the Clements et al. (2013) evaluation of the TRIAD/Building Blocks pre-K math intervention conducted in two large low-income school districts – Buffalo and Boston. With math scores scaled vertically in standard deviation units, Figure 2 shows that the math achievement of children in the control group grew by nearly a full standard deviation between the fall and spring assessment points during the pre-K year, and then by about a full standard deviation in the annual intervals between spring of the pre-K year and spring of kindergarten as well as between the springs of kindergarten and first grade. These are remarkable growth rates and far exceed the typical impacts found in intervention studies. Math achievement for children receiving the Building Blocks curriculum grew even faster during the pre-K year – about .50 sd faster – than controls, but less quickly than controls after that. The shrinking gap (i.e., fadeout) between the two groups appears to result more from "catch up" on the part of control-group children than from forgetting or related skill depreciation on the part of the Building Blocks children.

Figure 2: Math Achievement During and After the Pre-K TRIAD Program



Note: Rasch score standard deviation = 1

It is important to point out that the fadeout shown in Figure 2 occurred in the context of the early-grade curricula and instruction in Buffalo and Boston at the time of the interventions. As we elaborate below, the structure of both public preschool and K-12 education are

manipulable through policy, potentially in ways that reduce fadeout by enhancing the degree of dynamic complementarity.

Intensive model programs. The ability of intensive model programs to improve the life chances of disadvantaged children is illustrated by the well-known Perry Preschool and Abecedarian interventions. Evaluations of both programs employed random assignment, and follow-up studies have tracked participants well into their adult years.

Perry provided one or two years of part-day educational services and home visits delivered by highly educated teachers to a sample of low-income, low-IQ African American children aged three and four in Ypsilanti, Michigan, during the 1960s. Perry's curriculum fell into the "whole child" category discussed in Section 5.4 because it focused on quality of relationships and experiential learning The per-pupil cost of Perry in 2021 dollars was $23,800 – more than twice the per-pupil cost of the full-time Head Start program in the U.S. and four times the funding allocated to many state pre-K programs (Heckman et al. 2010, Table 8). Low-quality counterfactual conditions in home environments are reflected in low high school graduation rates for mothers (21%) and fathers (11%) and large family sizes (6.7, on average) (Schweinhart 1993).

The Abecedarian program, which began in 1972 and enrolled children of low-income, mostly African American women from Chapel Hill, North Carolina, was even more intensive than Perry. Mothers and children assigned to the Abecedarian "treatment" received year-round, full-time care for five years, starting with the child's first year of life. The program included transportation, individualized educational activities that changed as the children grew older, and very low child-teacher ratios (3:1 for the youngest children and up to 6:1 for older children). Abecedarian's curriculum was "skills-based," with a strong and persistent emphasis on language. Per-pupil costs of Abecedarian in 2021 dollars averaged $105,500 – a huge number. Although most of the children received some kind of non-parental care over the five-year project period, none was as intensive as the care provided by the Abecedarian treatment. The nature of counterfactual home environments for Abecedarian is reflected in the low completed schooling (10.2 years), IQs (85.5) and age at enrollment (19.9) of the mothers (Ramey et al. 2000).

Despite these differences, benefits for both programs far exceeded costs. Heckman et al. (2010) estimate benefit/cost ratios for Perry that range between 6:1 and nearly 9:1, with benefits driven in roughly equal measure by increases in earnings and reduced crime. In the case of Abecedarian, García et al. (2020) estimate benefit/cost ratios exceeding 6:1, with more than two-

thirds of the benefits driven by crime reductions and the remainder reflecting differences in adult health and the labor income of participants and their parents (the five years of full-time childcare enabled parents to establish and maintain more continuous and higher-paying careers).

Although mediational analyses have been conducted for both programs (e.g., Heckman et al. (2013) for Perry and Pages et al. (2020) for Abecedarian), it is useful to look across both projects and ask whether any broadly defined pathways are similar. Perry's impressive initial impacts on IQ scores (as shown in Figure 1) famously faded out by third grade, while the IQs of Abecedarian children continued to exceed those of control-group children through age 21 (Campbell et al. 2001, Schweinhart 1993). Overshadowed by Perry's fading IQ impacts is the fact that both Perry and Abecedarian produced consistently substantial impacts on verbal and quantitative achievement tests (Campbell et al. 2001, Schweinhart 1993) across childhood and adolescence. Cognitive skill growth may well have been an active ingredient in the longer-run successes of these programs.

As for noncognitive skills, Perry significantly reduced problem behavior, especially among boys, which may explain a substantial proportion of the program's effects on boys' crime and employment outcomes (Heckman et al. 2010). Perry also improved a host of measures of non-test-score academic skills and socio-emotional skills, some of which have been interpreted as measures of personality (Heckman et al. 2013). Fewer consistent impacts on K-12 noncognitive skills were apparent in Abecedarian. In fact, Haskins (1985) reported that the program generated an unexpected increase in teacher reports of children's aggressiveness in the early school years, although these effects appeared to fade with time. No impacts were found for an assortment of child self-ratings of behavior in early adolescence.

A final question is whether these two programs had positive impacts on school structures that might boost attainments, specifically avoiding special education classes and grade retention. Here again the verdict is mixed – Abecedarian had positive impacts on both of these outcomes, while Perry had positive impacts on neither (McCoy et al. 2017).

The Perry and Abecedarian evaluation literatures are consistent in supporting a cognitive pathway that operates through school achievement but not IQ itself. Given reductions in adult crime, one might expect to find support for reductions in problem behavior across childhood and adolescence, but while Perry supports that pathway, the more scattered evidence from Abecedarian does not. And whereas the consistently positive impacts of Perry on a broader range

of socio-emotional skills suggest other possible pathways, the Abecedarian evaluations failed to collect sufficiently comparable information to confirm or disconfirm them.

Have counterfactual conditions caught up with Head Start? The Head Start program is probably the most thoroughly researched, scaled-up, targeted program in the United States. Begun in 1965 as part of Lyndon Johnson's War on Poverty, in its early versions the Head Start program often spanned only the three summer months, but quickly expanded to include part- or full-day services over much of the year. Although academic competence was always a goal, so too was improved child health. From the very beginning, Head Start services included assessments of nutrition, vision, hearing, and speech, as well as tests for conditions such as tuberculosis and anemia.

It is important to remember these services when considering counterfactual environmental conditions in evaluations of the early Head Start cohorts. For the most part, conditions in which poor children, especially Black children, were growing up in the early years of the program are best described as deplorable. In the mid-1960s, prior to the introduction of the Food Stamp program and well before tax credit programs such as the Earned Income Tax Credit were introduced, government taxes and transfers tended to *reduce* the incomes of low-income families and increase child poverty. Racial discrimination in parts of the country denied Black women access to quality schools and hospital care, including childbirth in hospitals with a physician present (Almond 2006). These factors, plus the differences in family size and maternal schooling mentioned earlier, produced counterfactual conditions that made it much easier for a Head Start program to demonstrate effectiveness than it is for modern-day Head Start programs.

The short-run cognitive impacts of a number of Head Start programs were evaluated in the 1960s and 1970s (see Figure 1). Estimates of impacts vary widely, with none as high as the impacts of Perry and Abecedarian and some even negative. Most fall into the .25-.75 sd range. Longer-run impacts from the early decades of the Head Start program speak more directly to the policy issue of effectiveness.[7] Three examples: Ludwig and Miller (2007) took advantage of a discontinuity in how technical assistance was offered to officials in low-income counties during the very early years of the program, and found that Head Start availability was associated with

[7] A recent summary of this literature can be found in: https://journals.sagepub.com/doi/suppl/10.3102/0162373720948884/suppl_file/Supplementary_Table_A1.docx

reductions in child mortality and suggestive evidence that it was associated with increases in completed schooling.

Both Bailey et al. (2020) and Johnson and Jackson (2019) take advantage of geographic differences in the rollout of the Head Start program between 1965 and 1980. In one case (Bailey et al. 2020), authors link the rollout to differences in adult outcomes, based on Census Bureau data, while in the other (Johnson & Jackson 2019) they find differences (including sibling differences) in outcomes based on PSID data. Bailey et al. (2020) find differences in a broad set of educational outcomes as well as increases in labor supply (but not earnings) and reductions in poverty. Interestingly, the point estimates of their completed schooling impacts are smaller than in most of the ten other quasi-experimental Head Start studies they review (see their Figure 5), but their Census-based data provide much more statistical power to detect small impacts.

Johnson and Jackson (2019) find that the availability of Head Start was associated with increased educational attainment and earnings, as well as reductions in the likelihood of both poverty and incarceration in adulthood. Highly relevant to the issue of dynamic substitutability and complementarity, discussed in Section 3, is their finding that the impacts of Head Start, when combined with court-ordered K-12 school finance shocks, were larger than the effects of either factor alone.

This evidence leaves little doubt that the Head Start program was effective in its early decades. More recent evidence is mixed. Sibling-based studies using data from the NLSY Child Supplements have grown in number and the length of follow-up periods, as the project has continued to collect follow-up data. Two of the most recent studies are Deming (2009) and Pages et al. (2020). Deming (2009) covered cohorts born between 1976 and 1986, whereas Pages et al. (2020) are able to draw on an additional 10 years of the NLSY, which enables them to extend the outcome measurement period for Deming's cohorts as well as to apply Deming's sibling-based specification to cohorts born between 1986 and 1996.

Deming (2009) reaches optimistic conclusions regarding program impacts, finding that the siblings who participated in Head Start score nearly a quarter of a standard deviation higher on an index of young adult outcomes (e.g., completed education, crime, teen parenthood) than their siblings who did not participate in the program. An examination of shorter-run impacts on test scores failed to find consistent evidence of sustained effects. When they extend the measurement interval for Deming's adulthood outcomes, Pages et al. (2020) find no statistically

significant impacts on earnings, and mixed evidence of impacts on other adult outcomes. When they apply Deming's sibling comparison framework to ten more recent cohorts of children born to NLSY mothers, they find mostly negative Head Start impacts. Combining cohorts reveals generally null impacts on school-age and early adulthood outcomes. Clearly, the jury is still out when it comes to the impact of Head Start on participating children born 20 to 40 years after the program's inception.

Long-term outcomes are not yet available for cohorts born during the past 25 years, but shorter-run outcomes were tracked in a very high-quality RCT study of the Head Start program that began in the early 2000s (Puma et al. 2010). The National Head Start Impact Study (NHSIS) was able to draw a large probability sample of Head Start centers, and then worked with center directors to generate and randomly select children from waitlists (n=4667 in all) to be offered enrollment. The study included both three- and four-year olds applying for entry into sampled Head Start centers and conducted follow-up data collections through third grade.

Apart from rigorous random assignment, what distinguishes the Head Start Impact study is the quality and breadth of its measurement of academic and socioemotional skills, both prior to random assignment and in all follow-up waves. School achievement skills were measured with vertically scaled assessments. Both parents and teachers reported on a range of socioemotional capacities, including problem behavior, prosocial skills and, beginning in kindergarten, school engagement. Many of these skills were measured with several validated instruments.

ITT results from the study are summarized in Table 2 with simple counts of the number of impact coefficients that exceeded a low threshold – $p<.10$ in the absence of adjustments for multiple testing both at the end of the Head Start year and the end of 1st grade (3rd grade results are similar to 1st grade results). Because the evaluation included quite a number of assessments within the broad domains of literacy, math, and problem behavior, the table also shows the total number of assessments. Results are striking, showing that end-of-treatment impacts are significant only for measures in the literacy domain and for virtually none of the assessments at the end of 1st grade. With effect sizes at the end of the program for literacy measures averaging about .20 sd, it appears that children whose parents were offered Head Start slots enjoyed modestly higher growth in literacy skills. No impacts on math or behavioral measures were detected. At the end of 1st and 3rd grade, there was virtually no evidence of above-chance gains in any of the three domains.

Table 2: Ratio of statistically significant ($p < .10$) treatment impacts to outcomes examined in the Puma et al. (2012) Head Start Impact Study

	End of Head Start year	**End of 1st grade**
Literacy	6 of 8	1 of 11
Math	0 of 2	0 of 4
Problem behavior	0 of 9	1 of 9

ITT estimates are useful for policy questions such as the likely impact of expanding the number of slots in the Head Start program. Table 2 suggests that academic and behavioral gains are unlikely to persist even a few years beyond the Head Start year. Of course, the 24 outcome measures gathered in the 1st-grade follow-up might not capture the "dark matter" skills that matter most for long-term impacts (Gibbs et al. 2011, Heckman & Rubinstein 2001). But with the evaluation's intensive focus on measuring the broad range of early-grade skills that the research community identified as important, the dark matter channel to longer-run benefits from the current implementation of Head Start that might exceed the program's $10,000 per slot price tag seems far from certain.

Kline and Walters (2016) argue for a more sophisticated policy analysis of the Head Start Impact Study data. Increasing the number of Head Start slots won't increase social costs by the full ~$10,000 if children who would otherwise enroll in subsidized slots in other center-based programs are drawn into Head Start instead. Like Heckman et al. (2020), moreover, they underscore the importance of understanding counterfactual conditions when assessing impacts. Averaging across the two cohorts, Kline and Walters (2016) show that nearly a third of the lottery losers enrolled in other center-based programs, leaving 55% cared for in the home or in some other non-preschool setting. After addressing selection issues, they conclude that Head Start has larger positive effects on the end-of-program test scores of children who would otherwise have been cared for at home and insignificant effects on children who would otherwise have attended other preschools. They do not provide selection-adjusted estimates of impacts after children entered elementary school.

Two RCT studies of state pre-K programs. State-funded pre-kindergarten programs have expanded rapidly in the United States in the last 20 years. They now enroll nearly one-third of all

4-year-olds and are offered in 37 states (Phillips et al., 2017). As of 2017, 29 short-term studies, many based on birthday-cutoff RD designs, had been published, as had 39 studies of longer-run impacts (Phillips, 2017). Most of the longer-run studies are based on state records of pre-K enrollment matched to achievement test scores, and they use propensity-score matching methods without being able to control for pre-enrollment information. However, random-assignment evidence on longer-term impacts is available for two programs: – the Tennessee Voluntary Pre-K program and Boston's pre-K program, which we focus on in this brief review section.

The Tennessee Voluntary Pre-K program (TN VPK) is a large, state-funded pre-K program that annually enrolls about one-quarter of the state's four-year-olds. Although technically available to all four-year olds living in the state, TN VPK grants priority to children whose families qualify for the means-tested free or reduced-price lunch program. We therefore classify it as a "targeted" rather than a "universal" program.

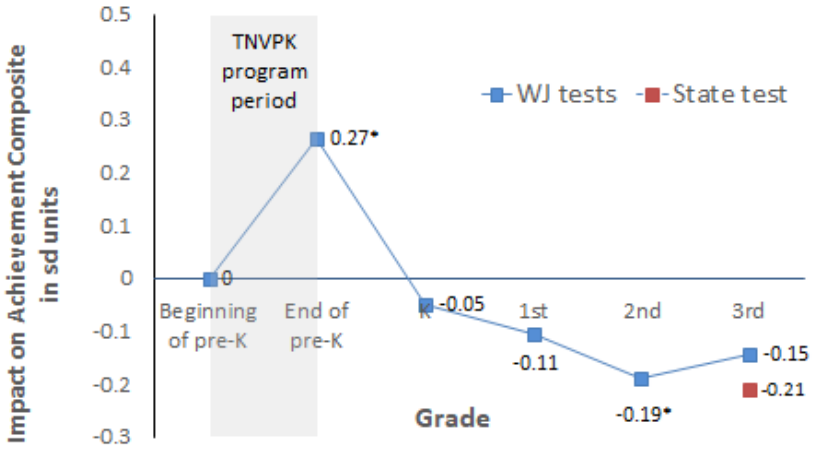
By conventional U.S. standards, the quality of the program is fairly high: It is a full-day program with relatively small class sizes (maximum of 20 students), and it requires both a licensed teacher and a teacher aide in each classroom (this keeps the child-to-adult ratio at or below 10 to 1). The curriculum must be chosen from a state-approved list.

What distinguishes the TN VPK evaluation is that the study, conducted by researchers at Vanderbilt University, randomly assigned children to TN VPK through a lottery system implemented at over-subscribed programs (Lipsey et al. 2018). Moreover, the researchers embedded their own series of surveys and assessments into the study, so that the evaluation was able to drawn on rich sources of test and survey report data.

Figure 3, which is taken from Watts et al. (2019), shows IV estimates of program impacts using assignment to pre-K as the instrument. At the end of the TN VPK program year, children attending the program outscored their comparison-group counterparts by about one-quarter of a standard deviation on an achievement test composite – a result consistent with many of the RD estimates of end-of-pre-K impacts. But these gains had disappeared completely by the end of kindergarten and became negative and statistically significant by second grade. A follow-up in sixth grade provides even strong evidence of perverse achievement and behavioral impacts (Durkin et al. 2022). School data showed few differences for absences, placement in special education or gifted-and-talented tracks, or behavioral outcomes. Survey data also revealed few differences in these measures, although TN VPK children had somewhat more behavior problems

and less positive feelings about school than their control-group counterparts in the early grades. As with the Head Start Impact Study, the pattern of fading achievement impacts and largely null behavioral impacts makes it seem unlikely that longer-run impacts will emerge.

Figure 3: TNVPK program impacts on achievement



Source: Watts et al. (2019)

A longer-term and quite different picture of the efficacy of targeted pre-kindergarten programs emerges from a recent analysis of the early years of Boston's program (Gray-Lobe et al. 2021). The authors use admissions lotteries for cohorts entering the program between 1997 and 2003 to estimate preschool impacts on K-12 school performance and behavioral outcomes, college entry, and college attainment. We categorize this study, like the Tennessee program, as "targeted" because children from disadvantaged backgrounds were heavily prioritized in the assignment system.[8]

Before we turn to impacts, a few remarks about program quality: Boston's pre-K program is recognized to be a very high-quality program, with a curriculum based on proven, play-based math, literacy and behavioral components, and with well-paid, highly credentialed teachers and, in some years, coaches to help guide teachers in implementing the curriculum (Duncan & Murnane 2014, Chapter 4). However, none of these components was in place between 1997 and 2003. In fact, an evaluation of these pre-K classrooms published in 2006 characterized them as

[8] Note, however, that the title of the study is "The long-term effects of universal preschool in Boston".

hobbled by mediocre instruction and as unsanitary and ill-equipped. All in all, the quality of the 1997-2003 pre-K program evaluated by Gray-Lobe et al. (2021) appears to be mediocre at best.

As with Table 2's Head Start summary, the general pattern of impacts in the Boston pre-K study is shown in Table 3 with counts of the number of coefficients that pass a $p<.10$ threshold. What stands out is the complete absence of impacts for the 24 measures of K-8 achievement scores and early school progress (e.g., avoiding grade failure and special education placement), coupled with above-chance achievement (e.g., state and SAT test-taking and scores) and behavioral impacts (e.g., suspensions, juvenile incarceration) in high school. The researchers find favorable impacts on 7 of 11 measures of high school graduation and college attendance. Some of these impacts are more favorable for boys than girls and for poor than non-poor children, but none of the poor/non-poor impact differences exceeded a $p<.05$ threshold.

Table 3: Ratio of statistically significant ($p < .10$) treatment impacts to outcomes examined in the Gray-Lobe et al. (2021) Boston Pre-K study

	K-12	**Post-secondary**
K-8 test scores	0 of 18	-
K-8 on-track progress	0 of 6	-
High school behavior problems	3 of 7	-
High school tests	7 of 21	-
Completed schooling	2 of 2	5 of 9

TN VPK and Boston pre-K evaluations produce puzzlingly disparate patterns. It seems clear that participation in Boston's mediocre-quality pre-K program produced favorable academic and behavioral impacts in adolescent and early adulthood. But none of the impacts from 24 measures of test scores and school progress taken prior to high school and available in administrative records passed muster at $p <.10$. As in the case of the TN VPK evaluation, Boston pre-K impacts on test scores through eighth grade were generally null, with as many point estimates negative as positive. Impacts on early grade repetition and special education placement

were also null. Qualitatively similar patterns (through 3rd grade) can be found in the NHSIS. We discuss implications of these patterns of effects in the concluding section of this chapter.

4.2 Participation in universal preschool programs

European countries, including the U.K., France, Germany, and all Nordic countries, invest heavily in universal preschool programs. Universal programs are also available in several U.S. cities and states as well as in Quebec, Canada. Investments in universal preschool are motivated largely by some of the research reviewed in Section 4.1, which suggests that some targeted preschool programs can boost child development and generate long-term impacts on school achievement and adult labor market participation. But that section also shows that the evidence for targeted programs is more mixed than what is reflected in many enthusiastic policy briefs, especially for long-run outcomes. Moreover, evidence for programs targeting disadvantaged children does not necessarily extrapolate to universal programs, which are implemented at scale and open to all.

Although many of the mechanisms that generate benefits to children are shared by universal preschool and targeted programs, two are unique to universal programs. First, children in universal programs may benefit more from positive peer effects. As discussed in Section 6, gaps in cognitive and social and emotional skills across family backgrounds are already substantial in early childhood. Disadvantaged children may benefit from playing and learning in interaction with more advantaged children. Consistent with this conjecture, Sojourner (2013) finds that kindergartners benefit from being randomly assigned to classrooms with high achieving peers in Project STAR.

Second, when entire cohorts enter K-12 schooling with higher skills as a result of their preschool experiences, teachers are able to build on those skills. This relates to the issue of dynamic substitutability and complementarity discussed in Section 3. It is difficult to provide proper instruction to individual children if children enter school with very different cognitive and socioemotional skills. Teachers often give priority to helping the low-achieving students achieve a certain minimum level of literacy and numeracy skills. This reflects a kind of dynamic substitution between preschool and formal schooling – the more a child learns in preschool, the less he or she will learn during the early school years, which may be why many studies find positive effects of preschool fading over time. In contrast, universal preschool can ensure that

most students have reached a minimum threshold of skills by the beginning of school, which makes it easier to teach all children at a level that takes full advantage of their capabilities, leading to dynamic complementarity.

Of course, publicly provided universal preschool is more costly than targeted programs, as it includes all children. The costs may not be outweighed by the benefits, especially if, as some of the evidence suggests, only disadvantaged children benefit from universal preschool. Another concern is that freely available public preschool programs can crowd out private providers. These public funds might have been spent more efficiently elsewhere. We will return to these concerns in Section 4.4.

During the last decade, several papers in economics have investigated the effects of universal preschool enrollment on child development. These papers, summarized in Table 4, differ substantially in terms of program content, context, study design and outcomes. However, all of them address concerns of non-random selection into preschool. They do this either by using government reforms that substantially increased universal preschool availability, providing variation in preschool access across space and time, or by using lotteries in areas with oversubscription. Moreover, all identify total impacts of preschool, reflecting not only effects of program participation itself but also potential changes in parental investments, parental stress or other parental behaviors in response to the program (see mechanisms described above), changes in school investments, or peer effects. We summarize the findings from this literature, discuss some of the insights gleaned from these studies and then point to directions for future research.

Our review distinguishes between studies focusing on 3- to 6-year olds in preschool and studies looking at the effects of center-based childcare on children ages 0-3. Attachment theory suggests that children's developmental needs change substantially between these phases of early childhood. It defines the attachment phase as the period from 6 months to two years of age. During this phase, early separation from the primary caregiver, typically the mother, may cause stress and anxiety in the child, with potentially adverse effects on children's development (Bowlby 1973). Moreover, the counterfactual mode of care often differs between these two phases of early childhood. In many European countries, most mothers return to the labor market before the child turns three. Thus, studies investigating the effects of universal preschool participation among children ages 3-6 are likely to be observing the effects of switching from informal or other forms of formal care to the subsidized universal care.

Table 4. Evaluations of Universal Preschool Programs

Where	Age	Preschool treatment	Counterfactual	Treatment effect	Family background heterogeneity	Years	Studies
Universal preschool ages 3-6							
Uruguay	4-5	n.a.	n.a.	Negative effects on grade retention and early drop-out in elementary school	Larger effects for children of mothers with lower-than-average education	90s	Berlinski, Galiani, and Manacorda (2008)
Argentina	3-5	Curriculum targeting communication skills, social and emotional skills, and mathematics	n.a.	Positive effects on third grade test scores	Larger effects for children living in more disadvantaged municipalities	90s	Berlinski, Galiani, and Gertler (2009)
Norway	3-6	College-educated preschool teacher and two assistants per 16 children. Limited curricular focus.	Informal care	Positive effects on labor market attachment, earnings and education at age 30	Larger positive earnings effects for children from low-income families. Negative earnings effect for children from high-income families.	70s	Havnes and Mogstad (2011, 2015).
Germany	3–6	12.5:1 student-teacher ratio. Teachers had completed vocational training. Limited curricular focus.	Family care by parents or grandparents	No average treatment effect on school readiness at age 6	Children from more disadvantaged backgrounds have substantially lower enrollment. Positive effects for children with low enrollment, and negative effects for children with high enrollment.	90s	Cornelissen, Dustmann, Raute and Schönberg (2018).
Georgia and Oklahoma, USA		10:1 student teacher ratio. Lead teachers had bachelor's degree in early education. Curricular focus.	Other care (formal and informal) and parental care	Positive effects on math test scores at the end of preschool, fourth grade and eighth grade. No effects on language in eighth grade.	Beneficial effects only for children of disadvantaged families	90s	Cascio and Schanzenbach (2013), Gormley and Gayer (2005), Wong et al. (2008), Fitzpatrick (2008)
Effects of center-based childcare for children ages 0-3							
Quebec, Canada	0-4	A mix of center-based care and licensed home-based providers. Only 60 percent of centers met training and education standards for teachers	Maternal care	Negative effects on children's health and noncognitive test scores. Negative long-term effects on adult health and life satisfaction, and higher crime rates	Stronger detrimental effects on non-cognitive tests among those already scoring low on noncognitive tests	2000s	Baker, Gruber, and Milligan (2008, 2019)
Bologna, Italy	0-2	Quality regulation concerning adult-to-child ratio: 1:4 at age 0 and 1:6 at age 1-2.	Parental or other family care	Negative effects on children's IQ and personality traits.	Large and significant IQ loss for children of more affluent households, while small and insignificant for children of less affluent households. Same pattern for personality traits.	2000s	Fort et al. (2020)
Denmark	3	A mix of center-based care and licensed home-based providers.	Parental care	No effects on non-cognitive skills at age seven, neither of center-based care, nor of	No differential effects across family background of center-based care. Home based care seemed to have detrimental	1990s	Datta Gupta, Simonsen (2010)

		Requirement of college-educated preschool teacher in center-based care		licensed home-based providers	effects for children of mothers with low levels of education		
Germany	1-2	Strict quality regulations concerning group size, the staff-child ratio, and staff qualifications	Maternal care	No average treatment effect on language and noncognitive and motor skills. Positive effects on motor skills	Children from more disadvantaged backgrounds have lower enrollment. Children with lower (higher) enrollment experience larger positive effects on noncognitive (motor) skills.	2000s	Felfe and Lalive (2018)
Norway	1-2	Strict quality regulations concerning group size, the staff-child ratio, and staff qualifications	Informal care	Positive effects on language and mathematics assessments at age 6–7	Stronger effects for children of disadvantaged families	2000s	Drange and Havnes (2019)
Japan	1-3	Teachers must have 2 years of relevant post-secondary education. Teacher-child ratio is 1:6.	Informal care and parental care	Positive effects on language at age 3. No effects on noncognitive skills.	Positive effects on noncognitive skills among children of mothers with low levels of education	2000s	Yamaguchi, Asai Kambayashi 2015
Spain	3	Teachers must have relevant college degree. Class size 20.	Family care by mother or grandmother	Positive effects of preschool enrollment on reading skills at age 15 and on grade retentions	Effects only for disadvantaged children	1990s	Felfe, Nollenberger and Rodríguez-Planas (2014)

In all cases, we will discuss heterogeneity of impacts across family backgrounds. Several studies suggest that gaps in skill development across family backgrounds can partially be explained by the degree to which parents are, or are not, able to create a home environment that stimulates learning and development (Guryan et al. 2008, Kalil et al. 2012). Thus, universal preschool may be particularly beneficial for children of parents with limited education, who tend to provide fewer learning opportunities at home than highly educated parents.

Effects of universal preschool ages 3-6. Several studies demonstrate that participation in universal preschool for children ages 3 to 6 can have positive effects on child development, especially in the case of disadvantaged children. These studies vary in the extent to which they investigate short- and long-term outcomes. Of course, it is important to determine whether the effects of preschool persist into adulthood. The drawback, though, is that long-term studies are bound to investigate effects of preschool systems in contexts that are decades old. Thus, studies looking at short- and medium-term outcomes are also important for understanding potential impacts of preschool expansions in today's societies. We begin our review with short- and medium-term studies, and then turn to a study that looks at long-term outcomes.

One of the first studies of universal preschool participation is from Argentina, where the public school system provided three years of preschool education for children ages 3 through 5. During the 1990s, the government of Argentina greatly improved access to preschool classrooms, increasing enrollment rates from 49 percent in 1991 to 64 percent in 2001, with large variation across regions. Using this variation in a difference-in-differences framework, Berlinski et al. (2009) demonstrate that one year of preschool increased average third-grade test scores by .23 of a standard deviation (Table 4). Effects were significantly larger for children living in more disadvantaged municipalities. The authors note that the preschool had a curriculum targeting communication skills, social and emotional skills, and mathematics. Unfortunately, however, they provide no information on class size, teacher qualifications or counterfactual conditions. This makes it difficult to apply lessons from the study to other contexts.

A study of a preschool expansion in Uruguay finds that beneficial effects of preschool on achievement can persist for a number of years. In the 1990s, the government of Uruguay initiated a rapid expansion of universal preschool for children ages 4 to 5, in response to major concerns about grade retention and early drop-out in elementary school (Berlinski et al. 2008). Most of the preschool classrooms were built and organized as a part of the public elementary schools. The

reform created variation in preschool access across space and time. Utilizing this variation in an instrumental variable framework and sibling fixed effects, Berlinski et al. (2008) demonstrate that preschool enrollment reduced both grade retention and drop-out. By age 15, treated children had completed 0.79 more years of education than their non-treated siblings and were 27 percentage points more likely to be still enrolled in school. These effects were driven by children of mothers with lower-than-average education. The authors demonstrate that the program was cost-effective. Even under the most conservative scenarios, the rate of return to investing in preschools was as high as 14%. As with the Argentinian study, the Berlinski et al. (2008) study provides very limited information about the nature of preschool classrooms and counterfactual conditions, which makes it hard to apply the findings to other contexts.

Cascio and Schanzenbach (2013) also find beneficial short- and medium-run effects of universal childcare in the U.S. In the 1990s, the states of Oklahoma and Georgia introduced universal preschool programs that had an academic focus and included comprehensive learning standards. The student-teacher ratio was 10:1, and the lead teachers were required to have bachelor's degrees in early education. The counterfactual was a mix of parental care and other care, both informal and formal. Cascio and Schanzenbach (2013) examine these preschool expansions in a difference-in-differences approach using other states as controls. They investigate impacts on fourth- and eighth-grade test scores in reading and mathematics for low- and high-income children. The evidence suggests no beneficial effects of program participation for children from high-income households, but substantial effects on the fourth-grade test scores of low-income children. For language, the effect on low-income children fades out by eighth grade, but for math it remains large and significant. The magnitude of the estimate suggests that preschool attendance increased the eighth-grade math scores of children from low-income households by almost a third of a standard deviation.

The evidence in Cascio and Schanzenbach (2013) is consistent with earlier studies that looked at the effects of each of the state programs in isolation on shorter term outcomes (Fitzpatrick 2008, Gormley & Gayer 2005, Wong et al. 2008). Using a regression discontinuity approach, Gormley and Gayer (2005) and Wong et al. (2008) find positive effects of the Oklahoma program on child development, measured at the end of preschool. And in an investigation of the Georgia program, using a cross-state difference-in-differences approach,

Fitzpatrick (2008) finds positive effects on fourth-grade test scores and the probability of being at grade level. All of these studies find larger treatment impacts for disadvantaged children.

Studies from Norway suggest that for disadvantaged children, the benefits of universal preschool can persist into adulthood (Havnes & Mogstad 2011, 2015). In the 1970s, the Norwegian government initiated a reform that substantially expanded publicly provided universal preschool. Preschools were required to have at least one college-educated teacher per 16 children aged 3–6, who typically worked closely with two assistants. The program was established in response to a need for high-quality childcare as mothers entered the labor market. The preschools were supposed to build academic and social and emotional skills, but the curriculum was vague, and classrooms were structured to promote socialization and free play, with limited planning and structure for learning activities. This "whole child" approach (see Section 5) builds on a belief that the best learning experiences for preschool-age children involve free play and activities that allow the children to take the initiative, without systematic attempts by teachers to relate those experiences to academic skills.

The reform led to considerable variation in preschool access for children 3–6 years old, both across time and between municipalities. Utilizing this variation in a difference-in-differences approach, Havnes and Mogstad (2011) demonstrate that attending preschool between the ages of 3 and 6 led to a substantial increase in completed schooling, labor market attachment, and earnings. An additional 17,500 preschool slots produced about 6,200 additional years of education, with effects concentrated more among children from lower-income families. In contrast, for children from high-income families, childcare participation led to a reduction in adult earnings (Havnes & Mogstad 2015). Additionally, the authors provide evidence suggesting that the new subsidized preschool crowds out informal care arrangements, with almost no net increase in maternal labor supply (Havnes & Mogstad 2011). Thus, the beneficial effects of preschool appear to reflect a shift that is largely from informal rather than parental care.

The studies described above provide limited insight into how families select into universal preschool. A recent study from Germany addresses this issue by carefully investigating heterogeneity in the effects of preschool enrollment (Cornelissen et al. 2018). It applies the marginal treatment effects framework, introduced by Björklund and Moffitt (1987) and extended in a series of papers by Heckman and coauthors (see e.g. Heckman & Vytlacil 2005, Heckman & Vytlacil 1999), relating the heterogeneity in the treatment effect to heterogeneity in the

propensity for preschool enrollment. It uses a reform during the 1990s that entitled every child in Germany to heavily subsidized preschool from age 3 to 6, which created variation in the availability of preschool slots both geographically and across cohorts. The authors investigate impacts on school readiness at age 6, which is a summary measure of reading, math and social and emotional skills. The analysis suggests a pattern of reverse selection on gains from preschool. Specifically, the marginal treatment effects framework demonstrates that fewer children from disadvantaged backgrounds attended preschool compared with children from advantaged backgrounds. Moreover, among those who were less likely to attend preschool, there was a large positive effect of preschool enrollment on school readiness. In contrast, effects were negative among those who were more likely to attend preschool.

To generalize from these findings, it is important to be familiar with key features of the German preschool program and the counterfactual. The German program was based on a whole-child approach, like the Norwegian program. But it targeted academic skills as well as social and emotional skills, through learning that was mostly informal and playful and carried out in the context of day-to-day social interactions between children and teachers. Moreover, the program was part-time – just four hours every morning. The student-teacher ratio was 12.5:1, and teachers were required to have completed vocational training relevant to their caregiving and teaching responsibilities. The counterfactual mode of care was family care, either by parents or grandparents. In this case, therefore, the effects of preschool reflecting a shift mostly from family care to part-time, play-based formal care.

Effects of childcare for children ages 0-3. While evidence for older preschool-aged children suggests that participation in universal preschool is beneficial for disadvantaged children, the evidence on universal childcare for younger children is more mixed. In fact, some studies suggest that it has only limited effects on child development, or that it may even be detrimental.

Baker et al. (2008) looked at the introduction of highly subsidized, universal childcare for children ages 0-4 in the Canadian province of Quebec in the late 1990s. This program was a mix of center-based care and care from licensed home-based providers. It imposed some quality requirements, such as training and education standards for teachers, but in the rush to expand the number of slots, only 60 percent of center teachers actually met these requirements. Baker et al. (2008) estimate difference-in-differences models comparing the outcomes of children in Quebec

and the rest of Canada before and after this reform. They find negative effects of childcare enrollment on aggression, motor skills, social skills and illness among children 0–4 years old. They also find large positive effects on maternal labor supply, which suggests that the counterfactual was maternal care. In a later follow-up study, they show that negative effects persist into adulthood, producing worse health outcomes, lower life satisfaction and higher crime rates (Baker et al. 2019).

Fort et al. (2020) investigates effects of enrollment in the universal childcare system of Bologna in Italy at ages 0-2. This was a center-based care program with an adult-to-child ratio of 1:4 at age 0 and 1:6 at age 1-2. Exploiting over-subscription and admission thresholds for identification, the authors demonstrate that enrollment caused large and significant IQ loss for children of more affluent households. For children of less affluent households the effects were small and insignificant. The same pattern emerged when investigating effects on personality traits.

Denmark offers universal childcare, which can be either licensed home-based care or center-based care. Municipalities are free to decide how the two different types of care are distributed, but are required to ensure that a sufficient number of slots are available for each age to meet local needs. Exploiting municipality-level variation in access to childcare for three-year-old children, Gupta and Simonsen (2010) use an instrumental variable approach to estimate effects on non-cognitive skills at age seven. They use the Strength and Difficulties Questionnaire index (SDQ) derived from parental survey data as a measure of non-cognitive skills. The SDQ is a commonly used behavioral measure in the child development literature that uses concrete questions (e.g., “does the child have at least one good friend” or “does the child often lose temper”). The authors find that enrollment in center-based care rather than being in parental care had no impact on children’s non-cognitive development (as reported by the parents), no matter the family background. Licensed home-based care providers, however, seemed to have detrimental effects on children of mothers with low levels of education.

Some studies suggest that center-based childcare has possible benefits for toddlers. The Norwegian public preschool system has long had a shortage of slots, especially for children ages 1-2, which prompted the Oslo municipality to introduce an enrollment lottery. Drange and Havnes (2019) utilize this lottery to study the effects of preschool enrollment on children ages 1-2. They find that preschool participation has positive but small impacts on school readiness (0.16

standard deviations higher scores on language and 0.11 standard deviations on mathematics assessments at age 6–7). The effects were considerably stronger for children from more disadvantaged backgrounds; children from low-income families improved their performance in language and math by .26 standard deviations, compared with modest or no effects for children from high-income families. The authors find limited effects of the lottery on maternal labor supply, suggesting that informal care is the relevant counterfactual. The beneficial effects for low-income children must be interpreted in the context of strict quality regulations concerning group size, staff-child ratios, and staff qualifications.

Center-based childcare also seems to be beneficial for young children in Japan, especially those from disadvantaged backgrounds. In the early 2000s, the Japanese government initiated a center-based childcare reform, resulting in large variation in childcare expansion across different regions. In an instrumental variables approach, Yamaguchi et al. (2018) use this variation to investigate effects of childcare at age 1-3 on developmental outcomes at age 2-3. They find that the treatment leads to 0.71 standard deviations higher scores on parent-reported language skills at age 2, but no significant average effects on parent-reported inattention and hyperactivity at age 3. There are no differential effects on language across family backgrounds, but childcare enrollment reduces inattention and hyperactivity among children of mothers with low levels of education, while it has no significant effects on children of highly educated mothers. Investigating mechanisms, the authors find that enrollment leads to a more positive home environment. It also improves the parenting quality of mothers with little education. The counterfactual in this study was a mix of parental care and informal care. The center-based childcare was strictly regulated; teachers were required to have at least two years of relevant college education, and the teacher-child ratio was 1:6.

A study from Spain suggests that positive effects of childcare at young ages can be long-lasting (Felfe et al. 2015). In the early 1990s, the Spanish government launched a sizeable expansion of publicly subsidized childcare for 3-year-olds, creating variation in childcare access across states. Felfe et al. (2015) take advantage of this variation in a difference-in-difference approach and find positive effects of preschool enrollment on reading skills at age 15 and on grade retention during primary school. Specifically, their intent-to-treat estimates suggest that reading test scores improved by 0.15 standard deviations, and the prevalence of grade retention in primary school declined by 2.5 percentage points. It is hard to interpret these findings, as they do

not capture the magnitude of the childcare expansion. The authors do not attempt to calculate treatment-on-the-treated. In a heterogeneity analysis, the authors find that the treatment effects are driven by disadvantaged children. These effects must be interpreted in a context in which the counterfactual care mode is family care by the child's mother or grandmother. The maximum class size was 20 children, and teachers had a relevant college education.

Much like Cornelissen et al. (2018), Felfe and Lalive (2018) investigate how families select into universal childcare. Beginning in 2005, Germany increased funding to school districts for expanding center-based early childhood care, leading to initial large variation in coverage rates across school districts. Felfe and Lalive (2018) first demonstrate that the counterfactual to center-cased care is maternal care. They then use the variation in coverage rates to study effects of preschool in a marginal treatment effects framework. Their analyses, like those of Cornelissen et al. (2018), suggest that there are no significant average treatment effects on language and noncognitive skills, but positive effects on motor skills. The marginal effects framework demonstrates that children from more disadvantaged backgrounds have lower rates of enrollment. Moreover, children with a low likelihood of enrollment experience positive effects on noncognitive skills, whereas children with a high likelihood of enrollment experience positive effects on motor skills. These beneficial effects of switching from maternal care to center-based care must be interpreted in the context of a childcare system with strict quality regulations concerning group size, staff-child ratios, and staff qualifications.

4.3 The neglected distinction between average and marginal returns

Returns at the margin are central to economic analysis, as is the contrast between average returns and marginal returns. It is thus surprising that so few of the empirical papers discussed in Sections 4.1 and 4.2 distinguish between marginal and average returns of public investments in preschool programs. [9]

The distinction between marginal and average returns matters when assessing the costs and benefits of publicly subsidized preschool. On the cost side, a key problem is that it can be

[9] Notable exceptions include Cornelissen et al. (2018) and Felfe and Lalive (2018). The marginal treatment effects approach adopted in these studies allow the authors to not only uncover treatment heterogeneity by observed characteristics, but also in terms of unobservables. While useful, these estimates do not permit a standard policy evaluation of marginal benefits and costs. The marginal treatment effects literature focuses on gross benefits of policies and recovers neither the marginal costs nor the perceived surplus associated with the programs being evaluated.

very inefficient to provide subsidies to infra-marginal families, i.e., families who would have used preschools in the absence of the subsidy. A good example is the heavily subsidized, universal preschool system in the Nordic countries. Despite the fact that most children in these countries have long been enrolled in a universal preschool center, the subsidy rates have increased over time in an attempt to raise participation rates even further. For the many infra-marginal families, the subsidy increase will only have income effects, making it identical to an equal-value cash transfer. Such transfers are costly because of the deadweight cost of taxation.

Infra-marginality problems plague almost all preschool subsidy policies because subsidies create a wedge between marginal and average returns, even if the benefits of attendance would be the same for all children. The basic insight from public finance is that subsidies should be targeted to individuals who would not engage in the desirable activity in the absence of the subsidy. But we don't want to provide subsidies to individuals who would have done so whether or not a subsidy was available.

The distinction between marginal and average returns is also important for an assessment of the benefit side of universal preschool programs. As discussed in Section 4.2, the effects of preschool show systematic heterogeneity across families with different incomes. Universal as well as some targeted programs have been shown to have the potential to produce beneficial outcomes for children from low-income families. By comparison, studies tend to report zero and sometimes negative impacts of universal programs on children from middle- or upper-class families. A possible explanation is that the quality of the counterfactual mode of care (i.e., the type of care the children would be exposed to absent the reform) is likely to differ systematically across the family income distribution. For example, the counterfactual mode of care for children in low-income families is often informal care arrangements (including relatives, unlicensed care givers, and other irregular care givers such as friends and neighbors). A shift from parental care to subsidized formal childcare may affect children differently than a shift from informal care, which is likely to be of inferior quality.

All of this may have important implications for the policy debate over universal preschool programs, as it appears likely that the average returns to providing subsidized universal preschool are unlikely to exceed the costs, even if the marginal returns to targeted subsidies to children from low-income families are large.

4.4 Discussion and future research directions

Universal preschool. The papers reviewed above provide somewhat mixed evidence for the effects of universal preschool. Explaining why the different studies yield different results is not an easy task, because the studies are from different countries and components of the preschool programs (e.g., the curriculum, teacher education, and child-staff ratios), the counterfactual to preschool (e.g., parent care, informal care, or other center-based care), the children's age and the populations differ across the studies. Moreover, many of the above-reviewed studies provide limited information on program content and counterfactuals. Another concern is that, for identifying causal relationships, most of these studies utilize policy reforms that led to rapid preschool expansions in certain geographic locations, using a difference-in-differences approach. The quality of care during and immediately after an expansion is not necessarily representative of the quality of care in centers with established routines and experienced caregivers.

Taken together, however, these studies provide quite convincing evidence that in the case of older preschoolers, preschool participation has the potential to be beneficial for development and to produce lasting effects for children from disadvantaged families. It is not clear why advantaged children often seem not to benefit from universal childcare. One reason may be that universal childcare teaches skills that children from more advantaged backgrounds would learn at home anyway. Another reason, discussed in Havnes and Mogstad (2015), is that universal childcare is essentially a subsidy for choosing care of a given quality. This could lead some mothers who would otherwise use care of lower quality to move to subsidized care, while some mothers who would otherwise choose higher-quality care instead opt for lower-quality subsidized care. If children from advantaged households tend to be in the latter category, while children from disadvantaged households are part of the first group, this might explain the differential effects.

More research is needed to gain a better understanding of the impacts of universal childcare and to provide policy recommendations. Specifically, we need:

- Studies that can help us better understand the mechanisms through which disadvantaged children benefit from preschool. As discussed above, positive effects of preschool can operate through many different mechanisms. In addition to the effects of the preschool itself, there may be changes in parental investment, behaviors and stress, peer effects, and school

investments. Unfortunately, the treatment-effect estimates outlined above provide only limited insights into possible mechanisms.

- Studies that can help us understand how to make universal preschool more effective. Specifically, what are the most important skills that universal preschool programs should focus on at various ages in order to produce lasting effects on child development? We will return to this question in Section 5.
- Studies that distinguish marginal and average returns of public investments, a topic that we discuss in more detail in Section 4.3.

Targeted programs. The literature on universal preschool programs consistently points to positive short-term impacts for disadvantaged children ages 3-5 and, in the case of the few studies that are able to estimate them, longer-run benefits as well. At the same time, the literature on specific programs targeting disadvantaged children is mixed at best, with some early studies pointing to large and long-lasting benefits, but most recent studies showing null to negative impacts several years after the end of the programs. To understand this apparent paradox, we need to recognize the challenges of drawing lessons from social experiments when control-group participants can access other, close-substitute programs or a higher-quality home environment (Heckman et al., 2020; Kline and Walters, 2018). In contrast to early Head Start studies, recent program evaluations like the Head Start Impact Study have shown modest initial effects that quickly faded out, which is consistent with the idea that the growth of center-based substitutes has reduced the apparent impacts of the program. Restricting the counterfactual to home-based care increases the apparent advantages of enrollment in the Head Start program (Kline & Walters 2016).

Although counterfactual conditions are not well described in most of the universal preschool evaluations, fragmentary evidence indicates that much of the counterfactual care is provided by families, which can lower the bar for the universal program to show benefits, particularly in the case of children from low-income households. A tentative conclusion is that interventions that draw disadvantaged children into the formal care sector from parental care are likely have a positive payoff. A corollary is that it is hazardous to draw conclusions about the potential value of general expansions of preschool enrollment among all disadvantaged children based on evidence from a targeted program that was implemented in an environment in which

children were drawn from center-based care alternatives. General expansions may well generate positive returns even if individual programs do not.

Which targeted programs are best? Evidence on targeted programs, virtually all of which are based in the U.S., is confusing. Perry and Abecedarian seem to show most clearly that high-quality (and high-cost) programs designed and run by researchers, and in the context of low-quality counterfactual conditions, can transform the lives of many of the enrolled children. They improve some combination of cognitive and noncognitive skills across childhood and adolescence, and even well into adulthood. Their costs may be high, but the value of the benefits they yield, specifically increased labor-market productivity and reduced crime, far exceeds those costs. This may also hold true for children who participated in Head Start during the first two decades of that program, which is particularly noteworthy because, in contrast to Perry and Abecedarian, Head Start is a fully scaled-up, government-run program.

What about more recent cohorts of children, for whom counterfactual conditions and safety-net policy environments are far more favorable than they were 50 years ago? The generally null elementary-school impacts in recent high-quality RCT studies like the Tennessee pre-K and National Head Start Impact Study are certainly not promising. To assess the relative costs and benefits, however, we need to consider the possibility of "sleeper effects." In the more promising results from the Boston pre-K study conducted by Gray-Lobe et al. (2021), longer-run impacts were apparent in high school and beyond despite the absence of impacts on outcomes measured earlier. Longer-run follow-ups for the TN VPK and NHSIS (as well as a longer-run follow-up in the Boston study) are an urgent research priority.

It is difficult to imagine that the longer-term Boston impacts were not fueled by some kind of short-term improvements, although it is quite possible that the "secret sauce" isn't captured by the administrative data available to evaluators. Attempts are currently underway to develop prediction models that can translate short-run impacts into longer-run projections (Athey et al. 2019, Bailey et al. 2020, García et al. 2020), but mere prediction is a poor substitute for a conceptual understanding and empirical confirmation of the processes by which recent cohorts of disadvantaged children may achieve upward mobility through targeted early childhood education programs.

5. Improving preschool effectiveness

Many public investments in preschool are motivated by research demonstrating that, for disadvantaged children, preschool programs can improve children's school achievement and even, in some cases, boost adult labor market productivity. But, as noted in Section 4, that evidence is mixed, especially in the case of more recent evaluations of scaled-up targeted programs. Several studies of programs targeting low-income children show that initial learning gains quickly fade out. Moreover, some studies, of both targeted and universal programs, suggest that preschool can have detrimental consequences for children's behavioral development. Unfortunately, this body of literature offers few insights into *why* the effects of preschool differ across studies because they provide few details on the structure of the preschool programs, the nature of teacher/child interactions and the counterfactual care arrangements of non-enrolled children.

In this section we review studies investigating how child outcomes are affected by various features of preschool programs – teacher education, classroom size, quality of teacher-student interactions, preschool curriculum, and the like. Not surprisingly, these studies are plagued by the challenge of manipulating one factor in the production technology while holding all others constant. Other investment factors often adjust to changes, as when an improved preschool curriculum may lead parents to invest less time and money in their children at home because they believe that they are learning so much in preschool. Accordingly, results from the studies discussed in this section should be understood as policy effects that include the adjustment effects of other investments rather than as estimates of production technologies.

Substantial variation in child learning across centers, even within the same preschool system, highlights the importance of this kind of research. In a random-effects analysis, Walters (2015) demonstrates that the extent to which children's cognitive development benefited from Head Start enrollment varied substantially across centers. A study by Rege et al. (2018) suggests that this is the case even in the universal preschool program in Norway, which ranks first among the OECD countries in public spending on preschool (Engel et al. 2015). In random-effects analyses, the authors demonstrate substantial cross-sectional variation in covariate-adjusted school readiness skills across centers. Differences in school readiness skills in centers between the 90th and 10th percentile of the center effect distribution amounted to over one half (0.55) of a standard deviation.

5.1 Defining preschool quality

Well established theories of early childhood education and care suggest that frequent, consistent, sensitive and nurturing interactions with caregivers, along with engagement in age-appropriate activities and play, are important for the children's socioemotional and cognitive learning (Burchinal 2018). Children's early learning is most efficient when they engage in play and discovery together with other children, as teachers intentionally introduce planned games and activities scaffolding children's exploration of core themes (Hirsh-Pasek et al. 2009). These theories stress that children need to be active and engaged, interact with others, and experience the activities as meaningful (Hirsh-Pasek et al. 2009).

The early childhood education and care literature stresses that playful activities should stimulate children's development of language, mathematics and socioemotional skills. As discussed in Section 3, these are often referred to as school readiness skills, as numerous studies have documented that they are strong predictors of success at the start of formal schooling and of longer-run academic achievement and social adjustment (e.g. Duncan et al. 2007). In that section we also discussed the concern raised by Bailey et al. (2020) that targeting these kinds of skills in preschool and home environment interactions during the early childhood period may be insufficient to generate persistent impacts, because some of these skills are soon mastered by children in the comparison groups. Traditional school readiness skills such as literacy and numeracy are all manipulable through preschool programs, but many of these are learned quickly by all children once they enter formal schooling.

Preschool quality is typically defined along two dimensions —structure and process. *Structural quality* comprises features of the childcare setting itself and is often measured by teacher education, teacher experience, child-staff ratios and class size. *Process quality* reflects the direct experiences of children in the preschool setting and includes factors such as the sensitivity and responsiveness of caregivers to children's individual needs, the more general nature of interactions with teachers and peers, and the pedagogical approaches and materials available for learning, which constitute important elements of curricula. As appealing as these ideas are, we will see in Section 5.2 that improving many elements of structural quality typically has a modest or no effect on children's development. Moreover, few attempts to improve the quality of teacher/pupil classroom interactions translate into gains in children's school readiness. However,

improving process quality by implementing skill-focused curricula has been shown to improve school readiness in many cases.

5.2 Structural investments

A large nonexperimental literature in developmental psychology and early childhood education investigates how structural factors correlate with children's learning and development (e.g. Early et al. 2006, Mashburn et al. 2008). Below, we look at studies from economics that use various strategies for addressing selection. As Table 5 demonstrates, findings from these studies parallel counterpart evidence from elementary schools (see review in Jackson et al. (2014)) in showing the limited potential for policy to improve child development by merely investing in structural characteristics.

Although the studies we review address selection problems, they have methodological limitations. First, and as discussed above, they do not help us identify parameters in the education production function by investigating the impact of one structural parameter, holding others constant. At best, estimates reflect the total treatment impact of investing in a given structural measure, which can include secondary adjustments in other factors. Second, even a policy interpretation can be problematic because the studies utilize random allocation of students to diverse preschools with different structural parameters. Estimates of factor productivity are biased if their structural parameters are associated with unobservable structural parameters that matter for child development. As far as we know, there are no randomized experiments investigating effects of structural parameters in preschool.

In one of the first studies investigating potential impacts on child development of improving structural quality in preschool, Blau (1999) uses data from the National Longitudinal Survey of Youth (NLSY) and relies on mother fixed effects for identification. He finds that increasing child-teacher ratios, improving teacher education or reducing class size have no treatment impact on children's skills as measured one year later. This could be because these aspects of the classroom do indeed have no effects, or because possible effects are canceled out by other aspects that change as these parameters change, but are not observed, for example parent behavior (Table 5).

Table 5: Evidence on impacts of structural investments on children's learning and development.

Where	Class size	Teacher-student ratio	Share of male staff	Director experience	Teacher experience	Teacher education/ training	Longer operating hours	Studies
USA	No effect	No effect	n.a.	n.a.	n.a.	No effect	n.a.	Blau (1999)
USA	n.a.	No effect	n.a.	No effect	No effect	No effect	n.a.	Currie and Neidell (2007)
USA	No effect	No effect	n.a.	No effect	n.a.	No effect	n.a.	Walters (2015)
Denmark	n.a.	Small positive	Small positive	n.a.	n.a.	Small positive	n.a.	Bauchmüller et al. (2014)
Denmark	n.a.	n.a.	Small positive	n.a.	n.a.	n.a.	n.a.	Gørtz et al. (2018)
Norway	n.a.	No effect	Small positive	n.a.	No effect	No effect	n.a.	Drange and Rønning, 2017
Germany	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	Positive	Felfe and Zierow (2018)

Currie and Neidell (2007) find with NLSY data that former Head Start children have higher reading and vocabulary scores at ages 5-7 if they attended a program where Head Start spending was higher at age four. Interestingly, however, they find little evidence that aspects of structural quality, such as teacher-student ratios, teacher education, or teacher and director experience are associated with beneficial effects. This study was correlational but included an extensive set of state- and county-level controls. In a more rigorous empirical design, Walters (2015) finds similar results using random assignment data from the National Head Start Impact Study. Specifically, he shows that observed inputs can explain little variation across centers in the short-run treatment effects on cognitive development. Treatment effects are somewhat larger in centers with home visits and full-day service, although structural quality inputs, such as teacher qualification, child-staff ratios and director experience, explain none of the variation in treatment effects.

There are several recent studies on structural quality from Scandinavia. Drange and Rønning (2017) use unique data on preschool application and admission lotteries in Norway for identification. They find that enrollment in preschools with higher levels of teacher education, more teacher experience, or higher teacher-student ratios fail to predict higher achievement scores in math or language in first grade. A high percentage of male staff in preschool is predictive of language achievements in early school years, although this may reflect unobserved characteristics of preschool centers with larger numbers of male staff.

Bauchmüller et al. (2014) use Danish administrative register data to investigate how quality indicators in preschool predict grades at the end of primary school. Controlling for a rich set of variables, they find that teacher education and child-teacher ratios are significant predictors of grade outcomes, but the estimates are small in magnitude. Consistent with Drange and Rønning (2017), Bauchmüller et al. (2014) find correlational evidence suggesting that children benefit from a higher percentage of male staff. This relationship is also supported in another study from Denmark that utilizes a more careful empirical design relying on within-center differences in gender composition across time (Gørtz et al. 2018).

We have limited evidence on how preschool operating hours affect child development. One exception is a study by Felfe and Zierow (2018), which investigates the consequences of a substantial increase of full-day slots at the expense of half-day slots in the German preschool system. The expansion depended on centers' readiness to increase their supply of full-day slots

and their administrative capabilities. This resulted in substantial variation in the available full-day slots across centers and time. Using this variation for identification, the authors demonstrate that more hours in preschool had a negative effect on children's socioemotional development.

5.3 Investments in process quality

Several scholars have investigated how aspects of process quality are associated with student learning. An often-used measure is the Classroom Assessment Scoring System (CLASS) (Pianta et al. 2008), which measures three aspects of process quality: instructional support, which captures the extent to which teachers provide and scaffold in-depth learning; emotional support, which captures teacher's awareness of and response to the academic, emotional and social needs of their students; and classroom organization, which captures routines, teaching methods, and approaches to promoting student engagement and learning, in addition to behavioral management. Howes et al. (2008) find some positive associations between higher CLASS scores and gains in cognitive and noncognitive skills, but the magnitudes are small and, for the most part, not statistically significant. Other studies using CLASS or other measures of process quality, such as the Early Childhood Environment Rating Scale (ECERS; Harms et al. 1998), find similarly modest and often insignificant associations between process quality and child development (e.g. Keys et al. 2013, Mashburn et al. 2008, Weiland et al. 2013).

Estimates in these correlational studies of CLASS and ECERS may be biased because of selection, omitted variables, or even reversed causality. A recent study by Araujo et al. (2016) addressed some of these concerns. Based on video recordings from kindergarten classrooms, the study measures teacher quality using CLASS. Although elements of CLASS were not randomly assigned across classrooms, children were. The study demonstrates that CLASS-based teaching quality, measured in year t, predicts kindergarten learning outcomes in year t+1, both for academic and socioemotional skills. Specifically, the study finds that increases in classroom quality by one standard deviation result in 0.11, 0.11, and 0.07 standard deviation higher test scores in language, math, and executive functioning, respectively, more than twice the magnitude typically found in correlational studies.

The CLASS data in Araujo et al. (2016) suggest that students' learning experiences vary substantially across classrooms, depending on teaching quality. If we want to consider CLASS as a practical and manipulable policy tool, however, these estimates are too small to be of much

practical importance. Based on the estimates by Araujo et al. (2016), we would need to develop classroom quality interventions that change the CLASS score by two full standard deviations to improve student outcomes by about .2 standard deviations. Moreover, the treatment-induced CLASS variation in the Araujo et al. (2016) study may matter for reasons other than variation in the CLASS measure itself. Specifically, the CLASS measure may be correlated with other unmeasured teacher attributes, which could themselves affect child learning. Experimental intervention studies targeting process quality as measured in CLASS through teacher professional development training and coaching often document large increases in the CLASS measure, but no corresponding impacts on child outcomes (Pianta et al. 2017, Yoshikawa et al. 2015).

Altogether, the literature on CLASS and other existing preschool measures of process quality, such as the ECERS, suggests that these measures may not be very useful policy tools for improving preschool classrooms in ways that facilitate children's learning and development. In other words, children's outcomes appear to be more strongly related to aspects of process quality that are not well captured by CLASS or other existing measures of process quality (Burchinal 2018). Other avenues for improvement include curricula and professional development. We will review evidence on these in Sections 5.4 and 5.5.

5.4 Curriculum Investments[10]

Preschool programs differ substantially in the kinds of curricula they employ. Curricula set goals for the knowledge and skills that children should acquire in school, and support educators' plans for providing the day-to-day learning experiences needed to cultivate those skills. In thinking about curricular choices, it is useful to distinguish between so-called *whole-child* curricula, which organize classroom activities across the day and year, and more targeted, *skill-specific* curricula, which are often delivered as supplements to activities in whole-child curricula.

Whole-child curricula emphasize child-centered active learning, facilitated by a strategic arrangement of the classroom environment and teacher guidance (DeVries & Kohlberg 1987, Piaget 1976, Weikart & Schweinhart 1987). They are based on Vygotsky's concept of the "Zone of Proximal Development," which, in the case of preschool learning, is a function of the distance

[10] Parts of this section draws from Jenkins and Duncan (2017).

between a child's actual developmental level and the level of development that might be attained through problem-solving under guidance from, this case, preschool teachers (Vygotsky 1980). If this approach is done well, each child engages with components of the classroom environment independently, and the teacher's task is to scaffold learning with just the right amount of input – not so little that the child fails to learn, but not so much that the task becomes teacher-driven. Head Start centers and many other preschool programs in the U.S. are required to use a whole-child curriculum.

The social pedagogical tradition of preschools in Scandinavia and Germany aligns with the whole-child approach. This tradition emphasizes free play in mixed-age groups, humanistic values and the affective qualities of the teacher-child relationship. But there is no detailed and structured curriculum; instead, teachers facilitate learning through spontaneous engagement, interaction and play, and through crafts projects and story time (Bennett & Tayler 2006, Engel et al. 2015).

Ideally, the sequence of teacher inputs tailored to the individual needs and interests of the children in a whole-child approach will promote the cumulative development of cognitive and socioemotional skills over the course of the preschool years. However, as might be expected, it takes considerable skill on the part of teachers to implement a whole-child curriculum effectively. Another concern with this approach is that it gives preschool centers a great deal of freedom with respect to pedagogical content, which can give rise to large differences in learning across centers (Bennett & Tayler 2006, Engel et al. 2015, Rege et al. 2018).

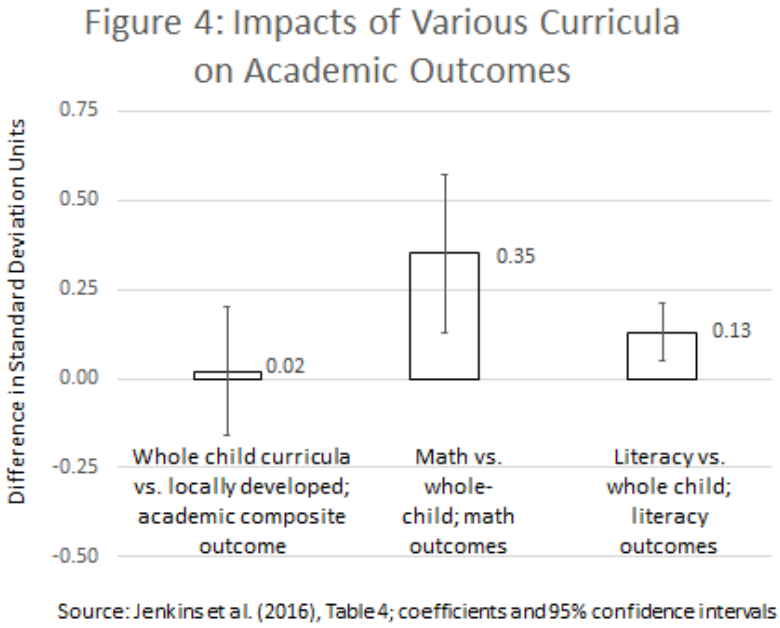
In contrast to the whole-child approach, proponents of skill-specific curricula argue that preschool children benefit most from sequenced, explicit instruction that focuses on specific academic or socioemotional skills, but is also provided in the context of play and exploration. Teachers are supposed to work intentionally and systematically to stimulate children's skill development using play-based approaches in both small and large group contexts. This approach is often referred to as the "school readiness tradition" because it takes an instrumental approach to strengthening children's skills in specific areas of development that have been linked to success at the start of formal schooling and to longer-run academic achievement and social adjustment (Bennett & Tayler 2006). There is ongoing debate about the effectiveness of the whole-child approach compared with more structured and academically-focused curricula (Montie et al. 2006, Weiland & Yoshikawa 2013) (Montie, Xiang, & Schweinhart, 2006; Weiland & Yoshikawa,

2013).

Preschool Curriculum Evaluation Research (PCER) Initiative Study. A comprehensive and rigorous assessment of curricula is the Preschool Curriculum Evaluation Research (PCER) Initiative Study. It comprised RCT-based evaluations of 14 early childhood education curricula evaluated at 18 different sites. Nearly 3,000 children in U.S. public preschools, Head Start programs, and private childcare were included in the evaluations. The centers that were part of the PCER study primarily served children from low-income families.

The curricula and comparison groups fell into several categories; we concentrate on the ones involving the random assignment of classrooms to implement whole-child or skill-specific curricula, with some of the comparison groups allowed to implement curricula that had been developed independently by preschool teachers. In two of the contrasts, one involving math curricula and the second involving literacy curricula, the comparison condition was whole-child curricula – either a popular curriculum called Creative Curriculum or High/Scope, which was used in the famous Perry Preschool program. In a third contrast, Creative Curriculum was compared with whatever lessons the teachers had developed on their own.

Using PCER data, Jenkins et al. (2018) first examined impacts on classroom processes. As might be expected, the math curricula boosted the amount of classroom math instruction relative to math instruction in classroom with whole-child-only curricula, but they did not improve general process quality as measured by the ECERS. The literacy curricula failed to boost the amount of observed literacy instruction relative to whole-child curricula only, and it did not improve overall quality as measured by the ECERS. Interestingly, the whole-child curricula generated large improvements – exceeding one-half standard deviation – on the ECERS relative to locally developed curricula, and they improved the amount of observed math and literacy instruction as well. The PCER study did not measure the content of these locally developed curricula, but by the rating standards used in PCER, their quality appears to be quite low.



Based on data in Jenkins et al. (2018, Table 2), Figure 4 shows the impacts at the end of the preschool program year of the various kinds of curricula on children's academic skills at the end of the program. Comparisons between whole-child and locally developed curricula show no differences in literacy and math skills, in a composite of these two skills domains, nor in a teacher rating of children's social skills. These null results are remarkable; despite the widespread use—and average per-classroom price tag of $2,000—of whole-child curricula, they appear to be no more effective at boosting school readiness than the assortment of activities that early childhood education centers develop on their own. To be clear, these results do not establish that whole-child approaches are without value – only that curricula based on these approaches do not appear to be effective when subjected to rigorous evaluation.

Adding a math supplement to a whole child curriculum produced large gains in math achievement (shown in Figure 4) but in neither literacy nor social skills. And finally, the collection of literacy supplements produced gains in literacy (shown in Figure 4) but not in math or social skills. Importantly, though, none of the gains observed across the preschool year persisted to the end of the kindergarten year, which is consistent with the more general evidence on fadeout in preschool program impacts presented in Section 4.

The PCER Initiative Study focused on skill-specific curricula in mathematics and language. Additionally, several experimental evaluations of supplemental curricula and teacher

training modules designed to improve children's socioemotional skills have demonstrated effectiveness when compared with usual classroom practice. One of the most successful curricula, Preschool PATHS, has shown short-run impacts on children's emotion knowledge, problem-solving skills, behavior, and self-regulation (Bierman et al. 2008, Morris et al. 2014).

Most of the studies that have investigated impacts of more skill-specific curricula have measured impact while kids are still in preschool or at the start of elementary school, and do not have any follow-up assessments. This is problematic because many early skills develop very quickly in counterfactual conditions across the early and middle-childhood periods. For example, nationally normed reading and mathematics tests show that children learn more than a full standard deviation of material between kindergarten and first grade, but considerably less in later grades (Hill et al., 2008). Thus, while these kinds of early cognitive skills may be among the most fundamental and malleable, the effects of interventions targeting these skills may also fade out most quickly owing to control group catch-up.

The Agder Project. A field experiment in Norway, known as the Agder Project, tested effects of an intervention that introduced a structured and comprehensive curriculum for five-year-olds into the universal preschool context of Norway, where the norm was a whole-child curriculum (Rege et al. 2022). The intervention consisted of a nine-month curriculum (Størksen et al. 2018) with age-appropriate, playful, intentional skill-building activities in mathematics, language, social skills and executive functioning, along with relevant teacher training. The field experiment included 701 five-year-olds in 71 preschool centers. Treated centers implemented the structured curriculum for the five-year-olds and received teacher training, while the control group continued with the usual approach. The study assessed the children's language, mathematics and executive functioning skills at baseline, post-intervention (end of preschool), and in a one-year follow-up (end of first grade). The testers for all assessments were trained, certified and blind to treatment status. During the treatment implementation, the investigators conducted an implementation and process evaluation, which demonstrated high intervention fidelity (Størksen et al. 2021).

Rege et al. (2022) demonstrate that the structured curriculum intervention generated a positive effect of about .12 standard deviations on a summary score of children's skills in math, language and executive functioning measured post intervention and this effect persisted one year following the end of the treatment. At the one-year follow-up, the treatment effect was

particularly pronounced in mathematics. Treatment impacts were larger in the follow-up than they were immediately post intervention. For example, in math it was .16 standard deviations at the end of treatment and .23 in the follow-up. These larger treatment effects in the follow-up suggest a dynamic complementarity, but the differences in treatment impacts are not significant. The Agder Project will follow the children throughout primary and secondary school to investigate long-term effects.

Fidjeland et al. (2022) investigate differential treatment effects across gender of the Agder structured curriculum intervention. The analysis is motivated by substantial differences in skills across gender even before children start formal schooling, where girls particularly outperform boys in literacy and language skills (Brekke et al. 2018, Lundberg et al. 2012), executive functioning (Lenes et al. 2020, Matthews et al. 2009), and social and behavioral skills (DiPrete & Jennings 2012). In the Norwegian context substantial gender differences are also found in mathematics (Brandlistuen et al. 2021). Moreover, studies have documented that girls and boys seemingly spend their time in preschool differently, with girls much more likely to engage in activities that promote school readiness and skills development (Early et al. 2010, Ruble et al. 2007, Tonyan & Howes 2003). This suggests that boys may benefit more than girls from a structured curriculum because they need more support and scaffolding from teachers to engage in stimulating activities (Størksen et al. 2015). Consistent with this conjecture Fidjeland et al. (2022) find that the curriculum effects in the Agder Project are entirely driven by the boys, with a treatment effect of .24 on the sum score and .33 for math in the one-year follow-up. For boys, there is also a large and significant effect on executive function post-treatment.

The Agder Project design uses the same skills assessments before and after treatment and in a one-year follow-up, which allows for investigating the cross-productivity of treatment induced changes in executive function (Thijssen et al. 2022). A large literature in psychology suggests that executive functions should be targeted in early childhood education programs because they are strong predictors of academic achievements and they may be foundational for children's capacity to navigate the transition to formal schooling (e.g. Blair & Raver 2015, Diamond 2013). The capacity to control thoughts, feelings and behaviors may be important to adapt to the demands of the formal school context and for forming new friendships. Still, there is little evidence on how preschool program-induced improvements in executive functions promote other skills in primary school. The child development literature (e.g. Bierman et al. 2008, Raver

et al. 2011) has identified executive functions as important mediators of intervention effects on school readiness. However, these studies measured executive functions and school readiness skills simultaneously, which sheds limited light on the cross-productivity of program-induced executive functions.[11]

Thijssen et al. (2022) investigate the self- and cross-productivities of executive function by combining the experimental data from the Agder Project with a skill-building model (Cunha & Heckman 2007). The analysis applies the decomposition framework in Heckman et al. (2013), assuming that intervention-induced increments in measured skills are independent of intervention-induced increments in unmeasured skills (as in Berger et al. 2020, Conti et al. 2016, Heckman et al. 2013, Kosse et al. 2020). Interestingly, the evidence suggests that the treatment induced learning in executive function during preschool had a significant cross-productive impact on math at the end of first grade. More research is needed, but these results provide suggestive evidence that investing in executive function in preschool can increase effectiveness of investments in formal schooling.

5.5 Investments targeting teacher effectiveness

Earlier in this section we described the ingredients of a high-quality pre-K environment. As noted, chief determinants of quality include not only the curriculum but also the skill with which teachers deliver it and their sensitivity in interacting with the children. Teachers who struggle with classroom organization and providing emotional or instructional support to their students, who struggle to keep children on task, or who simply lack general classroom skills are unlikely to effectively implement any of the curricula described in Section 5.4.

Researchers and practitioners have long called for professional development (PD) for teachers to improve the quality and impact of early learning programs. PD, which is designed to improve a teacher's knowledge and practice, can take a variety of forms and often includes adding a coaching component to training in implementing a curriculum. It can be challenging to separate the effects of the curriculum from the training and coaching that might have accompanied its implementation.

[11] McCoy et al. (2019) is an exception. They investigate the long-term effects of treatment-induced executive functions and academic skills on self-reported high school performance.

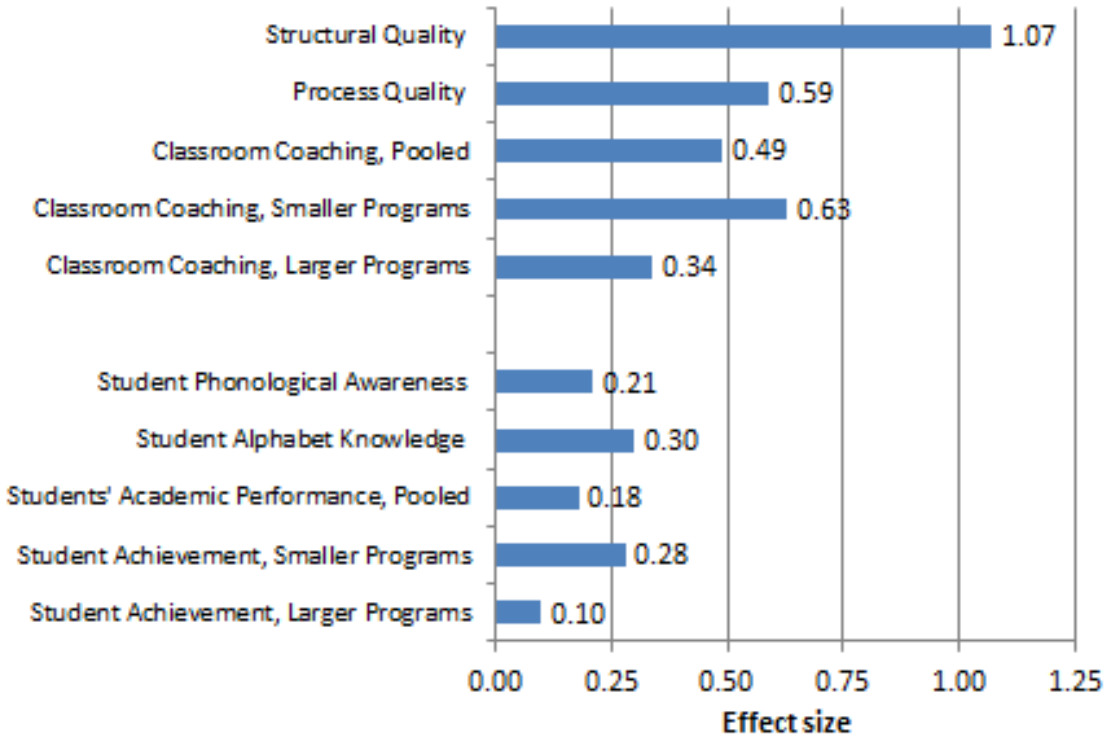
There is evidence that PD can improve teacher practice to at least a moderate degree. One recent meta-analysis (Werner et al. 2016) evaluated the results from 19 randomized controlled trials between 2003 and 2012 that were targeted at improving preschool teacher-child interactions and children's socio-emotional development in early childhood education programs serving low-income children. Nine of the 19 RCTs included in the meta- analysis focused particularly on caregiver sensitivity and responsiveness, and most of these also emphasized respect for children's autonomy, structure, and limit setting. Seven of the studies in the meta-analysis sought primarily to enhance verbal communication and peer interaction. Trials aimed at improving children's cognitive development were not included in this meta-analysis. The interventions were generally short-term, lasting several weeks to several months. The average effect size of the 19 randomized controlled interventions with combined outcome measures on all levels was .35. Programs with individual training sessions for caregivers generated significantly higher effect sizes (.41) than programs without individualized training (.09).

Another meta-analysis assessed the experimental impacts of language- and literacy-focused PD on teacher practice and on the literacy outcomes of preschool-age children (Markussen-Brown et al. 2017). This meta-analysis included 25 studies published between 1995 and 2014. Twenty-four out of the 25 studies come from U.S. samples; one is Canadian, but none is from Europe. Here, PD produced an effect size for process quality of .59 and an effect size for structural quality of 1.07, but no effect for educator knowledge. However, effect sizes for children's skills were considerably smaller. For instance, PD produced an effect size for phonological awareness of .21 and an effect size for alphabet knowledge of .30. The effect sizes were larger in programs serving economically-disadvantaged children.

Kraft et al. (2018) conducted a meta-analysis of the impacts of teacher coaching based on 60 randomized controlled trials conducted in the U.S. and other developed countries and published between 2006 and 2017 (Figure 5). Thirty-one of the 60 studies focused on teachers who worked in prekindergarten centers, while an additional 20 studies focused on teachers in elementary schools. This meta-analysis is especially useful in that it includes many RCTs that collected data on teacher practice as well as child test scores. Here, coaching programs include all in-service PD programs in which coaches or peers observe teachers' instruction and provide feedback to help them improve. Coaching is intended to be individualized, time-intensive, sustained over the course of a semester or year, context-specific, and focused on discrete skills.

The authors found large positive effects of coaching on teachers' instructional practices. Across all 43 studies that included a measure of instructional practice as an outcome, the analysis showed a pooled effect size of .49 SD.

Figure 5: The effects of professional development for teachers on classroom processes and child skills



However, the treatment impacts on student skills were far smaller. For the 31 studies that included measures of students' academic performance, the authors estimate that coaching raised student performance on standardized tests on average by .18 SD. The effects for both pre-K teacher outcomes and student outcomes were comparable to the effects for their K-12 counterparts. Also, the programs performed less well at scale: the authors estimate that smaller coaching programs improved classroom instruction by .63 SD and raised student achievement by .28 SD. As in the other meta-analyses, these effect sizes are statistically significantly larger than the size of effects on classroom instruction for larger programs (.34 SD) and on student achievement for larger programs (.10 SD). These findings illustrate the challenges of offering coaching at scale and the "voltage drop" (List 2022) commonly seen as programs expand and are further removed from researcher-driven high-quality implementation environments.

Although Kraft et al. (2018) lacked the underlying data to perform an instrumental variables analysis, they use a weighted linear regression framework to estimate that a 1 SD change in instruction is associated with a .21 SD change in achievement ($p = .16$). In other words, changes in student achievement appear to require relatively large improvements in instructional quality. This finding is consistent with a large body of literature documenting the weak relationship between educational inputs (instruction) and outputs (achievement) and helps explain why PD that results in more modest changes in teachers' instruction often does not lead to impacts on student achievement. Nonetheless, as the authors report, these effect sizes are larger than those of many other interventions in educational settings that are designed to raise student test scores. Kraft et al. (2018) also report data suggesting that traditional on-site coaching programs with a skilled coaching corps cost from $3,300 to $5,200 per teacher.

Finally, a recent quasi-experimental study conducted since the publication of these meta-analyses is informative for understanding PD at scale. Here, Rojas et al. (2020) leveraged a natural experiment in the large-scale pre-K system (n= 95 schools) in New York City (half of the children are low-income as defined by eligibility to receive free or reduced-price lunch) to assess the impact of PD combined with a targeted math curriculum called Building Blocks. The following supports were provided to teachers: (1) professional development delivered by the developers of Building Blocks and expert facilitators trained in its curriculum: teachers participated in 4 full-day training sessions (6 hours per day), and leaders participated in 3 half-day (4 hours per day) training sessions throughout the school year, and (2) on-site coaching for teachers by external coaches trained by Building Blocks certified trainers. On average, coaches observed the teacher in the classroom once a month for 1 hour and employed strategies such as modeling, providing feedback, and discussing implementation with the teachers. Coaches debriefed with educators during their visits, and they served as the direct point of contact for questions about the PD implementation.

Classroom-level outcomes were collected by trained graduate students, and the treatment was compared with control classrooms that had received an arts and music curriculum (similar in design and intensity to the math curriculum) as well as PD to support it. At the end of the first year of implementing the PD-enhanced math curriculum, the treatment showed positive impacts on the number and duration of teacher-led math activities and the measure of math classroom culture. The effect sizes were relatively large (with ESs = 0.95–2.73 standard deviations across

measures). These at-scale results are noteworthy, given that teacher attendance at the trainings was relatively low compared with prior effectiveness trials (Rojas et al. 2020). However, and perhaps most importantly, the PD had no discernible impact on the quality of the math activities that were offered. The math activity quality score assesses the extent to which teachers expanded children's conceptual understanding of math and extended children's mathematical thinking. As the authors note, improving early math instruction can be difficult for teachers; it requires them to know the content, understand children's thinking, engage in pedagogical practices that support learning, and see themselves as capable math teachers (Lee & Ginsburg 2009). It is possible that these positive qualities take more time to develop and do not take root until teachers have more experience with the curriculum.

Many have argued that of all the factors that influence a preschool's success, nothing is more important than teacher-child interactions (Hamre et al. 2017, Yoshikawa et al. 2013). Yet as the research reviewed in this section illustrates, this argument is weakened by the fact that interventions to improve teacher-child interactions or classroom instruction and organization are generally much more likely to have an impact on classroom processes than to produce meaningful improvements in child skill acquisition. The magnitude of the changes in classroom quality necessary to improve children's skills may be impossible to achieve at scale. Moreover, without knowing the costs of professional development (including the opportunity costs in terms of teachers' time) it is difficult to assess the efficiency of such approaches. It may be time to focus more attention on the most efficient and effective ways to improve children's learning and less on changing classroom processes.

5.6 Discussion and future research needs

The characteristics of the preschool environment (and, as we will see in the next section, the home environment) that correlate with children's subsequent school readiness and general well-being include the following: the teacher's use of complex language in consistent and responsive conversations with children; the teacher's ability to create interesting activities that engage children's attention and help to develop important academic, social and self-regulatory skills; and positive and emotionally supportive classroom environments, with an emphasis on teacher affirmation and warmth rather than disapproval (Burchinal & Farran 2020).

But how to translate these insights into effective programs? In other words, what programs or policies can be designed to promote these beneficial practices in preschools? Programs or policies have typically relied on two kinds of strategies: improving structural characteristics of the preschool and improving classroom processes.

National- or state-enforced structural regulations often focus on ensuring that classrooms meet minimal health and safety standards, establishing teachers' educational or credentialing requirements, and monitoring child-to-staff ratios. But while a structural approach might succeed in raising the "floor" of quality (e.g., assuring adequate levels of health and safety), most evidence on the impact of a teacher's education or credentials suggests that these things do not correlate consistently with gains in children's cognitive or socioemotional skills.

As for interventions to improve classroom processes: We are still struggling with the task of designing and implementing programs that improve the cognitive and emotional quality of teacher interactions with children and engage children in activities that boost school-readiness skills. Section 5.3 describes attempts to improve classroom quality by using rating systems based on systematic classroom observations, but these, too, have failed to improve child outcomes. Professional development, including coaching, for teachers shows some promise, though the treatment impacts are far greater for intermediate classroom process measures than for child learning outcomes, and the power of most programs drops as the programs scale up. Without knowing more about the costs and benefits of professional development and coaching at scale, it is hard to draw strong conclusions about their merit.

Prescribing curricula is a more promising approach. Research based on strong evaluation designs has identified several skill-based curricula that boost the respective skills of children who attend preschool. Unfortunately, most of these curricula studies have measured impact while kids are still in preschool or at the start of elementary school and do not have any follow-up assessments. We need more studies investigating long-term impacts of skill-based curricula interventions. This can help us better understand what skills curricula should target to have persisting effects.

In contrast to the skill-based curricular successes, curricula based on "whole-child" approaches have not yet proved to be effective. The theory behind whole-child curricula is in keeping with what is commonly thought to matter the most – child-centered active learning, facilitated by a classroom environment that is designed to promote engagement and curiosity. But

it takes considerable skill on the part of teachers to implement a whole-child curriculum effectively, and there is little evidence that scaled-up implementation of whole-child approaches improves academic or socio-emotional outcomes for preschoolers.

6. Improving the home environment and parental effectiveness

We turn now from environments children experience outside of the home and examine the role of the home environment. Gaps in children's skills based on family background appear long before most children enter preschool and kindergarten, which underscores the relevance of the home environment (Waldfogel & Washbrook 2011). Across disciplines, hundreds of studies have documented differences in the way advantaged and disadvantaged parents raise their children and have shown that these differences matter for children's development and adult success. Much of the theory about the importance of specific types of parental behavior comes from the field of developmental psychology. In this section, we review studies that try to estimate the causal effect of parent-child interaction in the home environment on child development and examine what the evidence says about the efficiency and effectiveness of interventions aimed at changing parental behavior.

We begin with a brief review of how the literature defines "quality" in the home environment. Numerous studies find positive correlations between these qualities and successful child development; however, far fewer studies have produced strong causal evidence. The basic idea is that much like teachers' roles in preschool environments, high-quality parenting is defined by consistent and responsive interactions that are sensitive to the child's developmental needs and abilities. In this view, the parent's job is to guide and encourage the child in a sensitive fashion, facilitating the child's continual mastery of new skills and competencies.

Of course, there are many constraints on parents' ability to provide a consistently sensitive and responsive learning environment that is perfectly attuned to the child's developmental needs. They can include a lack of information about child development and learning, material deprivation caused by low income, or stress that impinges on parental decision making and behavior. We review this literature and assess the effectiveness of interventions aimed at mitigating these varied constraints. We find merit in the idea that boosting family income can improve child outcomes. However, most existing programs intended to increase engagement among disadvantaged parents have met with little success, perhaps in part because their

complexity and demands seem to depress parental participation. We are more enthusiastic about newer programs that take a narrower and lighter-touch approach to changing parent behavior. We also leave open the possibility that new developments in technology may prove useful as tools to help support parent-child engagement in the home environment.

6.1 Defining quality of care at home

When children are young, parents are seen as key figures in their development. Through their time and attention, parents play important roles in cultivating children's skills and interests (Vygotsky 1967). The quality of the home environment, much like the preschool environment, can be characterized by structural and process factors. The most widely used measure of these factors in large national studies is the Home Observation for Measurement of the Environment (HOME), an observational and parent-report measure of the quantity and quality of material conditions, cognitive stimulation and emotional support children receive in the home (Bradley & Caldwell 1995). Structural conditions that are correlated with children's achievement and adjustment include absolute level of income (Duncan et al. 2010), as well as other quantitative dimensions of the home learning environment, such as the number of books and toys in the home; the amount and quality of food that is available to children; the regularity and sufficiency of sleep and other physical routines; and the safety of the home environment, including crowded conditions or exposure to second-hand smoke and other toxins.

The process measures captured by the HOME scale include both the quantity and quality dimensions of the behaviors reflecting "parenting." An example of quantity is how much time parents spend on a particular activity; sensitivity refers to how well-matched the parental response is to the child's bids for attention and other developmental needs. Two key dimensions of parenting behavior are parents' cognitive stimulation and emotional supportiveness. The former includes activities that foster cognitive growth, such as reading and playing, engaging in arts and crafts, and encouraging verbal interaction with the child. Emotional supportiveness refers to parents' affective responses to the child, such as warmth, affection and sensitivity, as well as refraining from harsh discipline or physical punishment (Bradley & Caldwell 1995).

Statistical studies suggest that parenting behavior as measured by the HOME scale accounts for a large share of the gap in children's cognitive and noncognitive skills by parental advantage (Cunha et al. 2006, Heckman & Masterov 2007, Waldfogel & Washbrook 2011). A

few studies suggest that spending time reading has a direct and causal effect on children's cognitive test scores (Kalb & Van Ours 2014, Price & Kalil 2019) and that the amount of time that parents spend in educational activities with their children is the single most important input for cognitive skill development (Fiorini & Keane 2014). Parental sensitivity is defined as parents' ability to perceive child signals, interpret these signals correctly, and respond to them promptly and appropriately (Ainsworth et al. 1974). Parental sensitivity is correlated with children's self-regulation, social functioning, and early cognitive skills (Eisenberg et al. 2001, Hane & Fox 2006, Kochanska 2002, Tamis-LeMonda et al. 2001). Studies have also shown that parent-child speech characterized by close-ended directives and prohibitions is not as strongly positively correlated with children's verbal development as is the use of open-ended questioning (Rowe 2018).

Research also describes the qualitative aspects of parenting in terms of an aggregate parenting style, and shows that the *authoritative* (versus *authoritarian*) parenting style is correlated with higher child skill development (see also Heckman and Mosso (2014) for a discussion of parenting). Authoritative parenting refers to a style of interaction in which high demands are placed on children, but combined with high levels of warmth and responsiveness. In contrast, authoritarian parenting sets strict limits but exhibits little warmth or dialog with children, and punishment tends to be harsh (Darling & Steinberg 1993, McCoby 1983). Doepke and Zilibotti (2019) provide a detailed economic treatment of parenting style as it relates to economic conditions in different countries and in different historical time periods.

Finally, discipline strategies represent a central component of socioemotional interaction between parents and children. Discipline is a key facet of the difference between authoritative and authoritarian parenting. Corporal punishment includes discipline strategies such as spanking, hitting with objects, and other actions that intentionally cause physical pain. Its use is associated with an authoritarian parenting style. In contrast, non-physical discipline strategies such as "time-out" and explanations for desirable behaviors are characteristics of authoritative parenting styles (Steinberg et al. 1994). Numerous correlational studies show links between corporal punishment, such as spanking, and a host of adverse cognitive and socioemotional child outcomes (Gershoff & Grogan-Kaylor 2016).

6.2 Economic resources

Hundreds of correlational studies have documented worse outcomes for children growing

up in low- as opposed to higher-income families in terms of physical and mental health, educational attainment and labor market success, risky behaviors, and delinquency (NASEM 2019). The causal mechanisms underlying these impacts on child development are a matter of debate. Theoretical models proposed by economists (Becker 1965, 1981, Del Boca et al. 2014, Heckman & Mosso 2014) view child quality as a multi-period product of the time and money that parents invest in their children. These parental investment decisions, in turn, are part of a larger optimization strategy involving parental decisions about time spent in paid employment, leisure and childcare, as well as expenditures on behalf of the children and themselves. Higher incomes from more generous transfer programs can increase parental time and money investments, although parents may also use some of the transfers to benefit their own consumption and leisure (Carneiro et al. 2021, Del Boca et al. 2016, Løken et al. 2012).

Some developmental psychologists focus on consequences of the fact that low-income families experience higher levels of stress in their everyday environments than more affluent families. This pressure, coupled with other stressful life events that are more prevalent in the lives of poor families, is thought to create high levels of psychological distress, including depressive and hostile feelings (Kessler & Cleary 1980, McLeod & Kessler 1990). Parents' psychological distress and conflict, in turn, are correlated with their parenting practices, which tend to be more punitive, harsh, inconsistent, and detached, as well as less nurturing, stimulating, and responsive to their children's needs. Such low-quality parenting is associated with elevated physiological stress responses among children and may ultimately harm their development (McLoyd 1990).

Evidence from the field of psychoneuroimmunology suggests that exposure to chronically elevated physiological stress may interfere with the development of poor children's stress response system and health, as well as the development of the regions of the brain responsible for self-regulation (Lupien et al. 2001). Nonexperimental studies have found that low-income children have higher levels of stress hormones than their more advantaged counterparts, and that early childhood poverty is associated with an increased allostatic load, a measure of physiological stress (Lupien et al. 2001, Turner & Avison 2003). These higher levels of physiological stress have been correlated with poorer cognitive as well as lower immunological functioning, the latter having long-term implications for a host of inflammatory diseases later in life (Miller et al. 2011).

Although research into the effects of income on these pathways is continuing, a clearer picture is emerging on the question of the causal effects of family income on child outcomes at

various point across childhood and into adulthood. Cooper and Stewart (2020) provide a general review of 67 RCT and quasi-experiment studies from developed countries that examine the impact of family income and resources on children's health, education and social and behavioral development. The vast majority (45 out of 61) of the reviewed studies find a significant positive effect of income across a range of children's outcomes, especially for children from low-income or disadvantaged families.

The nature and results from many of these studies can be gleaned from a recent review of the U.S. literature focused on children living in low-income families is provided in (NASEM 2019, Chapter 3). It distinguishes experimental and quasi-experimental studies of increases in cash income from impact studies of policies that transfer resources to families with children.

Beginning with cash income studies, the U.S. negative income tax experiments initiated under the Nixon administration provided the first random-assignment evidence of income effects on children. The large negative income tax payments reduced poverty and improved children's birth outcomes and nutrition, but had mixed effects on child outcomes such as school performance (Kehrer & Wolin 1979, Salkind & Haskins 1982). Impacts on school enrollment and attainment for youth were more uniformly positive (Maynard 1977, Maynard & Murnane 1979, Venti 1984).

One of the strongest quasi-experimental studies is based on data from the Great Smoky Mountains Study of Youth, which was originally designed to assess the need for mental health services among Eastern Cherokee and non-Indian, mostly White, children living in Appalachia (Costello et al. 2003). During the study, a gambling casino owned by the Eastern Cherokee tribal government opened on the tribe's reservation, which provided all members of the Eastern Cherokee tribe with an annual income supplement that grew to an average of approximately $9,000 by 2006 (Costello et al. 2010). The fact that incomes increased for families with tribal members relative to families with no tribal members provided researchers with an opportunity to assess whether developmental trajectories were more positive for tribal children than for nontribal children. The income supplements produced a variety of benefits for children in qualifying families – fewer behavioral problems such as conduct and psychiatric disorders (Costello et al. 2003), less alcohol and cannabis abuse, and less crime (Akee et al. 2010, Costello et al. 2010). Beneficial impacts on educational attainment were also found. Younger children and children with longer exposures to higher income had better outcomes.

Natural-experiment studies of the impacts of the U.S. Earned Income Tax Credit on child outcomes take advantage of the fact that federal EITC benefit levels increased substantially on several occasions between the late 1980s and the 2000s. It is important to bear in mind that the EITC affects family income through parental work effort, making it difficult to separate income effects from the effects of changes in parental employment (Hoynes & Patel 2018). Dahl and Lochner (2012) and Chetty et al. (2011) link EITC changes to children's achievement test scores over time and find that a $1,000 increase in family income raised math and reading achievement test scores by 6 percent of a standard deviation. Maxfield (2013) finds that a similar increase boosted the probability of a child would graduate from high school, earn a GED and complete one or more years of college by age 19. Taken together, the robust literature on the impacts of EITC-based increases in family income suggests beneficial impacts on children. Several other studies of U.S cash transfer programs also find positive impacts on a variety of child outcomes (e.g., the Aizer et al. (2016) evaluation of the Mother's Pension Program).

In addition to work on cash transfers of various kinds, there has been a great deal of research into the causal effects of what are sometimes called "near-cash" programs, especially those offering nutrition assistance and housing subsidies. The Supplemental Nutrition Assistance Program (SNAP program, formerly known as the Food Stamp Program) is by far the nation's largest near-cash program.[12] Hoynes and Schanzenbach (2015) also provide a summary of the literature examining causal links between SNAP participation and the nutrition and health outcomes of infants, children, and adults. In their 2016 study of possible long-term effects of food stamp coverage in early childhood on health outcomes in adulthood, Hoynes et al. (2016) focus on the presence or absence of a cluster of adverse health conditions known as metabolic syndrome. Scores on these indicators of emerging cardiovascular health problems grew worse as the timing of the introduction of food stamps shifted to later and later in childhood. It is impossible to determine the extent to which the adult health benefits of food stamp availability in very early childhood were generated by the nutritional advantages of the extra spending on food or by the more general increase in economic resources freed up for spending on other family needs.

Not all studies of in-kind program subsidies show beneficial impacts. (Jacob et al. 2015)

[12] https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap.

compare children in families that won the lottery allocating Section 8 housing vouchers in Chicago with children in families that lost that lottery. They find virtually no differences across a range of outcomes in educational attainment, criminal involvement, and health care utilization. Given the size of their lottery sample, these null effects are precisely estimated. Other studies of housing vouchers have found more positive impacts (Carlson et al. 2012a, Carlson et al. 2012b, Currie & Yelowitz 2000).

Although the Moving to Opportunity Program has generated the most interest in results for its treatment arm that combined housing vouchers with the requirement to relocate in a very low-poverty neighborhood, it also provided comparisons between control-group children and children in families receiving the economic resources embodied in conventional housing vouchers. Gennetian et al. (2012) find no differences across a range of schooling, health, and behavioral outcomes measured 10 to 15 years after the study began. The longer-run examination of college and labor market outcomes by Chetty et al. (2016) also failed to find statistically significant outcomes between these two groups, even for children who were younger (under age 13) when they entered the study. These results, when combined with the lottery results reported in Jacob et al. (2015), suggest that increasing resources through housing choice vouchers is unlikely to benefit children.

What conclusions does this evidence support about the role of income for the development of children in low-income families? The (NASEM 2019, p. 89) report provides a useful summary: “The weight of the causal evidence indicates that income poverty itself causes negative child outcomes, especially when it begins in early childhood and/or persists throughout a large share of a child’s life. Many programs that alleviate poverty either directly, by providing income transfers, or indirectly, by providing food, housing, or medical care, have been shown to improve child well-being.”

As mentioned in our discussion of Cooper and Stewart (2020), evidence on the developmental effects of income increases for children in families with middle- or upper-middle-class incomes is much weaker. Løken et al. (2012), for example, examine the effects of a positive income shock for low-, middle- and higher-income families. They use temporal and regional variations in the oil boom in Norway as a source of exogenous income variation. Their nonlinear instrumental-variables estimates show an increasing but concave relationship between family

income and children's outcomes. In other words, income appears to matter more for children of the poor than for children of the rich.

6.3 Boosting child skills by investing in targeted parental behaviors

Myriad intervention programs have been designed to boost children's skills by changing parental behavior. Not surprisingly, most of these interventions are offered to economically disadvantaged families with the goal of closing income-based gaps in children's skill development.

Before describing the leading parental behavior interventions and their impacts, it is useful to assess the leading theories about why income-based differences in parent behavior arise, because a key part of understanding why interventions succeed or fail depends on whether scholars have the theory of change right. Here we focus on information and other "process factors" (as distinct from family income, which was described in the prior section).

In addition, before engaging with this literature, it is useful to observe that many of the studies we will discuss depart from conventional economic analyses of early childhood education in their motivation, interpretation, and policy prescription. The premise of most conventional economic analyses is that the individual or the family is maximizing something, even if the "something" is unorthodox. For example, the way in which parents relate to their children is modeled by appealing to altruism, which means, formally, that the child's utility (or consumption) enters into the parent's utility function. Few economists are willing to concede that individuals simply do not know what they are doing. Instead, individuals are assumed to do what they expect is best given the constraints they face and the information they have. Yet there is ample opportunity for policy or interventions to improve individuals' outcomes, for example by alleviating constraints and market imperfections, reducing uncertainty, and lowering the costs of acquiring and processing information.

The departure from conventional economic analysis is often motivated by the argument that economic theory is too restrictive and at odds with the choice behavior. Here, some argue that cognitive biases and inaccurate beliefs are necessary to explain behavior and prescribe policy. However, it is worth noting that observed choice behavior may be consistent with many alternative specifications of preferences, constraints, and expectations. Sometimes the trade-offs that individuals face are subtle and difficult to gauge for researchers. Prices, opportunity costs,

and expectations are not necessarily parameters that are observed in data, but they affect behavior nonetheless. Thus, drawing conclusions about cognitive biases and inaccurate beliefs normally requires data other than observed choices and constraints.

The research addressed in this subsection tries to address this data challenge by surveying individuals about their preferences or expectations. Assuming that these stated preferences and expectations are accurately reported, they can be compared with observed choice behavior and (hopefully) objectively correct expectations to draw inferences about cognitive biases and inaccurate beliefs. If one concludes that such biases and inaccuracies exist, interventions are carried out to correct people's beliefs and improve their decision making.

When interpreting the findings from these studies, however, it is important to note that knowing which interventions can change behavior and perceptions does not necessarily tell us which interventions should be adopted to change a given outcome, such as a specific skill. One reason is that it is difficult to identify individual preferences and expectations. Another is that researchers are rarely fully aware of the parameters that govern individual choice behavior, such as payoffs and costs. For these reasons one may argue that interventions targeting individuals' beliefs and decision making is at best risky, and possibly harmful.

Finally, there is no guarantee that changing a specific parent behavior will lead to the expected change in child outcomes, and this is especially true in the realm of human capital development. For instance, one can probably get parents to read more to their children through a variety of incentives. But more reading will not translate into higher literacy skills if the reading is done reluctantly, with little skill or enthusiasm, by caregivers who are stressed or who dislike reading, or with diminishing marginal returns. In short, there is little reason to believe that motivating all parents to perform a task, or to do it more frequently, will result in that task being performed equally well by all parents. This point is especially germane to parenting interventions because programs are generally deemed to be more successful if they induce more parents to participate, and because the assumption tends to be that more of a given behavior is better (i.e., more math, more reading, more interaction).

Information, values, and preferences about parenting and child development. SES-based differences in investments in children stem in part from differences in the degree to which parents value or enjoy spending time engaged in sensitive interactions and enriching activities with children, as well as in the degree to which they are informed about these things. To conclude that

these factors drive SES-based differences in parental behavior, we would need strong evidence that information, values, and preferences differ by SES. However, most recent evidence suggests that U.S. parents across the economic distribution share similar beliefs about the importance of children developing skills that will prepare them for success in school and life and about the values they wish to instill in their children (Mayer et al. 2020).

Another important aspect of the parental belief system relates to parents' "terminal values" for their children, or those characteristics parents believe they need to instill in children to prepare them for life (Alwin 2001). Historically, high-SES parents have valued "independent thinking" or "self-direction" more than low-SES do, whereas low-SES parents have tended to value "obedience" and "conformity." Differences in parental values help account for the reproduction of social class because they influence in myriad ways how parents prepare their children for their academic and professional futures. But Ryan et al. (2020) show that rich and poor parents' ideas about the characteristics children need to succeed in life (i.e., the ability to work hard, be helpful, think for oneself) have converged substantially in the last three decades (see also Ishizuka 2019).

It is possible that although all parents share similar aspirations for their children's development and readiness for school, lower-SES parents may expect a lower return from their investments in their children relative to the payoff they receive from spending time in another activity. Research by Mayer et al. (2019) and by Cunha et al. (2013) shows that low-income parents do in fact expect a positive return from the time they spend in educational activities with their children. But even if parents perceive a positive return to one activity, such as time spent reading with a child, the perceived return to an alternative activity may be higher. In a representative survey of British parents of school-aged children in England, Attanasio et al. (2019) document that parents believe that a weekly parental time investment of 3 hours or a material investment of £30 weekly matters more than moving a child to a better school; this study also found that the perceived productivity of investments is not related to parental background.

Low-income parents may be less informed about the appropriate timing of key parental inputs, in terms of child development, and about children's developmental milestones. As a result, they may fail to promote important skills at a time when it would be most developmentally effective. Rowe (2008) offers descriptive evidence that low-income parents, compared with their higher-income counterparts, respond less often to their young children's utterances, at least in

part because they believe that adults cannot "make" babies talk. Kalil et al. (2012) showed that college-educated mothers were more likely than their less-educated peers to exhibit a "developmentally-tailored" pattern of time investment in their children – they invested more time in activities that were tailored to the child's developmental stage.

Income may also shape the extent to which parents view the time they allocate to their children in terms of direct utility rather than investment utility. Economists have long posited that highly educated parents, more than less educated parents, view time with children as an "investment behavior" that can increase children's future human capital (Guryan et al. 2008). The same theories suggest that highly educated parents may spend more time with their children because they derive more enjoyment from doing so. Kalil et al. (2020) found that for all mothers, spending time caring for a child is associated with more positive feelings than spending time in other activities. However, although highly educated mothers tend to spend more time engaged in childcare, the findings offered no support for the hypothesis that these mothers *enjoy* time spent caring for a child more than their less-educated counterparts do.

In sum, there is at best mixed evidence that information, values or preferences drive differences in parenting behavior across the socioeconomic distribution. Relative to higher-SES parents, lower-SES parents may have less nuanced ideas about how to promote child development and they may underestimate the benefits of time spent promoting child development, but parents in both of these groups value and understand the importance of engaging in enriching behavior, such as reading, with their children to fairly similar degrees and they appear to enjoy this time in equal measure.

Family and environmental stress. According to the "family stress model," economic pressures faced by low-income families, when coupled with other stressful life events that tend to be more prevalent in these families' lives, can create psychological distress that disrupts parent-child interactions (Kessler & Cleary 1980, McLeod & Kessler 1990). Poverty and chronic economic strain may also undermine parenting quality (Gershoff et al. 2007). Stress may tax parents' ability to acquire and process information in a way that supports their investments in their child's skill development (Lichand et al. 2021).

Research has substantiated many of the hypothetical links in the family stress model. Evans and Garthwaite (2014) find that EITC expansions during the 1990s are linked with low levels of stress-related biomarkers gathered in a large national health survey. According to the

family stress model, this psychological distress spills over into relationships throughout the family. Gassman-Pines (2011) demonstrated that among low-income mothers, non-standard, night-time work on one day – a daily stressor far more common among low-income parents – was associated with elevated parenting stress and higher levels of mother-child conflict the following day.

In recent years the family stress model's focus on environmental conditions and parental mental health has been broadened to include an understanding of how stress impacts neurobiological and cognitive functioning. In the context of scarcity, including a lack of money, parents may be more likely to make decisions that emphasize short-term rather than long-term gains, and invest less time in developmentally enriching activities as a result (Lichand et al. 2021, Mani et al. 2013, Mullainathan & Shafir 2013). Accordingly, the possibilities for purposeful, goal-directed parenting are greatly diminished.

Constraints on decision-making. Based on the evidence reviewed above, disadvantaged parents seem to want to do many of the same things as advantaged parents do, particularly things like reading to children and taking them on educational outings that are associated with more positive child outcomes, but they are less likely to do those things. The literature in behavioral science and behavioral economics, which departs from conventional economic theory in allowing for "irrational" decision-making, describes a set of "cognitive biases" that may give rise to this gap between knowing and doing.

Like many other decisions, parenting decisions are complex, and parents find them difficult simply because parenting is such a complex area. For this reason, parents are prone to rely on heuristics (cognitive "shortcuts") to simplify their decisions and make them "computationally cheap" (e.g. Gigerenzer & Selten 2001). Advantaged and disadvantaged parents, for a variety of reasons (e.g., differences in stress, differences in the composition of parents' social networks, differences in parents' own upbringing or experiences) may rely on different cognitive heuristics in making decisions, or they may experience the same cognitive heuristic differently, which might also result in different patterns of decision making. Here we describe two potentially important characteristics of parenting that make it especially susceptible to these "cognitive biases" and lead to differential adaptions to them by parental SES.

Parenting investments have uncertain returns. The payoff to many parenting decisions does not materialize until years into the future. Decisions about spending money and time on

children's schooling, extra-curricular activities, health-promoting behaviors, and other activities meant to improve child outcomes are decisions about investments with uncertain returns.

Research suggests that under these conditions, people systematically overweight present outcomes compared with future outcomes, and this leads to inconsistencies in individuals' time preferences (O'Donoghue & Rabin 2015). Present bias can cause parents to prioritize spending time on activities that provide immediate gratification rather than investing that time in their children, decisions they may regret once their children have grown up. There is no consensus on what causes differences in time preferences. Becker and Mulligan (1997) proposed that the more financial resources one has to support future planning, the lower the discount rate on the future; empirical evidence supports this hypothesis (Dohmen et al. 2010, Harrison et al. 2002, Lawrance 1991).

Parents make some decisions automatically. Parenting often requires quick, on-the-spot decisions. When a child runs towards a busy street, a parent must react rather than contemplate. When a child screams in the checkout lane because the parent says no to his request for candy, there is little time for the parent to reflect on what to do. The need to act quickly and on the spot results in automaticity. Automaticity is a response that occurs with minimal cognitive processing. It is a useful heuristic that reduces cognitive load. An automatic response can be beneficial if it is efficacious, but costly when it is not. Because automatic responses can be likened to habits and habits are hard to break, ineffective automatic responses can lead to ineffective parenting.

Parenting behaviors are correlated across generations (Van Ijzendoorn 1992). Thus, advantaged and disadvantaged parents may have different parenting habits owing to differences in their own upbringings in different socioeconomic circumstances. Any habit is hard to break, but if the automatic behavior is reinforced in the in-group or social network in which parents learn and practice parenting behaviors, it is even more difficult to change. Automaticity bias can therefore play a role in reinforcing SES-based differences in the cognitive stimulation and emotional support that parents provide, which are linked to children's skill development.

In sum, many mechanisms have been posited to explain the origins of SES-based parenting differences. These include meaningful SES-based differences in money and time constraints, parents' knowledge, values and preferences, family and environmental stress, and cognitive "shortcuts" that impact parents' daily decision-making. There is consistent evidence that low- and higher-SES parents interact with and invest in children differently because they

experience more stress than higher-income parents, both because that type of stress increases parents' depression and anxiety and because it can undermine parents' cognitive ability to focus on long-term, rather than short-term, goals. Additionally, recent research has found that the economic context of parenting itself, regardless of its impact on mental health, is associated with parental decision-making that focuses on present versus future gains and relies on habits rather than conscious problem solving.

Effectiveness of Policies and Programs at Changing Parental Behavior. One of the most common strategies for changing parental behavior is through interventions that attempt to affect parents' information about, values around or preferences for investments in children. Of the policy approaches to narrowing parenting gaps between higher-and lower-SES parents, home visiting programs are by far the most common, and one of their goals is to address these mechanisms. These programs typically target the mother-infant relationship, aiming to enhance child development by modeling appropriate care for infants, toddlers and preschool-aged children or providing direct instruction. This approach assumes that if parents are informed about the importance of certain parenting behaviors and instructed on how to engage in them, they will do so more often. The number of families served by home visiting programs increased most recently with the passage of the Maternal, Infant, and Early Childhood Home Visiting Program (MIECHV) as part of the Affordable Care Act (ACA) in 2010, and funding for MIECHV has been authorized through fiscal year 2022 (Michalopoulos et al. 2019). As the name suggests, home visiting models generally involve regular (i.e., weekly, bi-weekly, or monthly) visits to a family's home by a trained nurse or nurse practitioner. This practitioner may share information with the parent, often in the form of "tip sheets" about child development milestones and best practices for parenting behavior (see e.g. Doyle 2020). The practitioner may also serve as an ongoing source of support and connect the family to other resources in the community. In some program models, parents are invited to participate in group sessions with other parents to discuss parenting and gain social support.

A recent example from Ireland is the Preparing for Life Program (Doyle 2020), which incorporated home visiting, group parenting classes, and baby massage into an intensive 5-year intervention for economically-disadvantaged Irish families. In a randomized control trial of 233 families, the program was shown to have substantial treatment effects, raising children's cognitive and socioemotional/behavioral scores by two-thirds and one-quarter of a standard

deviation, respectively. Analysis of earlier waves of the study showed that Preparing for Life had significant treatment effects on parents' stimulation, interactions, and parenting strategies as measured by the quality of the home environment, time use, nutrition, and discipline strategies (Doyle 2020). The data suggest, though do not confirm, that the improvements in parenting account for some of the improvements in children's cognitive skill. However, the measures of parental investment do not significantly mediate the relationship between the treatment and children's socioemotional/behavioral scores (Doyle 2020). The question of whether targeting parenting as the most effective means to improving children's skill development thus remains an open one.

The U.S. Department of Health and Human Services Home Visiting Evidence of Effectiveness (HomVEE) review determines which home visiting models have enough evidence to meet the U.S Department of Health and Human Services (HHS) criteria for an "evidence-based early childhood home visiting service delivery model." This process is conducted annually, and as of December 2020, 21 of the 50 models that were reviewed met those criteria (see Table 1 at https://homvee.acf.hhs.gov/sites/default/files/2020-12/HomVEE_Summary_Brief.pdf)

Although states can choose from among a variety of evidence-based home visiting programs under MIECHV, three of the most frequently adopted and rigorously evaluated models are the Nurse-Family Partnership (NFP), the Early Head Start – Home Visiting program (EHS), and Healthy Families America (HFA) (see Ryan and Padilla (2019) for a detailed review of this literature with additional citations).

As Ryan and Padilla (2019) report, although each of these programs has demonstrated positive effects on some parenting outcomes, the effects vary across sites even within programs. For example, multiple evaluations of HFA have demonstrated modest effects on mothers' ability to interact in sensitive or stimulating ways with children, but only at some sites. Another home visiting program, Home Instruction Program for Preschool Youngsters (HIPPY), which specifically targeted parents' reading and educational activities with preschool-aged children, showed robust effects on children's school performance in the first cohort (Baker et al. 1998), but not in subsequent ones (Baker et al. 1999). More importantly, even the programs or program sites that have been found to enhance parenting practices show relatively modest effect sizes (e.g. Love et al. 2005, Olds et al. 1988) that often fade (Chazan-Cohen et al. 2013).

The most recent evidence on the effectiveness of home visiting programs comes from a

national evaluation of the four most commonly used MIECHV-funded programs: the three programs described above as well as the Parents as Teachers (PAT) program. Like previous evaluations of specific programs, the evaluation, called the Mother and Infant Home Visiting Program Evaluation (MIHOPE), found that home visiting programs are capable of increasing the quality of cognitive stimulation in the home and decreasing the frequency with which parents use harsh or aggressive disciplinary approaches, but these effects are modest in size, hovering near .10 of a standard deviation (Michalopoulos et al. 2019). Moreover, only about one-third of the parenting outcomes examined showed effects that were statistically significant.

Table 6 shows impact results from seven (out of 21) programs included in the December 2020 U.S. Department of Health and Human Services Home Visiting Evidence of Effectiveness (HomVEE) review of evidence-based programs. These seven represent programs for which a review was last updated in either 2020 or 2019. The table displays the number of treatment impacts in three domains for which significance reached $p < .10$ for all outcomes examined. The ratio of significant treatment impacts to outcomes examined is strikingly low.

[**Table 6:** Ratio of statistically significant ($p < .10$) treatment impacts to outcomes examined from the HOMVEE literature review

	Positive parenting practices	Child development and school readiness	Reductions in child maltreatment
Attachment and Biobehavioral Catch-Up	11/23	8/19	Not measured
Healthy Families America	28/131	12/56	20/209
Home Instruction for Parents of Preschool Youngsters	1/1	11/49	Not measured
Maternal Infant Health Program	Not measured	Not measured	0/18
Nurse Family Partnership	7/37	13/142	7/26
Parents as Teachers	3/92	7/67	0/4
Play and Learning Strategies (PALS) Infant	1/13	6/14	Not measured

Theoretically, there is merit in the idea of combining early education and parenting programs. This could be a promising approach if the programs complement each other. However,

so far the evidence on this approach is weak. As we pointed out in Section 3, Rossin-Slater and Wüst (2020) showed using Danish data that early childhood education and home visiting programs substitute for one another. Examining 321 cognitive and achievement impacts in a meta-analysis of 46 studies, Grindal et al. (2016) found no significant differences in outcomes between early childhood education programs with and without parenting components, although it appeared that programs with at least monthly home visits or with parent practicing stimulating behaviors had some impact. Early childhood education programs that also offered intensive home visits (i.e., more than one per month) had notably larger effects on young children's skills; however, the program costs and the challenges of enrolling parents and keeping them engaged in home visiting programs of that intensity have been noted elsewhere. The evidence does not bolster confidence that combinations of parent- and child-based programs will always (or even ever) be synergistic.

However, it may be that the primary mechanism these programs use to enhance parenting practices – providing information and instruction – is not the mechanism that differentiates the behavior of lower- and higher-income parents. Earlier in this section we reviewed literature that suggests that low- and high-income parents largely share the same parenting goals and values, and agree about the importance of engaging in enriching activities with their children. The best evidence suggests that low- and high-income parents differ primarily in the level of stress they experience while parenting and the impact of certain cognitive "shortcuts" on parenting decisions. Accordingly, new programmatic approaches that either focus on reducing stress directly or offer a less time-consuming (and potentially less stressful) approach may hold promise for future parenting interventions.

Family and environmental stress. All of the home visiting programs described above aim to alleviate parental stress and improve parental mental health, among other goals. Nonetheless, only HFA has demonstrated impacts on maternal mental health, and only at certain sites (see Ryan & Padilla 2019). The MIHOPE did find that home visiting, averaged across model programs, was associated with significant reductions in mothers' depressive symptoms, but the effects were small (Michalopoulos et al. 2019). These small and inconsistent findings are likely attributable to the programs' broad set of targets – they do not focus exclusively on addressing maternal mental health or reducing stress. In contrast, smaller, more targeted programs show that improving parents' mental health or reducing stress results in improved parenting behavior.

A program called “Mothers and Babies” is a 6-week cognitive-behavioral intervention with demonstrated efficacy in reducing depressive symptoms and preventing depressive episodes among perinatal women, which has been shown to elevate maternal sensitivity to their infants (McFarlane et al. 2017, Tandon et al. 2018). Another well-developed and rigorously evaluated intervention aimed at improving parental mental health is Family Foundations (FF). FF seeks to improve mental health by minimizing the strains associated with the transition to parenthood, increase coparental support, and decrease coparental undermining. At 6 months and one year, FF participants reported significantly fewer maternal and paternal depressive symptoms, decreased maternal anxiety, and increases in both mothers’ and fathers’ coparental support relative to controls (Feinberg et al. 2009). In interactions with children, parents showed increased sensitivity, support for child exploration, and positive affect, and intervention fathers showed less negativity. Remarkably, these effects were largely sustained at 3 and 6 years post-program (Feinberg et al. 2014).

This literature suggests that parenting interventions can improve the quality of parenting in low-SES families by reducing their emotional and interpersonal stress and improving their mental health. Parents’ ability to provide emotional support in their interactions with young children showed particular improvement. However, long-term home visiting programs that have diffuse goals are unlikely to produce such effects with consistency or at significant levels. Rather, programs that focus more specifically on improving parental mental health using clinical approaches hold the most promise for enhancing parenting behavior through this particular mechanism.

Lighter-touch approaches to delivering information and instruction. The home visiting programs described above typically experience substantial difficulty recruiting and retaining families (Duggan et al. 2018). Programs often recruit fewer than their target number of families and then provide far fewer home visits than programs expect. On average, program families in the MIHOPE participated in programs for only 8 months, even when the programs were designed to last several years, and 17% of program mothers never received a home visit at all (Duggan et al. 2018). Perhaps because of the hectic schedules and non-standard hours many low-income families have to contend with (Mytton et al. 2014), or because parents lack enthusiasm for the program model, with its demands on time and in-person interaction, it is difficult to deliver the programs with fidelity.

To address these issues, some new pilot parenting interventions have developed lighter-touch approaches to delivering information and instruction. York et al. (2019) use a text-messaging program for parents of preschoolers that provides tips on specific actions parents might take to increase their child's literacy skills. Their eight-month intervention of three texts per week increased students' scores on an assessment of early literacy by 0.11 SD (with a standard error of 0.054). These effects were concentrated among students who scored in the bottom half of the distribution of literacy test scores at baseline. York et al. were not able to directly measure the mechanism leading to the increase in test scores. Parents in the experimental group reported more engagement in the literacy activities that were closely associated with the content of the text messages. However, the increase was estimated imprecisely, and the self-reports of engagement may have been primed by the experiment itself rather than reflecting actual behavioral changes. Nonetheless, the treatment impacts on young children's language skills exceed those found in the far more intensive, and expensive, home visiting programs.

Cognitive shortcuts and decision-making. Another, newer type of parenting intervention targets specific, discrete parenting behaviors and, through behaviorally informed actions, addresses the cognitive shortcuts that may prevent parents from engaging in specific parenting practices. Mayer et al. (2019) tested a behaviorally informed intervention designed to increase the amount of time low-income parents spend reading with their children. Hypothesizing that present bias might be key, the researchers designed the intervention specifically to overcome this bias with a specific set of behavioral tools (goal setting, feedback, timely reminders, and social rewards) to "bring the future to the present" and help parents form a habit of regular book reading. These tools were all deployed using text messages, rather than through in-person visits from program staff, to make participating in the program relatively easy for low-income parents with hectic, unpredictable schedules and high levels of daily stress.

On average, the intervention had a very large treatment impact (~ 1 SD) on the amount of time parents spent reading with their children (the study used digital tools to measure time use objectively). But even more important was the study's finding that the intervention was substantially more effective for parents who were more present-biased (the study used standard techniques to assess present bias among all of the parents prior to the intervention; Andreoni and Sprenger (2012)). In short, parents who suffer from present bias are the very ones who benefit from an intervention designed to overcome it. Parents who were not present biased were already

reading to their children more frequently and the intervention had little impact on them. These findings suggest that a parent's failure to read to children is due in part to difficulty making temporal tradeoffs. These findings provide a blueprint for managing this cognitive bias. By using a set of known behavioral tools, parents can increase desired behaviors and improve their decision-making.

Similarly, (Kalil et al. 2021) implemented a behaviorally informed field experiment called Show Up to Grow Up designed to increase attendance and reduce the number of chronic absences at subsidized preschool programs in Chicago. This study sent personalized text messages to parents, targeting behavioral bottlenecks driving children's absences from preschool. According to outcome data from administrative records from preschools, the intervention increased attendance days by 2.5 (0.15 standard deviations) and decreased chronic absenteeism by 9.3 percentage points (20%) over an 18-week period. The text messages focused on, among other things, correcting parents' mistaken beliefs about Head Start by informing them of their child's actual number of absences over the prior months and by emphasizing the learning that children miss out on when they are absent. Here, we can think of incorrect beliefs as an informational constraint. The results of the intervention showed that the parents who benefited the most were the ones who reported lower preferences for attendance at baseline. These are the parents who are less likely than others to report that their child would be worse off in terms of academic and social skills if they missed many days of preschool. In short, parents who are less convinced of the importance of preschool benefit most from messages and reminders that emphasize its importance.

6.4 Discussion and future research needs

This chapter presents evidence showing that the markers of quality in the home environment parallel those in the preschool environment. High-quality parenting is defined by consistent and responsive interactions that are sensitive to the child's developmental needs and abilities. Numerous interventions, delivered to parents directly, have sought to help parents achieve this goal, based on the assumption that changes in parents' behavior will lead to changes in children's skill development. Some programs try to change parents' behavior directly through interventions such as home visiting programs that give parents information about child development and effective parenting strategies. Other programs try to change parents' behavior

indirectly by reducing their stress. In the realm of policy interventions, a variety of programs transfer economic resources to families with children. Higher incomes from more generous transfer programs can increase parental time and money investments and may also reduce parents' stress and improve their emotional well-being

What have we learned from these studies? Causal research on the impact of income increases on children living in low-income families has grown substantially in the past 15 years and supports the conclusion that income matters for these children. However, much of this literature is based on data that provide a look at the total effect of income, but not on the mechanisms by which the income effects occur. Future research should focus on understanding the roles of the investment and stress pathways behind the income effects.

Turning to interventions focused on improving quality in the home environment, we have summarized many examples of both intensive and light-touch parenting programs that may change some parenting behaviors, but rarely improve child well-being. A useful avenue for future research would be to identify more efficient and effective approaches that focus directly on improving children's skills instead of designing complex, expensive, and time-consuming interventions that take an indirect approach by targeting parents. At a minimum, future research should address the perennial problem of low rates of parent engagement and participation in most parenting programs operating at scale. It is time for a new paradigm that takes a more precise and tailored approach to changing parents' behavior in a way that reduces programs' demands on parents' time and attention.

Given the limited impacts of most scaled-up parenting interventions, researchers and practitioners should consider departing from the standard parent training programs that "teach" parents how to behave differently. There is no causal evidence that "more" of every type of parent investment is always better (e.g., time spent doing math activities with children) and this may even be counterproductive if parents themselves lack skill or confidence in these arenas. A better approach may be to focus on encouraging parents to engage in "some" rather than "none" of a key set of parental activities. At the same time, it is important to do everything possible to eliminate child abuse and neglect, as well as exposure to toxins and hazards, in the home, as we do in preschool settings.

Future research might consider the potential of high-quality educational technology solutions that might complement or even substitute for some parental inputs in the home

environment. Interventionists could then design and deliver programs designed to help parents with the things only a caregiver can do, such as talking and listening to children and fostering a nurturing and affectionate emotional climate in the home. Our review also suggests that the notion of constraints on parental behavior should expand to include not only information and credit constraints, but also mental stress and its deleterious effects on parents' emotional energy, attention, and decision-making.

7. Knowledge needs

The foundations for successful child development are established in early childhood. Beyond the provision of health care, the two main policy approaches for strengthening these foundations have been subsidized preschool programs and programs targeting the home environment. Our chapter reviews a large body of empirical work investigating whether these programs make a difference for children's development, and if so, how and under what conditions do they help, how cost-effective are they, and which programs are scalable? Although much progress has been made, existing research on early childhood education falls short of answering fundamental questions about what works for whom and why.

The body of work to date consists of a relatively small theoretical literature and a large, often quite atheoretical, empirical literature. Unfortunately, much of the empirical work presents a catalog of undigested effect estimates, often unrelated to economic policy, with little if any link to explicit theory or economic models, and lacking the data needed to understand the relevant mechanisms, such as counterfactual mode of care and measures of mediating factors.

A tighter link between theory, econometric methods and data is essential to compare and reconcile the mixed and sometimes conflicting empirical results across studies, and to understand when and why the impacts of home-environment and pre-school interventions fade out. Such an understanding is needed not only to explain how current policies affect children in the short and longer run, but also to prescribe new policies that can help children realize their potential in life.

The following subject areas are of particular importance if we are to achieve progress: *Counterfactual mode of care and parental investments.* As discussed well in Heckman (2010), there are two distinct tasks in causal inference and policy analysis: (a) defining counterfactuals and (b) identifying counterfactual outcomes from data. Early childhood education has made considerable progress on the second task by taking advantage of randomized controlled trials or

quasi-experiments. Less progress has been made on the first task, however. For example, the counterfactual mode of care is rarely described in sufficient detail. Moreover, parental responses to interventions are seldom thoroughly investigated. These pieces of information are essential to arrive at a precise definition of counterfactual states, and thus to interpret in economic terms the estimated program effects and to understand how these effects vary over time, or to draw comparisons with other studies.

Long-term effects of curricula investments. A substantial literature suggests that investing in skill-specific preschool curricula is a promising approach that warrants more research. Most of these studies have only measured impacts while kids are still in preschool or at the start of elementary school. We need more studies investigating long-term impacts of skill-based curricula interventions. This can help us better understand what skills curricula should target to have persisting effects.

Dynamics and follow-on investments. Most evaluations of early preschool and parenting programs fail to measure and account for the quality of follow-on investments in K-12 and beyond. This may be a major limitation if complementarity or substitutability is empirically important for child development, as some of the research that we reviewed suggests. Future evaluations of home and preschool interventions would benefit a great deal from measuring (and, ideally, manipulating) the nature and quality of subsequent school (and home) environments.

Combining early education and parenting interventions. Future research should also consider the merits of combining early education and parenting programs. This could be a promising approach if programs complement one another. However, as discussed in Sections 5 and 6, the evidence so far on this approach is weak.

Understanding parents' decisions. Research summarized in this chapter suggests that stress may be an important constraint on parents' attention and decision-making. Future research should focus on collecting better data on parental beliefs and constraints, enabling us to understand why parents make the choices they make and how they respond to interventions.

References:

Ainsworth, M. D. S., Bell, S. M., & Stayton, D. F. (1974). Infant-mother attachment and social development: Socialization as a product of reciprocal responsiveness to signals.

Aizer, A., Eli, S., Ferrie, J., & Lleras-Muney, A. (2016). The long-run impact of cash transfers to poor families. *American Economic Review, 106*(4), 935-971.

Akee, R. K., Copeland, W. E., Keeler, G., Angold, A., & Costello, E. J. (2010). Parents' incomes and children's outcomes: a quasi-experiment using transfer payments from casino profits. *American Economic Journal: Applied Economics, 2*(1), 86-115.

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics *Handbook of the Economics of Education* (Vol. 4, pp. 1-181): Elsevier.

Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population. *Journal of Political Economy, 114*(4), 672-712.

Alwin, D. F. (2001). Parental values, beliefs, and behavior: A review and promulga for research into the new century. *Advances in Life Course Research, 6*, 97-139.

Andreoni, J., & Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review, 102*(7), 3333-3356.

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics, 131*(3), 1415-1453.

Athey, S., Chetty, R., Imbens, G. W., & Kang, H. (2019). *The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely*. Retrieved from

Attanasio, O., Boneva, T., & Rauh, C. (2019). *Parental beliefs about returns to different types of investments in school children*. Retrieved from

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest, 21*(2), 55-97.

Baker, A. J., Piotrkowski, C. S., & Brooks-Gunn, J. (1998). The effects of the Home Instruction Program for Preschool Youngsters (HIPPY) on children's school performance at the end of the program and one year later. *Early Childhood Research Quarterly, 13*(4), 571-588.

Baker, A. J., Piotrkowski, C. S., & Brooks-Gunn, J. (1999). The home instruction program for preschool youngsters (HIPPY). *The Future of children*, 116-133.

Baker, M., Gruber, J., & Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy, 116*(4), 709-745.

Baker, M., Gruber, J., & Milligan, K. (2019). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy, 11*(3), 1-26.

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education, 18*(2), 141-157.

Bassok, D., Finch, J. E., Lee, R., Reardon, S. F., & Waldfogel, J. (2016). Socioeconomic gaps in early childhood experiences: 1998 to 2010. *Aera Open, 2*(3), 2332858416653924.

Bauchmüller, R., Gørtz, M., & Rasmussen, A. W. (2014). Long-run benefits from universal high-quality preschooling. *Early Childhood Research Quarterly, 29*(4), 457-470.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal, 75*(299), 493-517.

Becker, G. S. (1981). Altruism in the Family and Selfishness in the Market Place. *Economica, 48*(189), 1-15.

Becker, G. S., & Mulligan, C. B. (1997). The endogenous determination of time preference. *The Quarterly Journal of Economics, 112*(3), 729-758.

Behrman, J. R., & Taubman, P. (1989). Is schooling" mostly in the genes"? Nature-nurture decomposition using data on relatives. *Journal of Political Economy, 97*(6), 1425-1446.

Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Age, G. E. S.-R. S., . . . Purcell, S. (2012). The promises and pitfalls of genoeconomics. *Annu. Rev. Econ., 4*(1), 627-662.

Bennett, J., & Tayler, C. P. (2006). *Starting strong II: Early childhood education and care*: OECD.

Berger, A. (2011). *Self-regulation: Brain, Cognition, and Development*. Washington, DC: American Psychological Association.

Berger, E. M., Fehr, E., Hermes, H., Schunk, D., & Winkel, K. (2020). The impact of working memory training on children's cognitive and noncognitive skills. *NHH Dept. of Economics Discussion Paper*(09).

Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of public Economics, 93*(1-2), 219-234.

Berlinski, S., Galiani, S., & Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of public Economics, 92*(5-6), 1416-1440.

Bettinger, E., Hægeland, T., & Rege, M. (2014). Home with mom: the effects of stay-at-home parents on children's long-run educational outcomes. *Journal of Labor Economics, 32*(3), 443-467.

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., . . . Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79*(6), 1802-1817.

Bjorklund, A., Jantti, M., & Solon, G. (2005). Influences of nature and nurture on earnings variation. *Unequal chances: Family background and economic success*, 145-164.

Björklund, A., & Jäntti, M. (2009). Intergenerational income mobility and the role of family background. *Oxford handbook of economic inequality, 491*, 521.

Björklund, A., & Jäntti, M. (2012). How important is family background for labor-economic outcomes? *Labour Economics, 19*(4), 465-474.

Björklund, A., & Jäntti, M. (2020). Intergenerational mobility, intergenerational effects, sibling correlations, and equality of opportunity: a comparison of four approaches. *Research in Social Stratification and Mobility*, 100455.

Björklund, A., Jäntti, M., & Lindquist, M. J. (2009). Family background and income during the rise of the welfare state: brother correlations in income for Swedish men born 1932–1968. *Journal of public Economics, 93*(5-6), 671-680.

Björklund, A., Lindahl, M., & Plug, E. (2006). The origins of intergenerational associations: Lessons from Swedish adoption data. *The Quarterly Journal of Economics, 121*(3), 999-1028.

Björklund, A., & Moffitt, R. (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics*, 42-49.

Black, S. E., Devereux, P. J., Lundborg, P., & Majlesi, K. (2020). Poor little rich kids? The role of nature versus nurture in wealth and other economic outcomes and behaviours. *The Review of Economic Studies, 87*(4), 1683-1725.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and psychopathology, 20*(3), 899-911.

Blair, C., & Raver, C. C. (2015). School Readiness and Self-Regulation: A Developmental Psychobiological Approach. *Annual Review of Psychology, 66*(1), 711-731. doi:doi:10.1146/annurev-psych-010814-015221

Blau, D., & Currie, J. (2006). Pre-school, day care, and after-school care: who's minding the kids? *Handbook of the Economics of Education, 2*, 1163-1278.

Blau, D. M. (1999). The effect of child care characteristics on child development. *Journal of Human Resources*, 786-822.

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: a meta-analysis of the leading behavioral contributors to mortality. *Psychological bulletin, 130*(6), 887.

Bowlby, J. (1973). Attachment and loss: Volume II: Separation, anxiety and anger *Attachment and loss: Volume II: Separation, anxiety and anger* (pp. 1-429): London: The Hogarth press and the institute of psycho-analysis.

Bradley, R. H., & Caldwell, B. M. (1995). Caregiving and the regulation of child growth and development: Describing proximal aspects of caregiving systems. *Developmental Review, 15*(1), 38-85.

Brandlistuen, R. E., Flatø, M., Stoltenberg, C., Helland, S. S., & Wang, M. V. (2021). Gender gaps in preschool age: A study of behavior, neurodevelopment and pre-academic skills. *Scandinavian Journal of Public Health, 49*(5), 503-510.

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social forces, 92*(1), 109-140.

Bredtmann, J., & Smith, N. (2018). Inequalities in educational outcomes: how important is the family? *Oxford Bulletin of Economics and Statistics, 80*(6), 1117-1144.

Brekke, E. S., Lundetræ, K., & Reikerås, E. (2018). Gender differences in toddlers' language and participation in language activities in Norwegian ECEC institutions. *European Early Childhood Education Research Journal, 26*(3), 375-392.

Buhs, E. S., Ladd, G. W., & Herald, S. L. (2006). Peer Exclusion and Victimization: Processes That Mediate the Relation Between Peer Group Rejection and Children's Classroom Engagement and Achievement? *Journal of educational psychology, 98*(1), 1-13.

Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives, 12*(1), 3-9.

Burchinal, M. C., & Farran, D. C. (2020). What does research tell us about ECE programs. Foundation for Child Development. *Getting It Right: Using Implementation Research to Improve Outcomes in Early Care and Education, 13-36.*

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. *Developmental Psychology, 37*(2), 231.

Cappelen, A., List, J., Samek, A., & Tungodden, B. (2020). The effect of early-childhood education on social preferences. *Journal of Political Economy, 128*(7), 2739-2758.

Carlson, D., Haveman, R., Kaplan, T., & Wolfe, B. (2012a). Long-term earnings and employment effects of housing voucher receipt. *Journal of Urban Economics, 71*(1), 128-150.

Carlson, D., Haveman, R., Kaplan, T., & Wolfe, B. (2012b). Long-term effects of public low-income housing vouchers on neighborhood quality and household composition. *Journal of Housing Economics, 21*(2), 101-120.

Carneiro, P., García, I. L., Salvanes, K. G., & Tominey, E. (2021). Intergenerational mobility and the timing of parental income. *Journal of Political Economy, 129*(3), 757-788.

Carneiro, P., Løken, K. V., & Salvanes, K. G. (2015). A flying start? Maternity leave benefits and long-run outcomes of children. *Journal of Political Economy, 123*(2), 365-412.

Cascio, E. U., & Schanzenbach, D. W. (2013). The Impacts of Expanding Access to High-Quality Preschool Education. *Brookings Papers ON ECONOMIC ACTIVITY, 2013*(2), 127-192.

Cawley, J., Conneely, K., Heckman, J., & Vytlacil, E. (1997). Cognitive ability, wages, and meritocracy *Intelligence, genes, and success* (pp. 179-192): Springer.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: when the" have-nots" gain but the" haves" gain even more. *American Psychologist, 60*(2), 149.

Chazan-Cohen, R., Raikes, H. H., & Vogel, C. (2013). V. Program Subgroups: Patterns Of Impacts For Home-Based, Center-Based, And Mixed-Approach Programs. *Monographs of the Society for Research in Child Development, 78*(1), 93-109.

Chetty, R., Friedman, J. N., & Rockoff, J. (2011). *New evidence on the long-term impacts of tax credits.* Paper presented at the Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review, 106*(4), 855-902.

Child, N. S. C. o. t. D. (2010). *Early Experiences Can Alter Gene Expression and Affect Long-Term Development. Working Paper# 10*.

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal, 50*(4), 812-850.

Conti, G., Heckman, J. J., & Pinto, R. (2016). The effects of two influential early childhood interventions on health and healthy behaviour. *The Economic Journal, 126*(596), F28-F65.
Cooper, K., & Stewart, K. (2020). Does household income affect children's outcomes? A systematic review of the evidence. . *Child Indicators Research, 1-25.*
Cornelissen, T., Dustmann, C., Raute, A., & Schönberg, U. (2018). Who benefits from universal child care? Estimating marginal returns to early child care attendance. *Journal of Political Economy, 126*(6), 2356-2409.
Costello, E. J., Compton, S. N., Keeler, G., & Angold, A. (2003). Relationships between poverty and psychopathology: A natural experiment. *Jama, 290*(15), 2023-2029.
Costello, E. J., Erkanli, A., Copeland, W., & Angold, A. (2010). Association of family income supplements in adolescence with development of psychiatric and substance use disorders in adulthood among an American Indian population. *Jama, 303*(19), 1954-1960.
Cunha, F., Elo, I., & Culhane, J. (2013). *Eliciting maternal expectations about the technology of cognitive skill formation*. Retrieved from
Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review, 97*(2), 31-47.
Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education, 1*, 697-812.
Currie, J., & Neidell, M. (2007). Getting inside the "black box" of Head Start quality: What matters and what doesn't. *Economics of Education review, 26*(1), 83-99.
Currie, J., & Yelowitz, A. (2000). Are public housing projects good for kids? *Journal of public Economics, 75*(1), 99-124.
Dahl, G. B., & Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review, 102*(5), 1927-1956.
Dahl, G. B., Løken, K. V., Mogstad, M., & Salvanes, K. V. (2016). What is the case for paid maternity leave? *Review of Economics and Statistics, 98*(4), 655-670.
Danzer, N., & Lavy, V. (2018). Paid parental leave and children's schooling outcomes. *The Economic Journal, 128*(608), 81-117.
Darling, N., & Steinberg, L. (1993). Parenting style as context: An integrative model. *Psychological bulletin, 113*(3), 487.
Del Boca, D., Flinn, C., & Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies, 81*(1), 137-185.
Del Boca, D., Flinn, C., & Wiswall, M. (2016). Transfers to households with children and child development. *The Economic Journal, 126*(596), F136-F183.
Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1*(3), 111-134.
Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics, 132*(4), 1593-1640.
DeVries, R., & Kohlberg, L. (1987). *Programs of early education: The constructivist view*: Longman Publishing Group.
Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168.
DiPrete, T. A., & Jennings, J. L. (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research, 41*(1), 1-15.
Doepke, M., & Zilibotti, F. (2019). *Love, money, and parenting*: Princeton University Press.
Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review, 100*(3), 1238-1260.
Doyle, O. (2020). The first 2,000 days and child skills. *Journal of Political Economy, 128*(6), 2067-2122.
Drange, N., & Havnes, T. (2019). Early childcare and cognitive development: Evidence from an assignment lottery. *Journal of Labor Economics, 37*(2), 581-620.
Drange, N., & Rønning, M. (2017). *Child care center staff composition and early child development*. Retrieved from

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. OPRE Report 2018-76A. *Office of Planning, Research and Evaluation*.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428.

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

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

Duncan, G. J., & Murnane, R. J. (2014). *Restoring opportunity: The crisis of inequality and the challenge for American education*: Harvard Education Press.

Duncan, G. J., Ziol-Guest, K. M., & Kalil, A. (2010). Early-childhood poverty and adult attainment, behavior, and health. *Child Development, 81*(1), 306-325.

Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology*.

Dustmann, C., & Schönberg, U. (2012). Expansions in maternity leave coverage and children's long-term outcomes. *American Economic Journal: Applied Economics, 4*(3), 190-224.

Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., . . . Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly, 21*(2), 174-195.

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D.-M. C., Crawford, G. M., . . . Howes, C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 25*(2), 177-193.

Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., . . . Guthrie, I. K. (2001). The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development, 72*(4), 1112-1134.

Elder Jr, G. H. (1998). The life course as developmental theory. *Child Development, 69*(1), 1-12.

Engel, A., Barnett, W. S., Anders, Y., & Taguma, M. (2015). Early childhood education and care policy review.

Engzell, P., & Tropf, F. C. (2019). Heritability of education rises with intergenerational mobility. *Proceedings of the National Academy of Sciences, 116*(51), 25386-25388.

Evans, W. N., & Garthwaite, C. L. (2014). Giving mom a break: The impact of higher EITC payments on maternal health. *American Economic Journal: Economic Policy, 6*(2), 258-290.

Fagereng, A., Mogstad, M., & Rønning, M. (2021). Why do wealthy parents have wealthy children? *Journal of Political Economy, 129*(3), 703-756.

Feinberg, M. E., Jones, D. E., Roettger, M. E., Solmeyer, A., & Hostetler, M. L. (2014). Long-term follow-up of a randomized trial of family foundations: Effects on children's emotional, behavioral, and school adjustment. *Journal of family psychology, 28*(6), 821.

Feinberg, M. E., Kan, M. L., & Goslin, M. C. (2009). Enhancing coparenting, parenting, and child self-regulation: Effects of family foundations 1 year after birth. *Prevention Science, 10*(3), 276-285.

Felfe, C., & Lalive, R. (2018). Does early child care affect children's development? *Journal of public Economics, 159*, 33-53.

Felfe, C., Nollenberger, N., & Rodríguez-Planas, N. (2015). Can't buy mommy's love? Universal childcare and children's long-term cognitive development. *Journal of population economics, 28*(2), 393-422.

Felfe, C., & Zierow, L. (2018). From dawn till dusk: Implications of full-day care for children's development. *Labour Economics, 55*, 259-281.
Fidjeland, A., Rege, M., Solli, I. F., & Størksen, I. (2022). Reducing the Gender Gap in Early Learning: Evidence from a Field Experiment in Norwegian Preschools. *University of Stavanger, Manuscript*.
Fiorini, M., & Keane, M. P. (2014). How the allocation of children's time affects cognitive and noncognitive development. *Journal of Labor Economics, 32*(4), 787-836.
Fitzpatrick, M. D. (2008). Starting school at four: The effect of universal pre-kindergarten on children's academic achievement. *The BE Journal of Economic Analysis & Policy, 8*(1).
Fort, M., Ichino, A., & Zanella, G. (2020). Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families. *Journal of Political Economy, 128*(1), 158-205.
García, J. L., Heckman, J. J., Leaf, D. E., & Prados, M. J. (2020). Quantifying the life-cycle benefits of an influential early-childhood program. *Journal of Political Economy, 128*(7), 2502-2541.
Gassman-Pines, A. (2011). Low-income mothers' nighttime and weekend work: Daily associations with child behavior, mother-child interactions, and mood. *Family Relations, 60*(1), 15-29.
Gelber, A., & Isen, A. (2013). Children's schooling and parents' behavior: Evidence from the Head Start Impact Study. *Journal of public Economics, 101*, 25-38.
Gennetian, L. A., Sanbonmatsu, L., Katz, L. F., Kling, J. R., Sciandra, M., Ludwig, J., . . . Kessler, R. C. (2012). The long-term effects of Moving to Opportunity on youth outcomes. *Cityscape*, 137-167.
Gershoff, E. T., Aber, J. L., Raver, C. C., & Lennon, M. C. (2007). Income is not enough: Incorporating material hardship into models of income associations with parenting and child development. *Child Development, 78*(1), 70-95.
Gershoff, E. T., & Grogan-Kaylor, A. (2016). Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of family psychology, 30*(4), 453.
Gibbs, C., Ludwig, J., & Miller, D. L. (2011). *Does Head Start do any lasting good?* Retrieved from
Gigerenzer, G., & Selten, R. (2001). Rethinking rationality. *Bounded rationality: The adaptive toolbox, 1*, 12.
Goldberger, A. S. (1979). Heritability. *Economica, 46*(184), 327-347.
Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma an evaluation of Tulsa's pre-k program. *Journal of Human Resources, 40*(3), 533-558.
Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*(1), 79-132.
Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2021). *The Long-Term Effects of Universal Preschool in Boston*. Retrieved from
Grindal, T., Bowne, J. B., Yoshikawa, H., Schindler, H. S., Duncan, G. J., Magnuson, K., & Shonkoff, J. P. (2016). The added impact of parenting education in early childhood education programs: A meta-analysis. *Children and Youth Services Review, 70*, 238-249.
Gupta, N. D., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of public Economics, 94*(1-2), 30-43.
Guryan, J., Hurst, E., & Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic Perspectives, 22*(3), 23-46.
Gørtz, M., Johansen, E. R., & Simonsen, M. (2018). Academic achievement and the gender composition of preschool staff. *Labour Economics, 55*, 241-258.
Hamre, B. K., Partee, A., & Mulcahy, C. (2017). Enhancing the impact of professional development in the context of preschool expansion. *Aera Open, 3*(4), 2332858417733686.
Hamre, B. K., & Pianta, R. C. (2001). Early Teacher–Child Relationships and the Trajectory of Children's School Outcomes through Eighth Grade. *Child Development, 72*(2), 625-638. doi:10.1111/1467-8624.00301
Hane, A. A., & Fox, N. A. (2006). Ordinary variations in maternal caregiving influence human infants' stress reactivity. *Psychological science, 17*(6), 550-556.
Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale*: ERIC.

Harrison, G. W., Lau, M. I., & Williams, M. B. (2002). Estimating individual discount rates in Denmark: A field experiment. *American Economic Review, 92*(5), 1606-1617.

Haskins, R. (1985). Public school aggression among children with varying day-care experience. *Child Development*, 689-703.

Havnes, T., & Mogstad, M. (2011). No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy, 3*(2), 97-129.

Havnes, T., & Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of public Economics, 127*, 100-114.

Heckman, J., Hohmann, N., Smith, J., & Khoo, M. (2000). Substitution and dropout bias in social experiments: A study of an influential social experiment. *The Quarterly Journal of Economics, 115*(2), 651-694.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review, 103*(6), 2052-2086.

Heckman, J. J. (1995). Lessons from the bell curve. *Journal of Political Economy, 103*(5), 1091-1120.

Heckman, J. J. (2005). The scientific model of causality. *Sociological methodology, 35*(1), 1-97.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of economic literature, 48*(2), 356-398.

Heckman, J. J., & Masterov, D. V. (2007). The productivity argument for investing in young children: National Bureau of Economic Research Cambridge, Mass., USA.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of public Economics, 94*(1-2), 114-128.

Heckman, J. J., & Mosso, S. (2014). The economics of human development and social mobility. *Annu. Rev. Econ., 6*(1), 689-733.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review, 91*(2), 145-149.

Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica, 73*(3), 669-738.

Heckman, J. J., & Vytlacil, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences, 96*(8), 4730-4734.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172-177.

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. (2009). A mandate for playful learning in preschool: Applying the scientific evidence.

Holmlund, H., Lindahl, M., & Plug, E. (2011). The causal effect of parents' schooling on children's schooling: A comparison of estimation methods. *Journal of economic literature, 49*(3), 615-651.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly, 23*(1), 27-50.

Hoynes, H., Schanzenbach, D. W., & Almond, D. (2016). Long-run impacts of childhood access to the safety net. *American Economic Review, 106*(4), 903-934.

Hoynes, H. W., & Patel, A. J. (2018). Effective policy for reducing poverty and inequality? The Earned Income Tax Credit and the distribution of income. *Journal of Human Resources, 53*(4), 859-890.

Hyytinen, A., Ilmakunnas, P., Johansson, E., & Toivanen, O. (2019). Heritability of lifetime earnings. *The Journal of Economic Inequality, 17*(3), 319-335.

Ishizuka, P. (2019). Social class, gender, and contemporary parenting standards in the United States: Evidence from a national survey experiment. *Social forces, 98*(1), 31-58.

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annu. Rev. Econ., 6*(1), 801-825.

Jacob, B. A., Kapustin, M., & Ludwig, J. (2015). The impact of housing assistance on child outcomes: Evidence from a randomized housing lottery. *The Quarterly Journal of Economics, 130*(1), 465-506.

Jenkins, J. M., Duncan, G. J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschool teachers target skills or the whole child? *Economics of Education review, 65*, 107-125.

Jenkins, J. M., Farkas, G., Duncan, G. J., Burchinal, M., & Vandell, D. L. (2016). Head Start at ages 3 and 4 versus Head Start followed by state pre-k: Which is more effective? *Educational Evaluation and Policy Analysis, 38*(1), 88-112.

Jensen, A. R. (1998). The factor.

Johnson, R. C., & Jackson, C. K. (2019). Reducing inequality through dynamic complementarity: Evidence from Head Start and public school spending. *American Economic Journal: Economic Policy, 11*(4), 310-349.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*(3), 621-652.

Kalb, G., & Van Ours, J. C. (2014). Reading to young children: A head-start in life? *Economics of Education review, 40*, 1-24.

Kalil, A., Mayer, S., Delgado, W., & Gennetian, L. (2020). The Education Gradient in Maternal Enjoyment of Time in Childcare. *University of Chicago, Becker Friedman Institute for Economics Working Paper*(2020-131).

Kalil, A., Mayer, S. E., & Gallegos, S. (2021). Using behavioral insights to increase attendance at subsidized preschool programs: The Show Up to Grow Up intervention. *Organizational Behavior and Human Decision Processes*.

Kalil, A., Ryan, R., & Corey, M. (2012). Diverging destinies: Maternal education and the developmental gradient in time with children. *Demography, 49*(4), 1361-1383.

Kalmijn, M. (1998). Intermarriage and homogamy: Causes, patterns, trends. *Annual review of sociology, 24*(1), 395-421.

Kehrer, B. H., & Wolin, C. M. (1979). Impact of income maintenance on low birth weight: evidence from the Gary Experiment. *Journal of Human Resources*, 434-462.

Kessler, R. C., & Cleary, P. D. (1980). Social class and psychological distress. *American Sociological Review*, 463-478.

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., . . . Howes, C. (2013). Preschool center quality and school readiness: Quality effects and variation by demographic and child characteristics. *Child Development, 84*(4), 1171-1190.

Kline, P., & Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of Head Start. *The Quarterly Journal of Economics, 131*(4), 1795-1848.

Kochanska, G. (2002). Committed compliance, moral self, and internalization: a mediational model. *Developmental Psychology, 38*(3), 339.

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., & Falk, A. (2020). The formation of prosociality: causal evidence on the role of social environment. *Journal of Political Economy, 128*(2), 434-467.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547-588.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive psychology, 6*(2), 293-323.

Lawrance, E. C. (1991). Poverty and the rate of time preference: evidence from panel data. *Journal of Political Economy, 99*(1), 54-77.

Lee, J. S., & Ginsburg, H. P. (2009). Early childhood teachers' misconceptions about mathematics education for young children in the United States. *Australasian Journal of Early Childhood, 34*(4), 37-45.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124*(1), 83.

Lenes, R., Gonzales, C. R., Størksen, I., & McClelland, M. M. (2020). Children's self-regulation in Norway and the United States: the role of mother's education and child gender across cultural contexts. *Frontiers in Psychology*, 2563.

Li, W., Duncan, G. J., Magnuson, K., Schindler, H. S., Yoshikawa, H., & Leak, J. (2020). Timing in Early Childhood Education: How Cognitive and Achievement Program Impacts Vary by Starting Age, Program Duration, and Time since the End of the Program. EdWorkingPaper No. 20-201. *Annenberg Institute for School Reform at Brown University*.

Lichand, G., Bettinger, E., Cunha, N., & Madeira, R. (2021). *The psychological effects of poverty on investments in children's human capital*. Retrieved from

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly, 45*, 155-176.

List, J. A. (2022). *The Voltage Effect*: Penguin Random House.

Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., . . . Brady-Smith, 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.

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics, 122*(1), 159-208.

Lundberg, I., Larsman, P., & Strid, A. (2012). Development of phonological awareness during the preschool year: The influence of gender and socio-economic status. *Reading and writing, 25*(2), 305-320.

Lupien, S. J., King, S., Meaney, M. J., & McEwen, B. S. (2001). Can poverty get under your skin? Basal cortisol levels and cognitive function in children from low and high socioeconomic status. *Development and psychopathology, 13*(3), 653-676.

Løken, K. V., Mogstad, M., & Wiswall, M. (2012). What linear estimators miss: The effects of family income on child outcomes. *American Economic Journal: Applied Economics, 4*(2), 1-35.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science, 341*(6149), 976-980.

Manski, C. F. (2011). Genes, eyeglasses, and social policy. *Journal of Economic Perspectives, 25*(4), 83-94.

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (2017). The effects of language-and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly, 38*, 97-115.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., . . . Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732-749.

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of educational psychology, 101*(3), 689.

Maxfield, M. (2013). The effects of the earned income tax credit on child achievement and long-term educational attainment. *Michigan State University Job Market Paper*.

Mayer, S. E., Kalil, A., & Klein, N. (2020). Behavioral insights and parental decision-making *Confronting inequality: How policies and practices shape children's opportunities.* (pp. 99-117): American Psychological Association.

Mayer, S. E., Kalil, A., Oreopoulos, P., & Gallegos, S. (2019). Using behavioral insights to increase parental engagement the parents and children together intervention. *Journal of Human Resources, 54*(4), 900-925.

Maynard, R. A. (1977). The effects of the rural income maintenance experiment on the school performance of children. *The American Economic Review, 67*(1), 370-375.

Maynard, R. A., & Murnane, R. J. (1979). The effects of a negative income tax on school performance: Results of an experiment. *Journal of Human Resources*, 463-476.

Mazumder, B. (2008). Sibling similarities and economic inequality in the US. *Journal of population economics, 21*(3), 685-701.

McClelland, M. M., Cameron, C. E., McDonald Connor, C., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links Between Behavioral Regulation and Preschoolers' Literacy, Vocabulary, and Math Skills. *Developmental Psychology, 43*(4), 947-959.

McCoby, E. (1983). Socialization in the context of the family: Parent-child interaction. *Handbook of child psychology, 4*, 1-101.

McCoy, D. C., Gonzalez, K., & Jones, S. (2019). Preschool Self-Regulation and Preacademic Skills as Mediators of the Long-Term Impacts of an Early Intervention. *Child Development, 90*(5), 1544-1558.

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., . . . Shonkoff, J. P. (2017). Impacts of early childhood education on medium-and long-term educational outcomes. *Educational Researcher, 46*(8), 474-487.

McFarlane, E., Burrell, L., Duggan, A., & Tandon, D. (2017). Outcomes of a randomized trial of a cognitive behavioral enhancement to address maternal distress in home visited mothers. *Maternal and Child Health Journal, 21*(3), 475-484.

McLeod, J. D., & Kessler, R. C. (1990). Socioeconomic status differences in vulnerability to undesirable life events. *Journal of health and social behavior*, 162-172.

McLoyd, V. C. (1990). The impact of economic hardship on Black families and children: Psychological distress, parenting, and socioemotional development. *Child Development, 61*(2), 311-346.

Michalopoulos, C., Faucetta, K., Hill, C. J., Portilla, X., Burrell, L., Lee, H., & Knox, V. (2019). Impacts on family outcomes of evidence-based early childhood home visiting: Results from the mother and infant home visiting program evaluation. *Research, and Evaluation, Administration for Children and Families, US. DHHS.[Google Scholar]*.

Miller, G. E., Chen, E., & Parker, K. J. (2011). Psychological stress in childhood and susceptibility to the chronic diseases of aging: moving toward a model of behavioral and biological mechanisms. *Psychological bulletin, 137*(6), 959.

Miller, P., Mulvey, C., & Martin, N. (2001). Genetic and environmental contributions to educational attainment in Australia. *Economics of Education review, 20*(3), 211-224.

Mogstad, M., & Torsvik, G. (2021). Family Background, Neighborhoods and Intergenerational Mobility.

Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly, 21*(3), 313-331.

Morris, P., Mattera, S., Castells, N., Bangser, M., Bierman, K., & Raver, C. C. (2014). Impact findings from the Head Start CARES demonstration: National evaluation of three approaches to improving preschoolers' social and emotional competence. *Available at SSRN 2477974*.

Mullainathan, S., & Shafir, E. (2013). Scarcity. *Social Policy, 46*(2), 231-249.

Mytton, J., Ingram, J., Manns, S., & Thomas, J. (2014). Facilitators and barriers to engagement in parenting programs: a qualitative systematic review. *Health Education & Behavior, 41*(2), 127-137.

NASEM. (2019). *Environmental Engineering for the 21st Century: Addressing Grand Challenges*. Washington, DC: The National Academies Press.

National Scientific Council on the Developing Child. (2010). *Persistent Fear and Anxiety Can Affect Young Children's Learning and Development*: Harvard University, Center on the Developing Child.

Nicolaou, N., Shane, S., Cherkas, L., Hunkin, J., & Spector, T. D. (2008). Is the tendency to engage in entrepreneurship genetic? *Management Science, 54*(1), 167-179.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *American Psychologist, 67*(2), 130.

O'Donoghue, T., & Rabin, M. (2015). Present bias: Lessons learned and to be learned. *American Economic Review, 105*(5), 273-279.

Olds, D. L., Henderson Jr, C. R., Tatelbaum, R., & Chamberlin, R. (1988). Improving the life-course development of socially disadvantaged mothers: a randomized trial of nurse home visitation. *American journal of public health, 78*(11), 1436-1445.

Pages, R., Lukes, D. J., Bailey, D. H., & Duncan, G. J. (2020). Elusive longer-run impacts of head start: replications within and across cohorts. *Educational Evaluation and Policy Analysis, 42*(4), 471-492.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading research quarterly, 40*(2), 184-202.

Pekkarinen, T., Salvanes, K. G., & Sarvimäki, M. (2017). The evolution of social mobility: Norway during the twentieth century. *The Scandinavian Journal of Economics, 119*(1), 5-33.

Piaget, J. (1976). Piaget's theory *Piaget and his school* (pp. 11-23): Springer.

Pianta, R., Hamre, B., Downer, J., Burchinal, M., Williford, A., Locasale-Crouch, J., . . . Scott-Little, C. (2017). Early childhood professional development: Coaching and coursework effects on indicators of children's school readiness. *Early Education and Development, 28*(8), 956-975.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*: Paul H Brookes Publishing.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological bulletin, 84*(2), 309.

Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics, 47*(7), 702-709.

Poropat, A. E. (2014). Other-rated personality and academic performance: Evidence and implications. *Learning and Individual Differences, 34*, 24-32.

Price, J., & Kalil, A. (2019). The effect of mother–child reading time on children's reading skills: Evidence from natural within-family variation. *Child Development, 90*(6), e688-e702.

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., . . . Friedman, J. (2010). Head Start Impact Study. Final Report. *Administration for Children & Families*.

Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., & Ramey, S. L. (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied developmental science, 4*(1), 2-14.

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' preacademic skills: self-regulation as a mediating mechanism. *Child Development, 82*(1), 362-378.

Rege, M., Solli, I. F., Størksen, I., & Votruba, M. (2018). Variation in center quality in a universal publicly subsidized and regulated childcare system. *Labour Economics, 55*, 230-240.

Rege, M., Størksen, I., Solli, I. F., Kalil, A., McClelland, M. M., ten Braak, D., . . . Carlsen, M. (2022). The Effects of a Structured Curriculum on Preschool Effectiveness: A Field Experiment. *Journal of Human Resources, forthcoming*, 0220-10749R10743.

Ridley, M., & Pierpoint, G. (2003). *Nature via nurture: Genes, experience, and what makes us human* (Vol. 19): HarperCollins New York.

Rojas, N. M., Morris, P., & Balaraman, A. (2020). Finding Rigor Within a Large-Scale Expansion of Preschool to Test Impacts of a Professional Development Program. *Aera Open, 6*(4), 2332858420975399.

Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology, 46*(5), 995-1007. doi:10.1037/a0018880

Rossin-Slater, M., & Wüst, M. (2020). What is the added value of preschool for poor children? Long-term and intergenerational impacts and interactions with an infant health intervention. *American Economic Journal: Applied Economics, 12*(3), 255-286.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of child language, 35*(1), 185-205.

Rowe, M. L. (2018). Understanding socioeconomic differences in parents' speech to children. *Child Development Perspectives, 12*(2), 122-127.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2007). Gender Development. *Handbook of child psychology, 3*.

Ryan, R. M., Kalil, A., Hines, C., & Ziol-Guest, K. (2020). Trends in Parental Values in a Period of US Labor Market Change. *Journal of Marriage and Family, 82*(5), 1495-1514.

Ryan, R. M., & Padilla, C. (2019). Public policy and family psychology *APA handbook of contemporary family psychology: Applications and broad impact of family psychology, Vol. 2* (pp. 639-655): American Psychological Association.

Sacerdote, B. (2007). How large are the effects from changes in family environment? A study of Korean American adoptees. *The Quarterly Journal of Economics, 122*(1), 119-157.

Sacerdote, B. (2011). Nature and Nurture Effects on Children's Outcomes: What Have We Learned from Studies of Twins and Adoptees? *Handbook of social economics* (Vol. 1, pp. 1-30): Elsevier.

Salkind, N. J., & Haskins, R. (1982). Negative income tax: The impact on children from low-income families. *Journal of Family Issues, 3*(2), 165-180.

Schnitzlein, D. D. (2014). How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark. *Journal of population economics, 27*(1), 69-89.

Schweinhart, L. J. (1993). *Significant Benefits: The High/Scope Perry Preschool Study through Age 27. Monographs of the High/Scope Educational Research Foundation, No. Ten*: ERIC.

Sojourner, A. (2013). Identification of peer effects with missing peer data: Evidence from Project STAR. *The Economic Journal, 123*(569), 574-605.

Solon, G. (1999). Intergenerational mobility in the labor market *Handbook of labor economics* (Vol. 3, pp. 1761-1800): Elsevier.

Solon, G., Corcoran, M., Gordon, R., & Laren, D. (1991). A longitudinal analysis of sibling correlations in economic status. *Journal of Human Resources*, 509-534.

Steinberg, L., Lamborn, S. D., Darling, N., Mounts, N. S., & Dornbusch, S. M. (1994). Over-time changes in adjustment and competence among adolescents from authoritative, authoritarian, indulgent, and neglectful families. *Child Development, 65*(3), 754-770.

Størksen, I., Ellingsen, I. T., Wanless, S. B., & McClelland, M. M. (2015). The influence of parental socioeconomic background and gender on self-regulation among 5-year-old children in Norway. *Early Education and Development, 26*(5-6), 663-684.

Størksen, I., Ertesvåg, S. K., & Rege, M. (2021). Implementing implementation science in a randomized controlled trial in Norwegian early childhood education and care. *International Journal of Educational Research, 108*, 101782.

Størksen, I., Ten Braak, D., Breive, S., Lenes, R., Lunde, S., Carlsen, M., & Rege, M. (2018). Lekbasert læring–et forskningsbasert førskoleopplegg fra Agderprosjektet. *GAN Aschehoug*.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*(3), 748-767.

Tandon, S. D., Ward, E. A., Hamil, J. L., Jimenez, C., & Carter, M. (2018). Perinatal depression prevention through home visitation: a cluster randomized trial of mothers and babies 1-on-1. *Journal of Behavioral Medicine, 41*(5), 641-652.

Taubman, P. (1976). The determinants of earnings: Genetics, family, and other environments: A study of white male twins. *The American Economic Review, 66*(5), 858-870.

Thijssen, M. W. P., Rege, M., Solli, I. F., & Størksen, I. (2022). Cross-Productivities of Executive Functions. *University of Stavanger, Manuscript*.

Tonyan, H. A., & Howes, C. (2003). Exploring patterns in time children spend in a variety of child care activities: Associations with environmental quality, ethnicity, and gender. *Early Childhood Research Quarterly, 18*(1), 121-142.

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., . . . Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence, 42*, 83-88.

Turner, R. J., & Avison, W. R. (2003). Status variations in stress exposure: Implications for the interpretation of research on race, socioeconomic status, and gender. *Journal of health and social behavior*, 488-505.

Van Ijzendoorn, M. H. (1992). Intergenerational transmission of parenting: A review of studies in nonclinical populations. *Developmental Review, 12*(1), 76-99.

Venti, S. F. (1984). The effects of income maintenance on work, schooling, and non-market activities of youth. *The Review of Economics and Statistics*, 16-25.

Vygotsky, L. S. (1967). Play and its role in the mental development of the child. *Soviet psychology, 5*(3), 6-18.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*: Harvard university press.

Waldfogel, J., & Washbrook, E. (2011). Early years policy. *Child development research, 2011*.

Walters, C. R. (2015). Inputs in the production of early childhood human capital: Evidence from Head Start. *American Economic Journal: Applied Economics, 7*(4), 76-102.

Watts, T. W., Duncan, G. J., & Rivas, M. (2019). A Reanalysis of Impacts of the Tennessee Voluntary Prekindergarten Program. *EdWorkingPaper: 19-28*.

Weikart, D., & Schweinhart, L. (1987). The High/Scope cognitively oriented curriculum in early education. *Approaches to early childhood education*, 253-268.

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly, 28*(2), 199-209.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112-2130.

Werner, C. D., Linting, M., Vermeer, H. J., & Van IJzendoorn, M. H. (2016). Do intervention programs in child care promote the quality of caregiver-child interactions? A meta-analysis of randomized controlled trials. *Prevention Science, 17*(2), 259-273.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 27*(1), 122-154.

Yamaguchi, S., Asai, Y., & Kambayashi, R. (2018). Effects of subsidized childcare on mothers' labor supply under a rationing mechanism. *Labour Economics, 55*, 1-17.

York, B. N., Loeb, S., & Doss, C. (2019). One step at a time the effects of an early literacy text-messaging program for parents of preschoolers. *Journal of Human Resources, 54*(3), 537-566.

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., . . . D'Sa, N. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology, 51*(3), 309.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., . . . Zaslow, M. J. (2013). Investing in Our Future: The Evidence Base on Preschool Education. *Society for Research in Child Development*.