Estimating the Consequences of Norway's National Scale-Up of Early Childhood Education and Care (Beginning in Infancy) for Early Language Skills

Eric Dearing

Boston College

Norwegian Center for Child Behavioral Development

Henrik Daae Zachrisson

Norwegian Center for Child Behavioral Development

University of Oslo

Arnstein Mykletun

University of Tromsø

Norwegian Institute of Public Health

Haukeland University Hospital

Nordland Hospital Trust

Claudio O. Toppelberg

Judge Baker Children's Center

Harvard Medical School

*While most early childhood education and care (ECEC) programs taken to scale in the United States have served socially disadvantaged 3- to 5-years-olds, Norway scaled up universal ECEC from age 1. We investigated the consequences of Norway's universal ECEC scale-up for children's early language skills, exploiting variation in ECEC coverage across birth cohorts and municipalities in a population-based sample (*n = 63,350). Estimates from two-stage least squares (i.e., instrumental variable) regression and generalized difference-in-differences models indicated the scale-up of universal ECEC led to improved language outcomes, particularly for low-income children. As preschool programs at scale become increasingly common in the United States, our results from Norway help inform debate about the merits of universal versus targeted policies and should provoke discussion about the benefits of beginning ECEC programs as early as infancy.*

Keywords: *Child Development, Early Childhood, Econometric Analysis, Educational Policy, Longitudinal Studies, Policy Analysis, Poverty*

With scientists and policy decision makers lauding the promise of early childhood education and care (ECEC) for raising the achievement of disadvantaged children, there remains concern about whether large-scale, public-funded ECEC programs can replicate the results of small-scale interventions (European Commission, 2011; Farran, 2016; Magnuson & Shager, 2010; Obama, 2014; Organisation for Economic Co-operation and Development [OECD], 2006; Yoshikawa et al., 2013). There are also questions (Barnett, 2010, 2011) about whether scale-up should be universal (i.e., designed to serve all children) or targeted (i.e., exclusively serving disadvantaged children). To date,

scale-up efforts in the United States have largely been targeted programs for disadvantaged preschoolers (3- to 5-year-olds), but some nations now provide publicly funded, universal ECEC beginning in infancy. Norway is a case in point, having taken universal ECEC, beginning at age 1, to a national scale. As a complement to growing evidence about scale-up in the United States, we investigated the consequences of Norway's scale-up of ECEC for children's early language skills. In doing so, we examined whether the universal scale-up had differential consequences for children from low-income versus middle- and high-income families.

Preschool ECEC and the Achievement of Socially Disadvantaged Children

Given evidence that social inequality in achievement may be largely attributed to limited developmental stimulation in the home (e.g., Dearing, 2014; Hackman, Farah, & Meaney, 2010; Hoff, 2006; Shonkoff & Phillips, 2000; Yoshikawa et al., 2012), researchers across multiple disciplines have argued that high-quality ECEC may serve a compensatory function for children growing up in poverty (e.g., Dearing, McCartney, & Taylor, 2009; Magnuson & Shager, 2010). There are, in fact, a relatively large number of experimental and quasi-experimental studies addressing the effects of preschool attendance (typically for children age 3 or older) on achievement and meta-analyses of the results. For example, two meta-analyses of experimental and quasi-experimental preschool interventions targeting disadvantaged children in the United States demonstrate short-term and longer term achievement gains with effect sizes (Hedge's g) of about .20 (Camilli, Vargas, Ryan, & Barnett, 2010; Duncan & Magnuson, 2013), although often with some diminishing of the effects as children get older (Barnett, 2011).

Despite this evidence, it remains less clear whether (or in what direction) universal ECEC policies may affect achievement gaps between more and less advantaged children. To the extent that enrichment in one early learning environment compensates for deprivation in another (Sameroff & Chandler, 1975), universal ECEC coverage may narrow achievement gaps. Yet to the extent that enrichment in ECEC programs complements enrichment at home (Ceci & Papierno, 2005), achievement gaps between socially advantaged and disadvantaged children may be widened by universal provision. In an evaluation of a state-funded universal preschool program in the United States, ECEC had the largest achievement benefits for poor children but also important benefits for near-poor and middle-class children (Gormley, Gayer, Phillips, & Dawson, 2005; Gormley, Phillips, & Gayer, 2008). Both of these studies used regression discontinuity designs based on cut-off birth dates accounting for program admission to address selection bias.

In a meta-analysis of universal preschool program evaluations using quasi-experimental designs, there was mixed evidence, with the most consistent effects evident for disadvantaged children (van Huizen & Platenga, 2015). There is also evidence from Norway, where the present study was conducted, that income inequality among young adults was reduced by an increase in preschool coverage in the late 1970s, when these adults were between 3 and 7 years of age (Havnes and Mogstad, 2011). In this article, the authors used a difference-in-differences (DID) design based on variation in enrollment rates pre– and post–preschool reforms to identify the causal effect.

ECEC for Infants and Toddlers

Compared to the evaluations of ECEC programs for preschool-aged children, there is less evidence about the effectiveness of such programs for infants and toddlers. Yet the potential benefits of ECEC for infants and toddlers is of interest given the consequences of learning stimulation and environmental experiences for brain growth during the first 2 to 3 years of life (Fox & Rutter, 2010; Mustard, 2006; Nelson & Sheridan, 2011; Shonkoff & Phillips, 2000). And observational studies do, in fact, show that ECEC in the earliest years (0–3 years) can promote language and cognitive development. For example, in one U.S. study, more time in center care between 6 and 36 months of age predicted better language and school readiness scores at 3 years, controlling for background factors and child care quality (National Institute of Child Health and Human Development [NICHD], 2000). Also, in nationally representative U.S. data, Loeb, Bridges, Bassok, Fuller, and Rumberger (2007) found that math and reading achievement (measured at the start of kindergarten) was higher for children who entered center care prior to kindergarten, with effect sizes largest (Cohen's *d* approximately .11) for those who entered between ages 2 and 3 compared to earlier or later.

Within this line of nonexperimental work, there is also some evidence that high-quality child care is most strongly and positively associated with language skills and school readiness for low-income 3-year-olds (McCartney, Dearing, Taylor, & Bub, 2007), thereby reducing income-related achievement gaps in math and literacy through the elementary school years (Dearing et al., 2009). Yet larger ECEC effects for low-income children are not found consistently. Ruzek and colleagues (Ruzek, Burchinal, Farkas, & Duncan, 2014), for example, found higher quality infant and toddler ECEC was associated with better cognitive outcomes at 2 years, but not more so for low-income children than for others.

The potential in the early years for ECEC to narrow achievement gaps has also received increasing attention outside of the United States. In nonexperimental studies from Canada, ECEC attendance prior to 4 years predicted a narrowing gap in early language skills and academic readiness between children of low versus high socioeconomic status (Geoffroy et al., 2007; Geoffroy et al., 2010); children of mothers with low education who attended formal child care scored between 36% and 87% of a standard deviation higher on a range of achievement and cognitive tests after school entry, compared to their counterparts not attending ECEC. In a large U.K. sample, using a similar methodology, Cote, Doyle, Petitclerc, and Timmins (2013) found that center-based care (at 9 months) was associated with better cognitive skills at age 3 and 5 (about 10% of a standard deviation compared to attending other types of care) and was more than twice as strongly so associated for children of mothers

with low education. Yet this interaction was not evident for low family income, and the association was inconsistent at ages 5 and 7.

Notably, across most of these studies of ECEC between 0 and 3 years (in the United States and elsewhere), many low-income children attended informal care settings, not ECEC centers. Two exceptions, one from Canada and one from the United States, addressed larger scale implementations of center-based ECEC programs for children age 0 to 3. In Quebec, for example, a quasi-experimental study of the roll-out of universal early child care coverage found no effect on language skills at age 4 and negative impacts on behavioral, social, and motor outcomes, particularly for children entering as infants and toddlers and particularly for children of less educated parents (Baker, Gruber, & Milligan, 2008; Kottelenberg & Lehrer, 2014). It is notable, however, that the Quebec program covered relatively few 0- to 2-year-olds, involved relatively high adult to child ratios in care settings, and resulted in most infants and toddlers being cared for in accredited home care rather than in center care (Japel, Tremblay, & Cote, 2005).

In the U.S. study, Duncan and Sojourner (2013) used a large-scale randomized trial of a high-quality infant and toddler ECEC intervention (the Infant Health and Development Program) to extrapolate population-level effects of either a universal program or one specifically targeted at children from low-income families. The authors estimated that prior to school entry, in both cases, income disparities in cognitive abilities would be strongly reduced or eliminated, with treatment gains being substantial even 2 years after the end of the intervention (e.g., IQ score differences of 86% of a standard deviation between low-income children in the treatment and control groups).

ECEC Policy in Norway

In the present study, we build on the existing literature to examine Norway's national scale-up of universal ECEC for infants and toddlers. Scale-up of ECEC in Norway has been a gradual process across the past three decades, with the country having first set an aim of publicly subsidizing universal access to high-quality ECEC (beginning at age 1) in the 1980s. In turn, through the 1990s, Norwegian ECEC policy was focused on establishing federal regulations for quality of care (e.g., teacher educational requirements and an educational content framework) for both center-based care and family day care (Ministry of Education, 2014). And in 2002, Norwegian municipalities were mandated to provide access with "a right to child care for all children" in ECEC centers and/or family day care units, leading to an increasing number of ECEC slots for 1- and 2-year-olds.

Since the 2002 mandate, progression toward universal access has occurred incrementally as public spending has increased, and correspondingly, family fees for care have decreased. In 2004, a maximum fee for full-time care was introduced (NOK 2,750 or about US$400 per month at the 2004 exchange rate) and lowered further in 2006 (NOK 2,250 or about US$360 per month at the 2006 exchange rate), but sliding scales less than the maximum fee varied at the discretion of municipalities. Progress toward universal access also varied considerably across municipalities due to idiosyncratic local circumstances. Obstacles to increased ECEC slots included, for example, (a) lack of available spaces for building new centers, high building cost, and/or lack of available contractors (particularly for the major cities); (b) lack of qualified staff (particularly for the smaller municipalities); (c) concerns in some municipalities about overcoverage due to year-by-year variations in birth rates and for some municipalities conservative predictions of demand; and (d) local concerns about the availability of long-term earmarked funding from the government (Aspland Virak, 2006, 2009; Rindfuss, Guilkey, Morgan, Kravdal, & Guzzo, 2007). Nonetheless, across the 5-year time period investigated in the present study, the national coverage for 1- to 2-year-olds increased from 40% to 75% (Sæther, 2010; Scheistrøen, 2012).

With regard to age of entry, Norwegian parents with newborns receive up to 1 year paid leave; as a result, few children enter nonparental care prior to 9 months of age (Ministry of Children and Equality, 2014). After parental leave, parents have the choice of enrolling their children in publically subsidized ECEC or receiving cash benefits (approximately NOK 3,500 or US$560 per month at the 2006 exchange rate, 2006 being when many families in the present study were eligible) for staying home with their children until age 3.

With regard to quality, across the years of interest in the present study, national structural quality regulations in Norway stipulated that at least 30% to 35% of the staff should be ECEC teachers (with 3-year university college degrees) and that there should be a leader with ECEC teacher education in each center. Adult to child ratios of 1:3 for those younger than 3 and 1:6 for older children were also recommended but not enforced by law. In ECEC centers, most children younger than 3 are in infant-toddler groups of about nine children with three staff, one of them a trained teacher. Yet variations in group size and age composition are allowed, as long as the staff requirements are met.

The pedagogical content is guided by a "Framework" curriculum plan that sets out guidelines regarding the values and purpose of ECEC centers, their curricular objectives, and educational approaches (Ministry of Education, 2006). With regard to language development, which is of interest in the present study, the curriculum plan specifies that there should be stimulating verbal and nonverbal interactions in all everyday situations, age-appropriate use of learning materials, and a learning-rich environment including work with symbols, books, and reading. Thus, Norwegian ECEC is considered one national program, although the implementation of this

program varies between ECEC centers within the limits of the legal requirements for structural quality as well as the "Framework" plan.

Monitoring of ECEC quality in Norway is a collaborative responsibility of the ECEC centers and their municipalities, with a focus primarily on inspection of structural quality standards, safety and hygiene, and evaluation of the educational content (the latter varying in formality across municipalities). At the time of the present study, structural quality standards were mostly in accordance with regulations and recommended standards (Winsvold & Guldbrandsen, 2009) while not entirely met in all centers across the country (Brenna et al., 2010; OECD, 2015; UNICEF Innocenti Research Center, 2008). Observed quality has just recently been studied in Norway (Bjørnestad & Os, in press), showing that most classrooms meet only minimal quality standards as assessed with the Infant/Toddler Environment Rating Scale–Revised ITERS-R (Harms, Cryer, & Clifford, 2006), in particular with regard to hygiene, safety, and access to play materials.

The Present Study

We examined whether Norway's scale-up to national ECEC coverage beginning at age 1 (a) had consequences for children's early language skills and, if so, (b) differentially affected the language of children from low-, middle-, and high-income families and communities. Our focus on early language skills was driven by evidence that differences in early language help explain a considerable portion of the differences in school performance between lower and higher income children during elementary school (Durham, Farkas, Hammer, Tomblin, & Catts, 2007) and that early caregiving environments are critical to developing early language skills (Kuhl, 2004; Mustard, 2006; Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Shonkoff & Phillips, 2000; Snow, Burns, & Griffin, 1998).

For our analyses, we used population-based data ($n = 63,350$) on children living in more than 400 municipalities in Norway. Within these data, we focused our analyses on five birth cohorts of children who were sampled after the 2002 mandate for universal access. In these birth cohorts, we examined whether the ECEC policy scale-up (i.e., increasing availability of public ECEC over time) led to better mother-reported child language skills at age 3. More specifically, we estimated three types of statistical models that addressed complementary questions about the consequences of the scale-up.

First, using two-stage least square regression models, we attempted to isolate the causal effects of attending ECEC on individual children's language. Specifically, we used ECEC attendance rates within each child's municipality, and for her or his same birth cohort, as an instrument for that child's own ECEC attendance. The logic guiding these models was

that attendance rates provided an indicator of the availability of ECEC over time within municipalities. In turn, we expected that availability affected children's use of ECEC, for reasons unrelated to family selection, thereby allowing us to examine the effects of exogenous variance in ECEC use. Because we were also interested in whether ECEC use had differential impacts on children from low-, middle-, and high-income families, we estimated our models separately for these groups.

Second, we estimated DID models that exploited variations in ECEC use across birth cohorts and municipalities to estimate the causal effects of the policy scale-up at the level of municipalities. Specifically, we estimated the effects of changes in ECEC availability across cohorts, with the expectation that increasing availability should predict improved language skills within municipalities. Here again, we used ECEC attendance as an indicator of the likely availability of ECEC within municipalities for each birth cohort. And here again, we estimated models separately for low-, middle-, and high-income groups, although in this case with a focus on community-level income (i.e., within-municipality average family income for each birth cohort).

Third, we used fixed-effects regression to extend our study of the differential impacts of ECEC by income group. Specifically, we investigated whether the scale-up of universal ECEC in Norway predicted a narrowing of achievement gaps between low- and high-income children (in this case, language skill gaps). Our expectation was that municipalities that experienced a narrowing of ECEC-use gaps between low- and high-income children (i.e., municipalities in which increasing use of ECEC rose faster for low-income children than for high-income children) would also evidence the greatest narrowing of language skill gaps between these groups.

Method

Participants

Data are from the Norwegian Mother and Child Cohort Study (MoBa; for complete details, see Magnus et al., 2006, and www.fhi.no/morogbarn). Beginning in 1999, pregnant women in Norway who received routine exams at birth units delivering more than 100 births per year were invited to participate during their 17th-gestational-week visit. As of October 2010, 90,725 mothers of 108,639 children had enrolled and completed baseline assessments, which represented 42.1% of all eligible mothers in Norway. Written informed consent was obtained, and the study was approved by the Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate.

Questionnaires covering demographics, health, lifestyle, and child development were administered during the 17th, 22nd, and 30th weeks of gestation and at ages 0.5, 1.5, and 3.0 years (questionnaires are available online:

www.fhi.no/moba-en). Retention rates at 1.5 and 3.0 years were 72.4% and 59.3%, respectively. The present study uses data collected during pregnancy (demographics), at 1.5 years (child care use), and at 3.0 years (language skills). Linkage to the National Income and the Medical Birth Registries provided data on family income and infant health, respectively.

For the purposes of this study, we restricted the sample to 2002 through 2006 birth cohorts (*n* = 63,471 children) because the 2002 cohort was the first full cohort given the age 3 questionnaire and the 2006 cohort was the last for which registry data on family income were available by the child's age of 1.5 years. Moreover, by restricting analyses to these cohorts, we targeted a time of exceptional population increase in ECEC use during infancy due to policy changes. In Appendix A in the online supplemental material, we detail participation rates for the cohorts relative to population births in eligible birthing units.

Measures

Child language. Two separate language screening measures were used as indicators of broad levels of language ability, from language difficulties and delays to typical language ability. Mother reports were elicited to measure children's grammatical complexity at 3 years of age with the Norwegian version of a six-point scale previously used in a large-scale community study in the United Kingdom (Dale, Price, Bishop, & Plomin, 2003). The initial development of the measure was informed by the MacArthur Communicative Development Inventory: U.K. Short Form (Dionne, Dale, Boivin, & Plomin, 2003). A mother chose one of six statements that best described her child's ability, ranging from "Not yet talking" (one point) to "Talking in long and complex sentences, such as 'when I went to the park, I went on the swings' or 'I saw a man standing in the corner'" (six points). For general communication skills including receptive and expressive language, mothers answered six items from the communication domain of a normed Norwegian translation of the Ages and Stages Questionnaire (ASQ; Janson & Squires, 2004; Squires, Bricker, & Potter, 1999). The items included four original 36-month ASQ items, and one item each from the 18- and 48-month ASQ. Mothers responded on a three-point scale (*yes, a few times, not yet*) to communicative behavior descriptions (e.g., "When looking at a picture book, does your child tell you what is happening or what action is taking place in the picture?").

We combined these two measures (correlated by *r* = .62) by computing the mean of the standardized scores. We then transformed the final composite by $\log^{10}$ to correct for skewness, inverted values such that higher scores would reflect better language skills (in standard deviation units), and set the lowest score to equal zero. For more details about the distribution of items, psychometric properties, see Appendix B in the online supplemental material.

ECEC arrangements. Mothers reported the type of child care used at age 1.5 years, representing the child's primary care arrangement. Choices included "at home with mother or father," "at home with unqualified child minder," "family day care," and "center care." From these reports, we computed a dummy variable indicating whether children were in center-based ECEC versus any of the other arrangements that were not regulated to include educational content. We also analyzed the data by comparing children in center care exclusively with those in the "home with mother or father" group, but these models produced results that were statistically indistinguishable from the center-based care versus other arrangement analysis, and the policy relevance of our findings was maximized by comparing children in regulated center care with all other children in care arrangements that did not have a federally regulated educational curriculum.

Household income. We made use of annual tax records for each participating mother and for the fathers who had agreed to participate (77.6%). In cases wherein father's income was missing, we imputed this income using an expectation maximization algorithm, basing estimates on historical tax records on mother's income, assets, and debt dating back to 1993 as well as demographic information including total family income that was self-reported during pregnancy. We calculated a ratio of family income-to-needs, dividing total income by the OECD poverty line for each particular year (50% of the median income, adjusted for family size; OECD, 2011). The distribution of household income in the sample relative to the population distribution is provided in Appendix C in the online supplemental material.

For analyses, we divided families into three income groups: low income (< 25th percentile), middle income, and high income (>75th percentile). This approach allowed our adjustments for selection and our estimates of the exogenous components of ECEC use to vary across the three groups, with the assumption that both selection forces and rate of change in ECEC access due to policy change and municipality-specific factors likely differed for low- versus high-income families.

Covariates. Medical Birth Registry information was retrieved for child gender, birth weight (dichotomized: < 2,500 and ≥2,500 grams), Apgar score 5 min after birth, multiple birth (e.g., twins), and congenital syndromes (including Down syndrome, cleft lip and palate, and limb malformations). Parental education, partner status (single vs. partnered), non-Norwegian background, and number of siblings were reported by mothers at the 17th gestational week. Mothers reported on their anxious/depressive symptoms (Tambs & Moum, 1993) and partner/spouse relationship satisfaction (Rosand, Slinning, Eberhard-Gran, Roysamb, & Tambs, 2011) at 0.5, 1.5, and 3.0 years.

Statistical Analyses

We employed an array of analytic techniques to probe the causal effects of ECEC scale-up in Norway. Here, as our primary models, we focus the article on results from two-stage least squares (TSLS), generalized DID, and fixed-effect regression analyses, techniques that can provide quasi-experimental tests of causal hypotheses when correctly specified (e.g., Angrist & Pischke, 2009; Murnane & Willett, 2011).

For both our TSLS and DID models, we exploited variations in ECEC use across cohorts and municipalities as a proxy for variations in ECEC availability. Although we did not have access to administrative data on actual ECEC availability in Norway, we were able to estimate availability of ECEC within cohort by municipality clusters (i.e., each child was nested within a cohort of children based on birth year and municipality). Specifically, we used the proportion of children attending ECEC within each cohort by municipality cluster as an estimate of availability. To do so, we took a jackknife approach—iteratively excluding children one by one to compute the proportion—so that the cohort by municipality proportion in ECEC for child i excluded child i when calculating that child's corresponding proportion. In our TSLS models, proportion of children in ECEC served as the instrument; in our DID models, proportion of children in ECEC served as the treatment. Note that this approach led to an analytical sample of 63,350 children (from 63,471), dropping 121 children who were the only children represented within their cohorts by municipality cluster.

The general logic guiding these models was as follows. The rollout of universal ECEC in Norway created an empirical opportunity because children born at different times in different municipalities had varying access to ECEC for reasons beyond family selection. Within a short historical window, there is little reason to expect the language in the child population to improve across birth cohorts unless influenced by far-reaching environmental changes in Norway, such as influential policy shifts. Moreover, the interaction of (a) when children were born by (b) where families lived (i.e., birth cohort by municipality variations in ECEC scale-up) likely rendered the progression toward universal provision of ECEC difficult for families to forecast, and local idiosyncratic causes of ECEC scale-up rates within each municipality helped rule out the influences of other national policy changes or population trends. Below, we provide details of our TSLS and DID model specifications and key assumptions.

TSLS (instrumental variable) analyses. The first and second stages of our TSLS models were estimated as Equation 1 and Equation 2, respectively. In the first stage (Equation 1), ECEC use was regressed on our instrument (i.e., the proportion attending ECEC in the corresponding cohort by municipality cluster for child i) and our covariate set, indicated with Ws. In the second stage (Equation 2), children's language scores were regressed on predicted values of ECEC from the first stage, plus covariates.

$$ECEC_i = \pi_0 + \pi_1 ProportionECEC_{1i} \\ + \pi_2 W_{1i} + \ldots + \pi_{1+r} W_{ri} + \upsilon_i.$$

Equation 1

$$Language_i = \beta_0 + \beta_1 ECEC'_i + \beta_2 W_i \\ + \ldots + \beta_{1+r} W_{ri} + \mu_i.$$

Equation 2

To remove omitted between-municipalities heterogeneity, we estimated our TSLS models with municipality fixed effects. Our covariate set included all child and family variables in Table 1 as well as cohort by municipality averages for these variables. To examine variations by family income, we estimated the TSLS model separately for children in low-, middle-, and high-income families.

For these TSLS models, we examined three key assumptions, one of which is related to instrument strength and two of which are related to instrument validity: (a) the instrument should be strongly associated with the treatment (i.e., F-test statistic of 10 or greater), (b) the instrument(s) should be independent of factors (other than treatment) that influence the outcome variable, and (c) the instrument should influence the outcome variable only through the treatment (i.e., the exclusion restriction). Satisfying the first assumption, in all cases, F values from our first-stage models exceeded the criterion of 10 (see online supplemental material Table S4).

We also found that the study covariates were balanced across levels of our instrument. In two-way (cohort by municipality) fixed-effects regression models, nearly all associations between cluster-specific levels of the covariates and cluster-specific levels of ECEC participation were very small and null (see online supplemental material Table S5), with only two exceptions being evident: for children in high-income families, ECEC participation was negatively related to parent education and preterm births. Moreover, consistent with the exclusion restriction, there was no evidence that our instrument was associated with language scores once controlling for ECEC use and none of our covariates provided alternative pathways through which the instrument affected child language (see online supplemental material Table S6).

DID analyses. We estimated generalized DID models in the present study to examine whether rate of expansion of ECEC availability was, in turn, predictive of rate of improvement in children's language skills (for similar empirical approaches to the study of ECEC policy expansion, see Bassok, Fitzpatrick, & Loeb, 2014; Cascio & Schanzenbach, 2013). DID designs include two orders of differencing—a pre-post treatment difference and a comparison of the size of pre-post treatment differences across groups with varying levels of exposure to the treatment—and may be generalized to quasi-experiments in which there are more than two groups (e.g.,

TABLE 1

Descriptive Statistics for Child Language, Early Childhood Education and Care Use, and Study Covariates (n = 63,350)

	M (SD) or %	Range	Percentage missing
Early language skills	7.86 (1.04)	0.00–8.58	40.38
Early childhood education and care variables			
Use by 18 months	45.73%		27.25
Child covariates[a]			
Boy	51.00%		0.00
Apgar score at 5 min	9.40 (0.78)	0.00–10.00	0.00
Preterm/low birth weight	4.30%		<1.00
Multiple birth (e.g., twins)	3.51%		0.00
Congenital syndromes	5.04%		0.00
Parent and family covariates			
Maternal age	29.59 (4.58)	14.00–47.00	3.59
Income-to-needs	2.16 (0.78)	0.00–10.00	0.00
Years of parent education[b]	15.02 (2.44)	8.00–18.00	4.45
Number of siblings	0.79 (0.90)	0.00–15.00	0.00
Single parent	4.61%		2.54
Non-Western background	10.11%		5.52
Maternal anxiety/depression	1.25 (0.30)	1.00–4.00	32.43
Relationship satisfaction	1.79 (0.72)	1.00–6.00	14.48

[a]Information about all parent variables was taken from the questionnaire at the 17th gestational week for the Norwegian Mother and Child Cohort Study, and from the 6-month interview for the Behavior Outlook Norwegian Developmental Study. [b]Years of parent education is for the most educated parent in the household.

in the present study, the cohort by municipality clusters that differ in proportion of ECEC availability) and continuous treatment variables (e.g., in the present study, the proportion of children in ECEC within a cluster); in essence, generalized DID models are two-way fixed-effects models (e.g., in the present study, cohort by municipality). To examine variations by community income, we estimated the DID models separately for the poorest 25% of the clusters (low income), the middle 50% of the clusters (middle income), and the richest 25% of the clusters (high income). Specifically, for each income group, our DID equation took the following form, where *Language* for cohort by municipality cluster *cm* was regressed on proportion of children in ECEC for cluster *cm*, and γc and δm were vectors of cohort and municipality fixed effects:

$$Language_{cm} = \alpha + \beta_1 ProportionECEC_{cm} + \gamma_c + \delta_m + \varepsilon_{cm}.$$

Equation 3

Within-municipality regression estimates of language gap changes. In addition to our TSLS and DID models, we estimated a within-municipality fixed-effects regression that was also specified to estimate DID. In this case, we were interested in examining associations between income-related disparities in ECEC use and, in turn, income-related disparities

in early language skills. And more specifically, we examined whether within-municipality changes in ECEC-use disparities predicted changes in language skill disparities.

Prior to estimating the fixed-effects regression model, the first-level differencing involved computing within-municipality difference scores for the 2002 and 2006 cohorts. For both rate of ECEC use and language skills, we subtracted the average for low-income children from the average for high-income children. For example, for differences in average language skills, the difference score for cohort *c* and municipality *m* was computed as $LanguageGap_{cm} = HighIncomeLang_{cm} - LowIncomeLang_{cm}$. Next, to create a set of DID indicators, we subtracted the language gap for the 2002 cohort from the language gap for the 2006 cohort. That is, for municipality *m* we computed $\Delta LanguageGap_m = LanguageGap_{2006m} - LanguageGap_{2002m}$. And in a similar fashion, we created DID scores for ECEC use as well as the study covariates.

In turn, to examine whether narrowing of ECEC-use gaps over time (i.e., smaller difference scores for the 2006 cohort compared with the 2002 cohort) predicted narrowing language gaps, we estimated a (within-municipality) fixed-effects regression model as presented in Equation 4; $\Delta LanguageGap_m$ was the change score indicating the extent to which language score differences between low- and high-income children narrowed or widened across cohorts for municipality *m*, $\Delta ECECGap_{1m}$ was the change score

indicating the extent to which ECEC-use differences between low- and high-income children narrowed or widened across cohorts for municipality m, and W_{1i} to W_{ri} were municipality-level averages for the child and family covariates presented in Table 1.

$$\Delta LanguageGap_m = \pi_0 + \pi_1 \Delta ECECGap_{1m} + \pi_2 W_{1i} + \ldots + \pi_{1+r} W_{ri} + \upsilon_i.$$

Equation 4

For this analysis, we included only the 173 municipalities that had a reasonable representation of lower income and higher income households (we chose 10 households in each of these categories as the threshold, although the results were similar if we included municipalities with fewer low- and high-income households).

Missing data. Despite complete data on some key indicators such as family income, there were considerable missing data due to attrition. Most notably, 27.25% of children were missing ECEC data at 1.5 years, and 40.37% of children did not have complete language data at 3.0 years. Although likelihood of having missing values on the study covariates or language outcome was by and large unrelated to the "treatment" of interest, likelihood of attrition was higher for more socially disadvantaged families and children with congenital syndromes. To account for this attrition, our statistical models were estimated using multiple imputation for missing values, combining estimates and standard errors according to Rubin's Rules (Rubin, 1987). Given the sample size and complexity of our models, we were limited to using five imputations (see online Appendix E for further details about missing data).

Results

Descriptive Statistics

In Table 1, we provide descriptive statistics for the study variables. In addition, in Figure 1, we plot the proportion of children in ECEC for each birth cohort. In the figure, it is evident that ECEC use increased across cohorts by nearly 30 percentage points or more for children from low-, middle-, and high-income households. Correspondingly, for low-income children, use of parental care had the most rapid declines across cohorts, while for middle- and high-income children, declining use was fairly similar for parental care and family day care.

TSLS Models

In Table 2, we summarize results from our TSLS models estimating the effects of ECEC use at 18 months on language at 3 years. For low-income children, the estimated effect of ECEC use was statistically significant and more than twice as large as the estimates for middle- and high-income children. Based on the TSLS estimate, low-income
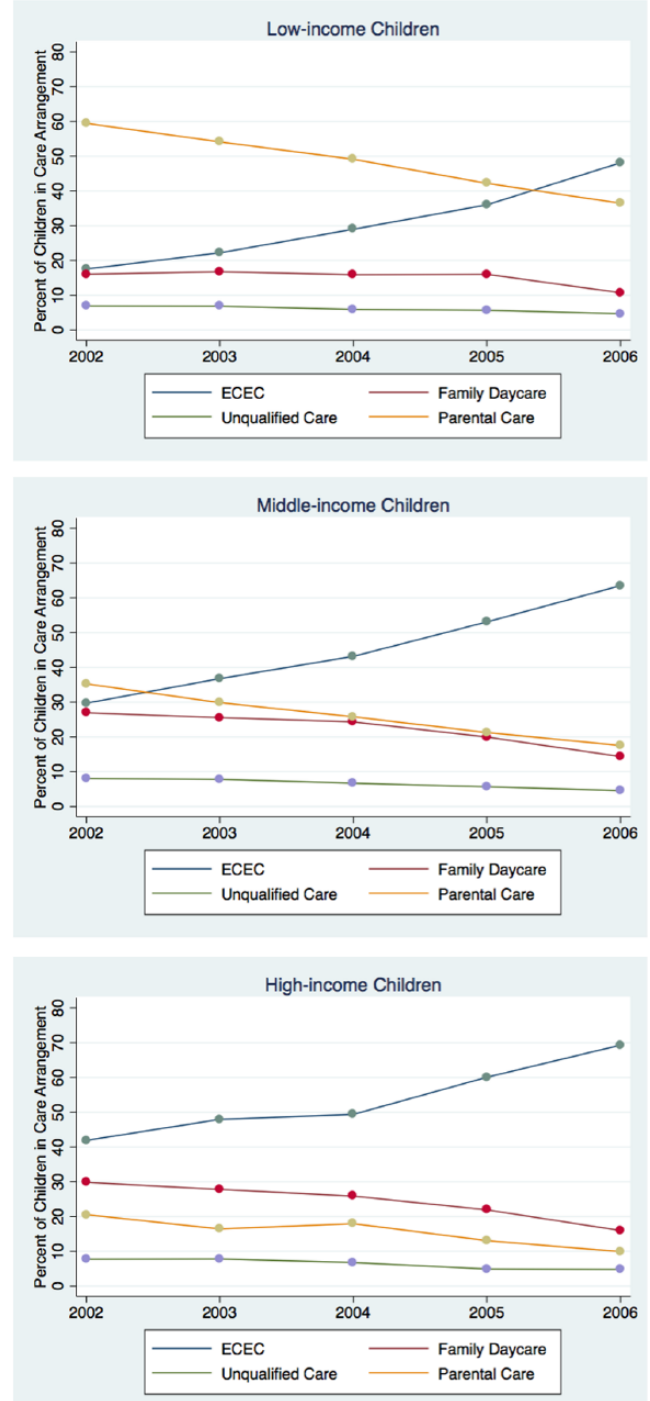

FIGURE 1. *Trends in center-based early childhood education and care and other arrangements across birth cohorts.*

children who were in ECEC at 18 months had language scores at 3 years that were, on average, 89% of a standard deviation higher than those for low-income children who did not attend ECEC. For middle- and high-income children, on the other hand, ECEC use was associated with approximately 31% and 29% of a standard deviation difference in

TABLE 2

Predicting Language Scores at 36 Months from Early Childhood Education and Care Use at 18 Months: Instrumental Variable Models

	Low income (n = 17,395)	Middle income (n = 31,073)	High income (n = 14,882)
	b [95% CI]	b [95% CI]	b [95% CI]
Early childhood education and care use at 18 months	.89*** [0.28–1.49]	.31* [–0.05–0.67]	.29 [–0.12–0.70]

Note. In the two-stage least squares models, *low, middle,* and *high income* refer to family income levels. Low-income families were at the 25th percentile or lower, and high-income families were at the 75th percentile or higher.
*p < .10. ***p < .01.

TABLE 3

Predicting Language Scores at 36 Months from Early Childhood Education and Care Coverage at 18 Months: Difference-in-Difference Models

	Low income (n = 495)	Middle income (n = 1,006)	High income (n = 489)
	b [95% CI]	b [95% CI]	b [95% CI]
Early childhood education and care coverage at 18 months	.35*** [0.10–0.60]	.21 [–0.11–0.52]	–.04 [–0.44–0.36]

Note. Sample sizes for the difference-in-difference models indicate the number of cohort-by-municipality clusters of children. In these models, *low, middle,* and *high income* refer to cluster-level averages of family income. Low-income communities were those at the 25th percentile or lower, and high-income communities were those at the 75th percentile or higher.
***p < .01.

language scores, respectively; the estimate was null for high-income children and only approached significance for middle-income children. It is worth noting, however, that even for the low-income children, the 95% confidence interval was quite wide, including values ranging from 28% to 149% of a standard deviation; given similarly large confidence intervals for the middle- and high-income groups, the estimates for the three income groups were not statistically distinguishable from one another.

DID Models

In Table 3, we summarize results from our DID models. While these estimates did not significantly differ from zero for children in middle- and high-income communities (i.e., children in cohort by municipality clusters that were, on average, middle and high income), increasing ECEC availability significantly predicted improved language for children in low-income communities. Regarding effect size, it is critical to note that the coefficients in these models represent the estimated changes in language scores (in percentages of a standard deviation) given a change of 0% to 100% "availability" (i.e., no children within a cluster versus all children in a cluster attending ECEC). Thus, given a one standard deviation

change in our ECEC availability indicator (i.e., a 15.4 percentage point increase), we would expect a 5.39% standard deviation increase in the average language scores of children within clusters. It is also important to note that confidence intervals for the three income groups overlapped, indicating they were not statistically distinguishable from one another.

Within-Municipality Regression Estimates of Language Gap Changes

As a final analytical step, we estimated ordinary least squares regression models examining whether municipalities that evidenced narrowing gaps in ECEC use between higher and lower income children (from 2002 to 2006) also evidenced narrowing gaps in language scores (see Table 4). We estimated unadjusted models and, in turn, models that controlled for (a) within-municipality mean levels of the covariates in Table 1 and (b) changes across cohorts in the covariates. In total, 58.29% of municipalities demonstrated narrowing gaps between the proportions of lower versus higher income children who were in ECEC at 18 months, and these narrowing gaps in ECEC use were positively predictive of narrowing gaps in language scores. In the two models adjusting for covariate levels or changes, every 10

TABLE 4

Municipality-Level Association Between Changes in Early Childhood Education and Care–Use Gap and Changes in Language Skill Gap Between Lower and Higher Income Households

Change in Early Childhood Education and Care–Use Gap	Change in Child Language Gap
	b (SE)
Unadjusted OLS estimates	.14** (.06)
OLS adjusted for average within-municipality covariate levels	.16*** (.07)
OLS adjusted for within-municipality covariate rates of change	.16*** (.06)

Note. OLS = ordinary least squares.
** *p* < .05. *** *p* < .01.

percentage point narrowing of the gap in ECEC use between lower and higher income households was associated with a 1.56 percentage point narrowing in the language skill gap.

Sensitivity and Robustness Checks

In addition to our primary models reported here, we conducted a number of analyses to examine the sensitivity of our results to respecification and, more generally, to examine the robustness of our main findings. In Appendix F in the online supplemental material, we provide a brief overview of results from these sensitivity and robustness checks.

Discussion

To date, most evidence from the United States on ECEC at scale comes from targeted programs for disadvantaged preschool children (3- to 5-year-olds); international evidence on universal scale-up beginning at younger ages provides a useful addition to the cumulative knowledge. In the present study, we investigated the consequences of Norway's national scale-up of universal ECEC, beginning at age 1, for children's early language skills. In doing so, we gave special attention to differential consequences for children from low-, middle-, and high-income families. In a population-based sample, we found that scale-up of Norway's universal ECEC led to improvements in children's early language skills, with low-income children's evidencing this most robustly. Our results were, by and large, consistent with the hypothesis that attending large-scale public ECEC is improving low-income children's language skills in Norway and thereby may be narrowing early achievement gaps between low- and high-income children. More specifically, our results provided three complementary pieces of evidence on the effects of universal ECEC scale-up in Norway.

First, in TSLS regression models, we found that attending ECEC at 18 months was predictive of better language skills at age 3, primarily for low-income children. On average, low-income children attending ECEC were estimated to have language skills approximately 90% of a standard deviation higher than low-income children not in ECEC. A word of caution is required when interpreting the size of this effect, however, given that our 95% confidence interval for low-income children ranged from 29% to 149% of a standard deviation. Despite this wide range, the estimates reached statistical significance, and even the lower bound of nearly one third of a standard deviation would be of considerable practical importance given the long-term risks associated with limited early language skills (Durham et al., 2007).

Second, in DID models, we found that the increasing availability of ECEC in Norway led to improvements in the average language skills within low-income municipalities. The size of these improvements were, however, considerably smaller than those estimated for the effects of ECEC use on individual children's language scores—a 15 percentage point increase in the availability of ECEC predicted slightly more than a 5% of standard deviation increase in the average language scores within low-income municipalities. One reason for this seemingly small effect was likely the fact that not all children in low-income municipalities were, themselves, living in low-income families; based on our TSLS models, we would expect less robust effects of ECEC on the language skills of children in middle- and high-income families compared with those children in low-income families. Moreover, our initial DID models did not address for whom ECEC availability was increasing most rapidly (e.g., was availability increasing at similar or faster rates for low- vs. high-income children as a function of cost structures?).

Third, therefore, we estimated regression models focused on the question of for whom ECEC attendance increased within municipalities. More specifically, we compared municipalities according to disparities in ECEC attendance between low- versus high-income children and the implications of changes over time in these disparities for language outcomes. Municipalities in which use of ECEC increased more rapidly over time for low-income children than for high-income children also evidenced the greatest narrowing of language skill gaps between these groups of children. In 2006, in the population-based sample we examined, there remained an ECEC-use disparity of approximately 20 percentage points between low- and high-income children (see Figure 1); our regression model estimates indicated that closing this gap would narrow income gaps in early language skills by a little more than 3 percentage points. While these effects are modest in absolute terms, they should be evaluated with attention to the fact that strong early language skills are excellent predictors of long-term achievement and well-being outcomes (e.g., Durham et al., 2007; Farkas &

Beron, 2004). Thus, rather small changes in early language due to universal ECEC scale-up might still prove to have considerable implications for these children's life chances and for hopes of reducing social disparities in Norway.

Optimism about the potential benefits of Norway's ECEC program is justified by the fact that our findings are consistent with other Norwegian studies. There is evidence that the scale-up of ECEC for preschoolers in Norway in the late 1970s improved life chances into early adulthood (Havnes & Mogstad, 2011); in addition, a recent study from Norway's capital, Oslo, showed that children who, due to a lottery, entered ECEC on average 6 months earlier than their counterparts (i.e., about 15 months of age rather than 19 months) scored 12% of a standard deviation higher on math and reading tests at age 6 (Drange & Havnes, 2015). Yet it is critical to recognize that any hope of reducing social disparity via ECEC relies on strong rates of participation in public ECEC among disadvantaged families. While participation among low-income children grew rapidly during the time period we studied, only a little more than half of the municipalities in Norway actually narrowed rates of ECEC use between low- and high-income families. With national rates of ECEC use for 1- to 2-year-olds now near 80%, disparities may be decreasing but remain a concern as Norway has recently increased its ECEC policy focus on participation rates among socially disadvantaged families (OECD, 2015).

Extending the Cumulative Knowledge About ECEC at Scale: Relevance in the United States and Internationally

Beyond Norway, our findings of improved early language skills for low-income children during ECEC scale-up are consistent with those demonstrated in randomized trials of infant and toddler ECEC (Duncan & Sojourner, 2013) and nonexperimental studies of high-quality infant and toddler care, including those studies that have demonstrated larger effects of high-quality care for children in low-income families (Geoffroy et al., 2007; McCartney et al., 2007). Our findings are also consistent with evaluations of universal preschool programs at scale in the United States for which positive achievement gains have been most pronounced among low-income children but also evident among middle-income children (Gormley et al., 2005; Gormley et al., 2008). Our findings extend these existing lines of work because we have examined such effects within a large-scale, national implementation of quality-regulated ECEC, beginning at age 1, when brain maturation, language learning, and cognitive development are rapid.

Compared with effect sizes from previous studies, our estimated effects were smaller, however. Some studies, in fact, report considerably larger effect sizes. One case in point are the projections based on analyses of low birth weight children in the United States (Duncan & Sojourner, 2013); at age 3, estimated population gaps in IQ would be closed by 87% assuming a 2-year universal ECEC program. Somewhat smaller effect sizes were reported in Duncan and Magnuson's (2013) meta-analysis of preschool interventions (i.e., 21% of a standard deviation), particularly for studies dating after 1980 (i.e., 16% of a standard deviation). Also relevant are effect sizes for nonexperimental studies of infant and toddler care, ranging from 9% to 16% of a standard deviation for school readiness scores for each additional year children attend center care in the NICHD Study of Early Child Care and Youth Development, for example (NICHD, 2000).

Given that Norway's ECEC program lasts until school entry (the year children turn 6), it is possible that our estimates at age 3 are smaller than they would be at school entry. However, when comparing our estimated effects of early ECEC in Norway to preschool interventions in the United States, the supportive sociopolitical context of Norway must be considered (Dearing & Zachrisson, 2017). Beyond paid parental leave, Norway offers free health care (including well-being clinics for parents and children) and a more family-friendly labor market than does the United States (e.g., 5 weeks paid holiday). In addition, there are lower levels of material deprivation associated with child poverty in Norway compared with the United States (UNICEF Innocenti Research Center, 2012). It is reasonable to assume that the counterfactual condition to regulated ECEC for low-income children (e.g., being cared for by parents or unqualified child minders), is to a lesser extent than in the United States, associated with disadvantaged developmental contexts. Thus, the baseline for children's language development may well be higher; indeed, in the population-based data analyzed in the present study, language score differences between low- and high-income children are only about 15% of a standard deviation, on average. In Canada, a national context somewhat more comparable to Norway than is the United States, evidence on ECEC appears mixed: ECEC attendance for disadvantaged children prior to age 4 predicted 36% to 87% of a standard deviation higher scores on school achievement in Canada (Geoffroy et al., 2007; Geoffroy et al., 2010), but this finding is in stark contrast to evaluations finding null (or negative) effects on behavior of ECEC policies in Quebec (Baker et al., 2008; Kottelenberg & Lehrer, 2014).

Our results should also be interpreted in light of the national curriculum ("Framework" plan). While structural quality appears high, this curriculum provides very general guidelines for pedagogical practice, which likely varies considerably across centers. Relatively speaking in international comparisons, Norwegian ECEC teachers may often emphasize free play, minimize staff-child interactions, and avoid formalized direct instructions (OECD, 2015), potentially leading to low instructional quality. Considering the emphasis given to high-quality curriculum as a core "active ingredient" in ECEC promoting cognitive development (Duncan & Magnusen, 2013), it is reasonable to expect that a more

academically focused and structured curriculum could lead to larger effect sizes for language development than we detected.

As a final note about the relevance of our findings in the United States and internationally, it is worth considering that while there is some evidence of negative behavioral consequences of early, extensive, and continuous ECEC (e.g., Belsky, 2001; Huston, Bobbit, & Bentley, 2015), this does not appear to be the case Norway (Dearing, Zachrisson, & Nærde, 2015; Zachrisson, Dearing, Lekhal, & Toppelberg, 2013). One reason may be that paid parental leave policy makes 1 year the most common age of entry into ECEC in Norway, whereas many children in the United States enter some form of nonparental care by 9 months (Halle et al., 2009). More generally, however, the internal validity of many studies demonstrating negative consequences of ECEC has been called into question (Dearing & Zachrisson, 2017).

Limitations and Strengths of the Present Study

One of the more serious limitations to the present study was the high rates of attrition in this population-based sample, with evidence that more disadvantaged and higher risk children were less likely to be retained through age 3. As described here, and in our supplementary materials, our results proved robust across approaches to handling this limitation (i.e., multiple imputation and listwise deletion). While our primary approach (multiple imputation) is recommended for bias reduction with high levels of attrition that is correlated with study covariates (e.g., Graham, 2009), it is possible that our estimates of the effects of ECEC scale-up do not, in fact, apply to the most seriously disadvantaged or developmentally at-risk children in Norway.

It is also important to note that in the population-based sample, our language measure was based on maternal reports. There is, however, excellent evidence of the validity of parent reports of child language at age 3 via concurrent and predictive correlations with direct assessment tools (e.g., Feldman et al., 2005; Thal, O'Hanlon, Clemmons, & Fralin, 1999). Even so, the two maternal report measures employed in the population-based study are screening tools for language problems/delay.

Finally, it is critical to note that our instrument in the TSLS models and treatment variable in the DID models was, in essence, a proxy indicator. Ideally, we would have exploited public administrative data on publicly funded ECEC availability within municipalities; however, for purposes of protecting participants' anonymity, municipalities were not identifiable in MoBa, and therefore we could not link public ECEC data to MoBa. However, MoBa's being a population-based sample boosts our confidence in our estimates of ECEC availability. Moreover, an important strength of the present study was our ability to employ a variety of methods aimed at probing the causal hypothesis and determining whether the results of our TSLS and DID models were robust to alternative specifications.

While randomized experiments are the safest method for ensuring internal validity, methodologists increasingly encourage cause probing in nonexperimental work. Best practice recommendations highlight the importance of (a) using statistical methods and designs that help rule out omitted variables bias (Duncan, Magnusson, & Ludwig, 2004; Foster, 2010; McCartney, Bub, & Burchinal, 2006; Shadish, Cook, & Campbell, 2002) and (b) examining robustness and sensitivity across multiple methods that rely on somewhat different assumptions regarding plausible alternative hypotheses (Jo & Vinokur, 2011; Morgan & Winship, 2015; Murnane & Willett, 2011; Shadish et al., 2002). In the present study, we examined alternative specifications of our instrumental variable, DID, and within-municipality regression models, as summarized in the online appendices. Our results, by and large, proved robust across this array of alternative methods.

Conclusion

Even as preschool programs at scale become increasingly common in the United States and elsewhere, debate should continue about the relative costs and benefits of universal versus targeted ECEC policies (e.g., do targeted public programs produce similar (or larger) gains among socially disadvantaged children than do universal programs?). Yet our findings in Norway juxtaposed with the few universal preschool program evaluations in the United States should help push that debate away from targeted approaches that exclude middle-class and higher income children. Furthermore, whereas most policy discussions in the United States are centered on preschool children, we believe these findings should provoke more conversation about the value of ECEC at scale for the younger children. Although our findings are limited to shorter term outcomes, with fade-out a legitimate concern (Bailey, Duncan, Odgers, & Yu, 2017), the present study increases evidence that nations can implement publically subsidized and regulated ECEC programs for very young children at scale with a potential benefit of narrowing achievement gaps.

Acknowledgments

References

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Aspland Virak. (2006). *Analyse av barnehagetall pr 20.09.2006* [Analyses of ECEC numbers by Sept 20th 2006]. Retrieved

from http://www.regjeringen.no/en/dep/kd/dok/rapporter_planer/rapporter/2007/analyse-av-barnehagetall-pr-20092006.html?id=456310

Aspland Virak. (2009). *Analyse av barnehagestatistikk—status for utbygging og ventelister pr. 20. september 2009* [Analysis of ECEC statistics—status for expansion and waiting lists by Sept 20th 2009]. Retrieved from http://www.regjeringen.no/upload/KD/Vedlegg/Barnehager/Rapporter%20og%20planer/Sluttrapport_analyse%20av%20barnehagetall%20per%2020.09.09.pdf

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39. doi: 10.1080/19345747.2016.1232459

Baker, M., Gruber, J., & Milligan, J. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, *116*(4), 709–745. doi: 10.1086/591908

Barnett, W. S. (2010). Universal and targeted approaches to preschool education in the United States. *International Journal of Child Care and Education Policy*, *4*(1), 1–12. doi: 10.1007/2288-6729-4-1-1

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, *333*, 975–978. doi: 10.1126/science.1204534

Bassok, D., Fitzpatrick, M., & Loeb, S. (2014). Does state preschool crowd-out private provision? The impact of universal preschool on the childcare sector in Oklahoma and Georgia. *Journal of Urban Economics*, *83*, 18–33. doi: 10.1016/j.jue.2014.07.001

Belsky, J. (2001). Emanuel Miller Lecture—Developmental risks (still) associated with early child care. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*, 845–859.

Bjørnestad, E., & Os, E. (in press). Quality in Norwegian child-care for toddlers using ITERS-R. *European Early Childhood Education Research Journal*.

Brenna, L. R., Bjerkestrand, M., Broström, S., Fagerli, B., Hernes, I., Hornslien, Ø., … Tørresdal, B. (2010). *Med forskertrang og lekelyst. Systematisk pedagogisk tilbud til alle førskolebarn* (NOU 2010:8) [With urge for research and playfulness. Systematical educational programs for all preschool children (Research report 2010:8)]. Retrieved from http://www.regjeringen.no/nb/dep/kd/dok/nouer/2010/nou-2010-8.html?id=616123

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

Cascio, E. U., & Schanzenbach, D. W. (2013). *The impacts of expanding access to high-quality preschool education* (No. w19735). Cambridge, MA: National Bureau of Economic Research.

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*, 149–160. doi: 10.1037/0003-066X.60.2.149

Cote, S. M., Doyle, O., Petitclerc, A., & Timmins, L. (2013). Child care in infancy and cognitive performance until middle childhood in the Millennium Cohort Study. *Child Development*, *84*(4), 1191–1208. doi: 10.1111/cdev.12049

Dale, P. S., Price, T. S., Bishop, D. V., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research*, *46*, 544–560. doi: 10.1044/1092-4388(2003/044)

Dearing, E. (2014). The state of research on children and families in poverty: Past, present, and future empirical avenues of promise. In K. McCartney, H. Yoshikawa, & L. Forcier (Eds.), *Improving the odds for America's children* (pp. 203–216). Cambridge, MA: Harvard Education Press.

Dearing, E., McCartney, K., & Taylor, B. A. (2009). Does higher quality early child care promote low-income children's math and reading achievement in middle childhood? *Child Development*, *80*(5), 1329–1349. doi: 10.1111/j.1467-8624.2009.01336.x

Dearing, E., & Zachrisson, H. D. (2017). Concern over internal, external, and incidence validity in studies of child-care quantity and externalizing behavior problems. *Child Development Perspectives*, online first. doi: 10.1111/cdep.12224

Dearing, E., Zachrisson, H. D., & Nærde, A. (2015). Age of entry into early childhood education and care as a predictor of aggression faint and fading associations for young Norwegian children. *Psychological Science*, *26*(10), 1595–1607. doi: 10.1177/0956797615595011

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development*, *74*, 394–412. doi: 10.1111/1467-8624.7402005

Drange, N., & Havnes, T. (2015). *Child care before age two and the development of language and numeracy: Evidence from a lottery*. IZA Discussion Paper No. 8904. Bonn, Germany: IZA Institute of Labor Economics.

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives*, *27*(2), 109–132. doi: 10.1257/jep.27.2.109

Duncan, G. J., Magnuson, K. A., & Ludwig, J. (2004). The endogeneity problem in developmental studies. *Research in Human Development*, *1*, 59–80. doi: 10.1207/s15427617rhd0101&2_5

Duncan, G. J., & Sojourner, A. J. (2013). Can intensive early childhood intervention programs eliminate income-based cognitive and achievement gaps? *Journal of Human Resources*, *48*(4), 945–968. doi: 10.1353/jhr.2013.0025

Durham, R. E., Farkas, G., Hammer, C. S., Tomblin, J. B., & Catts, H. W. (2007). Kindergarten oral language skill: A key variable in the intergenerational transmission of socioeconomic status. *Research in Social Stratification and Mobility*, *28*, 294–305. doi: 10.1016/j.rssm.2007.03.001

European Commission. (2011). *Communication from the commission: Early childhood education and care: Providing all our children with the best start for the world tomorrow*. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0066:FIN:EN:PDF

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research*, *33*(3), 464–497. doi: 10.1016/j.ssresearch.2003.08.001

Farran, D. C. (2016). *We need more evidence in order to create effective pre-K programs*. Evidence Speaks Reports, 1. Washington, DC: Brookings Institution.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at

ages 2 and 3 years. *Child Development*, *76*(4), 856–868. doi: 10.1111/j.1467-8624.2005.00882.x

Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology*, *46*(6), 1454–1480.

Fox, N. A., & Rutter, M. (2010). Introduction to the special section on the effects of early experience on development. *Child Development*, *81*(1), 23–27. doi: 10.1111/j.1467-8624.2009.01379.x

Geoffroy, M. C., Cote, S. M., Borge, A. I., Larouche, F., Seguin, J. R., & Rutter, M. (2007). Association between nonmaternal care in the first year of life and children's receptive language skills prior to school entry: The moderating role of socioeconomic status. *Journal of Child Psychology and Psychiatry*, *48*(5), 490–497. doi: 10.1111/j.1469-7610.2006.01704.x

Geoffroy, M. C., Cote, S. M., Giguere, C. E., Dionne, G., Zelazo, P. D., Tremblay, R. E., … Seguin, J. R. (2010). Closing the gap in academic readiness and achievement: The role of early childcare. *Journal of Child Psychology & Psychiatry*, *51*(12), 1359–1367. doi: 10.1111/j.1469-7610.2010.02316.x

Gormley, W. T., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science*, *320*, 1723–1724. doi: 10.1126/science.1156019

Gormley, W. T., Jr., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, *41*, 872–884. doi: 10.1037/0012-1649.41.6.872

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. doi: 10.1146/annurev.psych.58.110405.085530

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: Mechanistic insights from human and animal research. *Nature Reviews Neuroscience*, *11*, 651–659. doi: 10.1038/nrn2897

Halle, T., Hair, M. N., Weinstein, D., Vick, J., Forry, N., & Kinukawa, A. (2009). *Primary child care arrangements of US infants: Patterns of utilization by poverty status, family structure, maternal work status, maternal work schedule, and child care assistance*. Washington, DC: Child Trends and Office of Planning Research and Evaluation.

Harms, T., Cryer, D., & Clifford, R. M. (2006). *Infant/Toddler Environment Rating Scales* (rev. ed.). New York, NY: Teachers College Press.

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. doi: 10.1257/pol.3.2.97

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. doi: 10.1016/j.dr.2005.11.002

Huston, A. C., Bobbit, K. C., & Bentley, A. (2015). Time spent in child care: How and why does it affect social development? *Developmental Psychology*, *51*, 621–634. doi: 10.1037/a0038951

Janson, H., & Squires, J. (2004). Parent-completed developmental screening in a Norwegian population sample: A comparison with US normative data. *Acta Paediatrica*, *93*, 1525–1529. doi: 10.1111/j.1651-2227.2004.tb02641.x

Japel, C., Tremblay, R. E., & Cote, S. (2005). "Quality counts: Assessing the quality of daycare services based on the Quebec Longitudinal Study of Child Development." *Choices*, *11*, 1–42.

Jo, B., & Vinokur, A. D. (2011). Sensitivity analysis and bounding of causal effects with alternative identifying assumptions. *Journal of Educational and Behavioral Statistics*, *36*, 415–440. doi: 10.3102/1076998610383985

Kottelenberg, M. J., & Lehrer, S. F. (2014). Do the perils of universal childcare depend on the child's age? *CESifo Economic Studies*, *60*, 338–365. doi: 10.1093/cesifo/ifu006

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843. doi: 10.1038/nrn1533

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*, 52–66. doi: 10.1016/j.econedurev.2005.11.005

Magnus, P., Irgens, L. M., Haug, K., Nystad, W., Skjaerven, R., & Stoltenberg, C. (2006). Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology*, *35*, 1146–1150. doi: 10.1093/ije/dyl170

Magnuson, K., & Shager, H. (2010). Early education: Progress and promise for children from low-income families. *Children and Youth Services Review*, *32*, 1186–1198. doi: 10.1016/j.childyouth.2010.03.006

McCartney, K., Bub, K. L., & Burchinal, M. (2006). Selection, reflection, detection. *Monographs of the Society for Research in Child Development*, *71*(Serial No. 285), 105–126. doi: 10.1111/j.1540-5834.2006.00407.x

McCartney, K., Dearing, E., Taylor, B. A., & Bub, K. L. (2007). Quality child care supports the achievement of low-income children: Direct and indirect pathways through caregiving and the home environment. *Journal of Applied Developmental Psychology*, *28*(5/6), 411–426. doi: 10.1016/j.appdev.2007.06.010

Ministry of Children and Equality. (2014). *Family benefits*. Retrieved from http://www.regjeringen.no/en/dep/bld/topics/family-policies/family-policy.html?regj_oss=1&id=670514

Ministry of Education. (2006). *Rammeplan for barnehagens innhold og oppgaver* [Framework plan for content and duties in kindergarten]. Retrieved from http://www.regjeringen.no/nb/dep/kd/dok/lover_regler/reglement/2006/rammeplan-for-barnehagens-innhold-og-opp.html?id=278626

Ministry of Education (2014). *Early childhood education and care policy*. Retrieved from http://www.regjeringen.no/en/dep/kd/Selected-topics/kindergarden/early-childhood-education-and-care-polic.html?id=491283

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). New York, NY: Cambridge University Press.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. New York, NY: Oxford University Press.

Mustard, J. (2006). Experience-based brain development: Scientific underpinnings of the importance of early child development in a global world. *Paediatrics and Child Health*, *11*, 571–572.

Nelson, C. A., & Sheridan, M. A. (2011). Lessons from neuroscience research for understanding causal links between family

and neighborhood characteristics and educational outcomes. In G. J. Duncan, & R. J. Willett (Eds.), *Whither opportunity* (pp. 27–46). New York, NY: Russell Sage.

National Institute of Child Health and Human Development Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development*, *71*, 960-980. Retrieved from http://www.jstor.org/stable/1132337

Obama, B. (2014). *President Barack Obama's state of the union address 2014*. Retrieved from http://www.whitehouse.gov/the-press-office/2014/01/28/president-barack-obamas-state-union-address

Organisation for Economic Co-operation and Development. (2006). *Starting strong II: Early childhood education and care policy*. Paris: Author.

Organisation for Economic Co-operation and Development. (2011). Poverty. *Society at a glance 2011: OECD social indicators*. Paris: Author.

Organisation for Economic Co-operation and Development. (2015). *Norway—Early childhood education and care policy review*. Retrieved from http://www.oecd.org/norway/Early-Childhood-Education-and-Care-Policy-Review-Norway.pdf

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*, *77*(2), 266–280. doi: 10.1111/j.1467-8624.2006.00869.x

Rindfuss, R. R., Guilkey, D., Morgan, S. P., Kravdal, O., & Guzzo, K. B. (2007). Child care availability and first-birth timing in Norway. *Demography*, *44*, 345–372. doi: 10.1353/dem.2007.0017

Rosand, G. M., Slinning, K., Eberhard-Gran, M., Roysamb, E., & Tambs, K. (2011). Partner relationship satisfaction and maternal emotional distress in early pregnancy. *BMC Public Health*, *11*, 161. doi: 1471-2458-11-161

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Ruzek, E., Burchinal, M., Farkas, G., & Duncan, G. J. (2014). The quality of toddler child care and cognitive skills at 24 months: Propensity score analysis results from the ECLS-B. *Early Childhood Research Quarterly*, *28*, 12–21. doi: 10.1016/j.ecresq.2013.09.002

Sæther, J. P. (2010). *Barn i barnehage—foreldrebakgrunn og utvikling de seneste årene* [Children in center care—Parental background and recent developments]. Oslo: Statistics Norway.

Sameroff, A., & Chandler, M. (1975). Reproductive risk and the continuum of caretaking casualty. In F. D. Horowitz (Ed.), *Review of child development research* (Vol. 4). Chicago, IL: University of Chicago Press.

Scheistrøen, J. (2012). *Undersøking om foreldrebetaling i barnehagar, januar 2012* [Investigation of parent fees in child care centers, January 2012]. Retrieved from http://www.ssb.no/utdanning/artikler-og-publikasjoner/undersoking-om-foreldre-betaling-i-barnehagar-januar-2012

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton, Mifflin.

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods. The science of early childhood development*. Washington, DC: National Academy Press.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Squires, J., Bricker, D. D., & Potter, L. (1999). *The ASQ user's guide* (2nd ed.). Baltimore, MD: Brookes.

Tambs, K., & Moum, T. (1993). How well can a few questionnaire items indicate anxiety and depression? *Acta Psychiatrica Scandinavia*, *87*(5), 364–367.

Thal, D. J., O'Hanlon, L., Clemmons, M., & Fralin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(2), 482–496.

UNICEF Innocenti Research Center. (2008). *The child care transition: A league table of early childhood education and care in economically advanced countries*. Innocenti Report Card 8. Florence, Italy: Author.

UNICEF Innocenti Research Center. (2012). *Measuring child poverty: New league tables of child poverty in the world's richest countries* (Report Card 10). Florence, Italy: Author.

van Huizen, T., & Platenga, J. (2015). *Universal child care and children's outcomes: A meta-analysis of evidence from natural experiments*. Discussion Paper Series 15-13. Utrecht, the Netherlands: Utrecht School of Economics, Tjalling C. Koopmans Research Institute.

Winsvold, A., & Guldbrandsen, L. (2009). *Kvalitet og kvantitet—kvalitet i en barnehagesektor i sterk vekst* [Quality and quantity—Quality in a child-care sector in rapid growth]. Oslo, Norway: NOVA.

Yoshikawa, H., Aber, J. L., Bergman, L. R., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioral health of children and youth. *American Psychologist*, *67*(4), 272–284. doi: 10.1037/a0028015

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*. Ann Arbor, MI: Society for Research in Child Development. Retrieved from http://fcd-us.org/resources/evidence-base-preschool

Zachrisson, H. D., Dearing, E., Lekhal, R., & Toppelberg, C. O. (2013). Little evidence that time in child care causes externalizing problems during early childhood in Norway. *Child Development*, *84*(4), 1152–1170. doi: 10.1111/cdev.12040

Authors

ERIC DEARING is a professor of applied developmental psychology in the Lynch School of Education at Boston College and a senior researcher at the Norwegian Center for Child Behavioral Development. His research is focused on the role of children's lives outside of school for their success in school, with a special interest in the ways family, early education and care, and neighborhood conditions affect children's achievement and psychological well-being.

HENRIK DAAE ZACHRISSON is senior researcher at the Norwegian Center for Child Behavioral Development and a professor at the Center for Educational Measurement at the University of

Oslo. His research is on consequences of child care and social inequality for children's development.

ARNSTEIN MYKLETUN is a professor in the Department of Community Medicine at the University of Tromsø, a senior researcher in the Department of Mental Health and Suicide at the Norwegian Institute of Public Health, and a researcher at the Centre for Research and Education in Forensic Psychiatry and Psychology at the Haukeland University Hospital, Bergen, Norway. He is also head of research at the Center for Work and Mental Health, Nordland Hospital Trust, Bodø, Norway. His research is in the areas of epidemiology, public health, psychiatry, and work medicine.

CLAUDIO O. TOPPELBERG is a research scientist at Judge Baker Children's Center and an assistant professor at Harvard Medical School. Dr. Toppelberg's research in child/adolescent psychopathology has two foci: (a) the relations of language, neuro-cognitive, and emotional/behavioral development and (b) the development of immigrant and dual-language children and national childhood policies that affect both.