CESifo
AREA
CONFERENCES 2022

Economics of Education

Munich, 2 – 3 September 2022

Childcare, social skills, and the labor market

Mikko Silliman and Juuso Mäkinen

CESifo

Childcare, social skills, and the labor market

Mikko Silliman, *Harvard University*[*] Juuso Mäkinen, *Aalto University*

August 22, 2022

Link to most recent version

Abstract

This paper provides the first large-scale evidence linking the labor-market effects of childcare programs to social skills measured in adulthood. First, we evaluate the effects of Finland's first national public childcare program and find that small average effects of public childcare access mask considerable heterogeneity; public childcare levels the playing field, reducing the persistence between parent and child income. Second, we show that treatment effects on income are most correlated with treatment effects on adult measures of social competence and almost uncorrelated with effects on fluid intelligence. These results suggest that sustained effects on social skills may be particularly important in explaining the effects of childcare. Additionally, we show that childcare affects earnings by changing the types of jobs people do, and that these effects are only partly explained by shifts in later educational choices.

Keywords: early childhood, social skills, tasks, labor markets, education policy
JEL Classification: J08, I24, J24

[*]*silliman@g.harvard.edu; juuso.p.makinen@aalto.fi.* Mikko is particularly thankful to his doctoral advisors Raj Chetty, David Deming, and Marty West for supporting the project from its inception and to Matti Sarvimäki at VATT for helping to arrange access to administrative data. We also thank Isaiah Andrews, Orazio Attanasio, Andres Barrios-Fernandez, Mark Chin, Christopher Cleveland, Charles Gale, Heikki Hiilamo, Andrew Ho, Hermanni Hälvä, Ramin Izadi, Martti Kaila, Dana McCoy, Luke Miratrix, Maarit Olkkola, Amanda Pallais, Michele Pellizarri, Tuomas Pekkarinen, Miika Päällysaho, Matti Sarvimäki, Diane Schanzenbach, Uta Schönberg, Marko Terviö, Joonas Tuhkuri, Tuuli Vanhapelto, Anna Zink, and Arkadijs Zvaigzne for helpful feedback along the way. This work has benefitted from the comments of participants at seminars at CESifo, Harvard (Economics, Education, and Public Policy), Helsinki (Economics, Education, and Public Policy), IZA, Labore, NHH, the Paris School of Economics, and the Turku School of Economics.

Although there is a consensus regarding the importance of early childhood in shaping long-term outcomes (Heckman, 2006; Yoshikawa et al., 2013; Almond et al., 2017; Black et al., 2017), we still know little about how these effects operate (Duncan et al., 2022). A number of studies have found that childcare programs affect short term outcomes across various domains, that these effects disappear in the medium term, but then re-emerge in measures of adult outcomes (Deming, 2009; Chetty et al., 2011; Heckman et al., 2013; Bailey et al., 2020; Li et al., 2020). A common explanation for this pattern is that the persistent effects of childcare operate through socio-emotional rather than cognitive skills (see, for example, Heckman et al., 2013). However, since linking the effects of childcare to measures of socio-emotional skills from adulthood has been challenging, this hypothesis lacks strong empirical support.

In the first part of the paper, we evaluate the effects of Finland's first national public childcare program on educational and labor market outcomes. In the second, we leverage unique data from the Finnish Defence Forces measuring various dimensions of skills for fifteen nearly full cohorts of Finnish men to study how and why childcare shapes these long-term outcomes.[1]

Our focus is the *Childcare Law of 1973*, which established the first national public childcare program in Finland. Without sufficient resources for the government to provide public childcare in all municipalities right away, only some municipalities could receive funding for public childcare in the first years following the law. Since there were already municipality-provided public childcare programs in urban areas before 1973, we focus on rural municipalities for which the reform provided access to public childcare for the first time. In these areas childcare was provided primarily by mothers as well as a patchwork of informal and private services prior to public childcare access, though even these informal options were often unavailable. We compare cohorts born in the first set of rural municipalities to receive public childcare (treatment) to the same cohorts born in similar municipalities that only received public childcare in later years (comparison). Just a few years after the introduction of the policy, about 35 percent of eligible cohorts in the treatment municipalities attended public childcare while there remained no public childcare in comparison municipalities.

Our empirical approach captures the causal effects of public childcare access if the outcomes of cohorts in treated and comparison municipalities would have progressed in a parallel manner absent the *Childcare law of 1973*. Supporting a causal interpretation, outcomes of children in treatment and comparison municipalities progressed in a parallel manner before 1973, both on average as well as among individuals born to similar types of families. Moreover, the introduction of public childcare does not coincide with regional trends or changes in the composition of families in treated versus comparison municipalities. Since we observe annual municipality-level spots in public childcare

[1] In a companion paper we examine the role of public childcare access on maternal labor market outcomes over the life-cycle, and find economically significant increases in maternal labor market participation that persist to retirement, well past their children's childcare eligibility (Mäkinen and Silliman, 2022).

rather than individual enrollment, this strategy estimates the results of access rather than enrollment.

After the introduction of public childcare, cohorts in treatment municipalities eligible for public childcare experience no average improvements in educational attainment, measures of adult skills, or labor market outcomes. However, these average estimates mask considerable heterogeneity between children growing up in different childhood environments. Our estimates suggest that access to public childcare reduced the association between parent and child income percentile rank by 0.05, with children from the poorest fifth of families improving their income rank by more than two percentiles between the ages of thirty-five and forty, and children from the richest fifth of families experiencing negative effects of almost a comparable magnitude. We find a similar pattern of effects for educational, employment, and marriage outcomes. Effects for females are similar to those for males. These results are robust to the inclusion of controls, to our choice of estimation sample, and to alternative estimators.

Next, we use exceptionally detailed data from the Finnish Defence Forces covering eighty percent of the male population to study whether the effects of childcare on economic outcomes might be explained by lasting effects on social skills (Deming, 2009; Heckman et al., 2013). To root our analysis in theory and facilitate interpretation, we aggregate our measures of skills to three constructs: one which measures social competence, a second which measures academic skills typically developed in school, and a third which measures visual-spatial skills – a major component of fluid intelligence. We then assess the plausibility that each of these candidate mechanisms explains the long-term effects of childcare.

As for labor market outcomes, public childcare levels the playing field in terms of skills across all domains we measure. Providing preliminary evidence that lasting effects on social skills may be particularly important in explaining the effects of childcare on economic outcomes, the effects on social skills are substantially larger than the effects on visual-spatial skills. However, the specific contours of the pattern of effects underlying these linear estimates may differ for each skill and labor market outcome.

To better understand how skill effects map to labor market effects, we study the covariance between treatment effects on labor market outcomes and treatment effects on skills. While these types of exercises are common in the context of teacher value-added (Chetty et al., 2011; Jackson, 2018), such covariances between treatment effects are challenging to estimate in contexts where only a handful of estimates for each outcome are produced. To overcome this challenge, we follow an insight from Angrist et al. (2022) and study the relationships between shorter and longer-term treatment effects across subgroups. To both tie our hands as we form a large number of granular subgroups – reducing researcher degrees of freedom – and to maximize treatment effect variation across subgroups, we base these groups on predicted treatment effect heterogeneity using the machine learning framework from Chernozhukov et al. (2021). We then estimate split-sample correlations between treatment effects on long-term outcomes and treatment effects on mediating outcomes across these subgroups. We complement these estimates of treatment effect correlations with results from a decomposition-based approach to mediation (Imai et al., 2010; Heckman et al., 2013).

Treatment effects on income are most correlated with treatment effects on social competence (r = 0.50), less correlated with treatment effects on academic skills (r = 0.28), and almost uncorrelated with treatment effects on visual-spatial skills (r = 0.07) – the closest measure to fluid intelligence or IQ. These results are robust to several potential biases, and treatment effect correlations between years of education and skills exhibit a similar pattern. Further, decomposition-based estimates corroborate these results, suggesting that skills account for upwards of fifty (seventy) percent of effects on income (education), and that conditional on social skills fluid intelligence has no explanatory power for the long-run effects of public childcare. Interestingly, this pattern of results is strikingly different from the raw correlations, where these three skills exhibit roughly similar correlations with adult income rank. These results suggest that, amongst candidate mechanisms, lasting effects on social skills are most likely to underlie the labor market effects of childcare access.

Social skills could either affect economic outcomes by directly increasing productivity in interpersonal tasks (Weidmann and Deming, 2021) or indirectly by improving educational outcomes (Heckman et al., 2013; Johnson and Jackson, 2019). Since childcare in 1970's Finland had no emphasis on academic learning, the effects we see on academic skills and education both suggest the possibility of indirect effects of social skills operating through education. This idea is also supported by the overlap in effects on both social and academic skills. However, results from decomposition exercises suggest that social skills play a role in explaining effects on income above and beyond academic skills or education. To further understand how these effects on skills shape earnings, we study whether childcare affects earnings by changing the types of occupations people where people work, or by shifting people's productivity in jobs they would have done anyway. We show that childcare shifts the mean income of people's occupations without changing their income rank within occupations. Further, we see that people's skills in early adulthood are linked to the task-content of their work nearly twenty years later (Acemoglu and Autor, 2011a; Speer, 2017), and that skill effects are intimately linked to effects on occupational tasks. Together, these results suggest that childcare shifts earnings by changing the types of jobs where people work.

The variation in the effects we find by family income may be explained by the quality of early childhood socialization in the absence of access to public childcare (Kline and Walters, 2016). In our context, access to public childcare increases maternal labor force participation, suggesting that public childcare at least partly substitutes for maternal care. Moreover, effects on maternal labor force participation are correlated with effects on children's skills. This could be because without resources for a nanny to make maternal work possible, public childcare is more likely to substitute for maternal care in poor families, often settings with less parental attention (Guryan et al., 2008;

Falk et al., 2021). In contrast, for higher income families public childcare is more likely to substitute for the personalized and attentive care of a nanny. To further probe the role of early childhood interactions in explaining our results, we study heterogeneity related first-born status (Price, 2008; Black et al., 2018), a characteristic known to be tied to parental attention within families. We replicate the finding from Black et al. (2018) showing higher levels of skills amongst first-borns, and then show that access to public childcare lowers the first-born advantage – particularly in terms of social competence. Additionally, our machine learning predictions of treatment effect heterogeneity suggest that public childcare is likely to benefit children with lower quality socialization in their homes – and particularly those whose mothers are likely to work regardless.

The results from this study extend the existing literature in several ways. First, our study provides some of the first large-scale evidence that public childcare can shift long-term measures of social skills. Existing research in economics and psychology provides a handful of estimates of the effects of childcare on detailed measures of socio-emotional skills measured in childhood (Weiland and Yoshikawa, 2013; Drange and Havnes, 2019; Ichino et al., 2019; Cappelen et al., 2020). In adulthood, treatment effects on socio-emotional skills are typically proxied by behavioral outcomes such as dropout, teenage pregnancy, and crime (Deming, 2009; Heckman et al., 2013; Sorrenti et al., 2020). Our study examines the effects of childcare on detailed measures of adult social and emotional skills measured at age nineteen for nearly full Finnish male cohorts. Our results suggest that access to public childcare affects skills across a range of domains, particularly those linked to social competence, and that these skills persist through adulthood.

Perhaps most importantly, we show that the effects of public childcare on long-term outcomes are most strongly related to effects on adult measures of social skills, compared to those on academic skills or fluid intelligence. These results lend empirical support for the hypothesis that behavioral or socio-emotional skills drive the effects of childhood programs on long-term outcomes (Deming, 2009; Heckman et al., 2013; Bailey et al., 2017). Further, our results suggest that social skills affect economic outcomes both directly (Weidmann and Deming, 2021) as well as through a dynamic complementarity with education (Heckman et al., 2013; Johnson and Jackson, 2019). Connecting effects on social skills to long-term outcomes, our paper provides a basis for optimism that recent interventions targeting social skills can lead to meaningful economic gains in adulthood (Alan et al., 2019; Berger et al., 2020; Cappelen et al., 2020; Kosse et al., 2020; Sorrenti et al., 2020; Algan et al., 2022). Our results also suggest that social skills are likely to affect income by changing the types of jobs people do (Speer, 2017). These results provide encouraging evidence that effective early childhood programs may be able to generate skills that help people keep up with tasks demanded by the future of work (Acemoglu and Autor, 2011b; Deming, 2017).

Together, our results help to inform the debate on whether to provide targeted or universal public childcare (Decker and Kelly, 2022), providing new evidence on the effects of a national public childcare program. As policy-makers increasingly look to public childcare as a promising tool to help improve economic well-being and reduce inequalities (European Union, 2019; Biden, 2021), empirical evidence on the effects of national childcare programs remains mixed. While evaluations of early Head Start cohorts and some other state-level programs in the United States find considerable benefits (Ludwig and Miller, 2007; Deming, 2009; Carneiro and Ginja, 2014; Barr and Gibbs, 2022), later cohorts appear to experience zero or even negative effects of attending Head Start (Pages et al., 2019). Similarly, evidence on universal childcare programs from Denmark, Canada, Germany, Italy, and Norway also suggests mixed effects of public childcare (Gupta and Simonsen, 2010; Havnes and Mogstad, 2015; Kottelenberg and Lehrer, 2017; Ichino et al., 2019; Cornelissen et al., 2018). In contrast, evaluations of small-scale high-intensity programs (e.g. Heckman et al., 2010) as well as city-level programs (Gray-Lobe et al., 2021) document large improvements in a range of outcomes.[2] Our results suggest that although Finland's first universal public childcare program produced small average effects, it leveled the playing field substantially. We show that this pattern of results is likely due to the relative quality of early childhood socialization in the absence of public childcare. These results suggest that childcare programs can generate positive effects by either targeting children from more disadvantaged home environments or investing in the quality of universal programs.

After discussing the institutional context (Section 1), we outline the data and measures we use (Section 2), detail the empirical approach we take and report the reduced form results (Section 3), and delve into the role of skills in explaining adult outcomes (Section 4). We conclude with a short overview of our findings, pointing to areas for future work.

1 Institutional context

The foundation for Finland's first national public childcare program – still in place today – was laid by the *Childcare Law* of 1973 (Law 36/1973).[3] After being in the works for nearly a quarter of a century (Alila et al., 2014), a proposal for a law concerning childcare was presented in parliament in 1972. With public childcare only available in cities, these legal proceedings emphasize the urgency

[2] See also Baker (2011), Elango et al. (2016), and Duncan et al. (2022) for overviews of prior work on public childcare.

[3] The seeds of childcare provision in Finland were laid in 1919 under the auspices of social services, and by the 1920's and 1930's the first laws formalizing the government role came into place. In 1922, the *Poverty-care Law* (Law 145/1922) provided a legal basis for national support for childcare–but primarily for those with special needs or disabilities (Alila et al., 2014). As Alila et al. (2014) describe, this law provided support primarily for children that were mentally disabled, blind, deaf, or physically disabled. Still focused on children with special needs and disabilities, the *National Childcare Funding Law* of 1927 (Law 296/1927) provided government funding to individual childcare centers through application on the basis of demonstrated need. And, in 1936, the *Child-protection Law* (Law 80/1936) stipulated that municipalities must make efforts to supply childcare or support private childcare provision for children growing up in poverty or in unsafe home environments. Political gridlock made it impossible to make progress on childcare for the next four decades.

of public support for childcare and highlight the "variability in quality and uneven geographic distribution of childcare" (Valtiopaivat 1972). Perhaps most importantly, both the parliamentary proceedings themselves as well as commentary from the time period (Hulkko, 1971) suggest the increasing labor market participation of women was an important factor behind the newfound support for public childcare. In parliament, advocates of the new law cited demographic and cultural changes that resulted in the demand for childcare had far outstripped the supply: "...employment rates of the mothers of young children have increased. The economic and demographic changes, as well as the increased time spent in education, have increased the demand for childcare" (Valtiopaivat 1972). Following decades of political gridlock, the law was just barely passed in parliament following an extended sitting.

The law was implemented quickly and on April 1st, 1973 the *Childcare Law* (Law 36/1973) made the provision of childcare a universal right, unified concepts surrounding childcare, and provided a transparent and simple mechanism for the government funding of childcare. To facilitate the expansion of public childcare, the national government agreed to provide funding for the fixed costs of establishing childcare centers and cover up to 80% of the annual costs, depending on the municipality's ability to pay. Given the enormous cost and time required to train qualified staff to supply the estimated 100,000-120,000 spots demanded, the government planned to grow childcare coverage by 5,000 annually until the year 1990 (Valtiopaivat 1972).

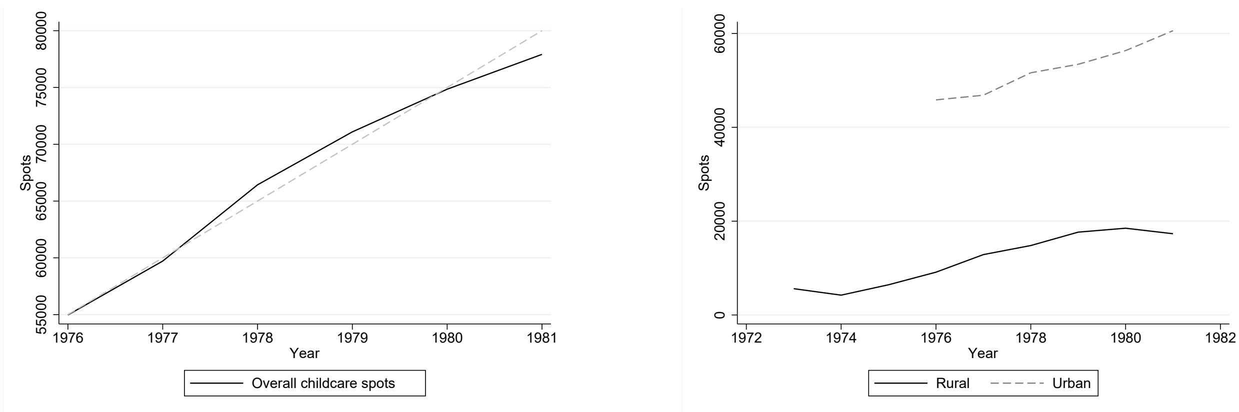
As shown in Figure 1a, the number of government funded daycare spots grows at almost exactly the planned rate of 5,000 a year after 1973. While there was some subsidized childcare in cities prior to 1973, the law made public childcare available in rural areas for the first time (Figure 1b). Given that these rural areas are where the expansion of childcare grew most, rural municipalities are the focus of this paper (Figure 1b).

Some of these municipalities had both half and full day care available, and some childcare centers also provided free lunch. Public childcare centers operated for ten months a year. Despite its roots in social services, the potential importance of childcare for child development was acknowledged already in this period: a publication from the Finnish Population and Family Welfare League argues that "the work implies participation in productive activity, since it constitutes the production of coming labor power" (Hulkko, 1971). Although the concepts surrounding quality in early childhood education and care directly after 1973 were still developing, maximum group-size limits were in place to ensure that childcare centers were sufficiently staffed.[4] Still, the services these childcare centers provided in these early years was likely of considerably inferior quality compared to modern

[4]In 1980, the first year that we can accurately locate childcare teachers in the census, we see that childcare teachers in this period were primarily young married women, often with childcare-age kids themselves (Appendix 4). Most of these women had completed some post-secondary education. For the latest cohorts in our sample, non-subsidized childcare options began to become available even outside municipalities that received public funding for childcare (in our comparison group).

Figure 1: The Expansion of Public childcare, 1973-1981

(a) New childcare spots (b) Urban and rural municipalities



Notes: These figures show data on the growth in public childcare spots following the *Childcare Law* of 1973. Figure (a) shows that the annual increase in childcare spots in the data corresponds to almost exactly 5,000 spots annually (scenario in gray) – the target number in the parliamentary proceedings from 1972. These years (1976-1981) are the only years that public childcare data is available for all municipalities (urban and rural). Figure (b) shows the annual number of public childcare spots by urban and rural status. Prior to 1973 there was almost no public childcare available in rural areas. This is set of municipalities is the focus of our paper.

childcare. It was only outside our window of study, after the next wave of childcare reform in Finland in 1983, that the role of childcare began to more formally shift from social care to child development and education (Alila et al., 2014).[5]

Prior to 1973, some urban municipalities had developed public childcare infrastructure, but families outside urban areas had little access to childcare. In these areas, childcare was provided primarily by mothers as well as a patchwork of informal and private services, though even these informal options were mostly unavailable. In affluent families, it was common for children to be taken care of by nannies. The notable exception to this was children with special needs or disabilities, for whom public childcare was provided first through the *Child-protection Law* (Law 80/1936).

[5]After its birth, the next major period of childcare reform took place between the years 1984-1996. During these years, childcare became a subjective right, first for children under the age of three (1990), and then for all children not yet in school (1996) (Alila et al., 2014, pg. 13). Further securing its position as a universal right integral to the operation of the Finnish welfare state, the legal basis for both home-care and private-care became linked to the *Childcare Law* of 1990. Today, the effects of public childcare access in Finland remains hotly debated by academics and policy-makers (Erola, 2018; Erola et al., 2020). This debate emerged after a pair of papers found participation in public childcare to be associated with positive outcomes (Karhula et al., 2017; Hiilamo et al., 2018) while another paper found there to be no association between learning outcomes and childcare participation (Saarinen et al., 2019). However, as one of the authors of these studies themselves notes, a potential reason for the discrepancies in these results is that these studies lack an experimental or quasi-experimental setup (Erola, 2018).

2 Data, concepts, and descriptive statistics

2.1 Data sources and outcomes

We link together various sources of data for cohorts born between 1962 and 1976. This allows us to include several cohorts aged three to six – the initial ages for childcare eligibility – before and after the *Childcare Law of 1973* was passed.

Childcare data. We begin with municipal-level data on public childcare. After the passing of the *Childcare Law of 1973*, data on childcare provision was collected annually by the research and planning division of the Association for Finnish Municipalities and reported in their annual reports on social spending and services for the years 1973-1981 (Association for Finnish Municipalities, 1974; 1975; 1976; 1977; 1978; 1979; 1980; 1981; 1982). After the year 1981 the statistics are no longer reported in a consistent format that would allow for year to year comparisons. We transcribed these manually from reports located at the archives of Statistics Finland. These reports include statistics on the number of spots for children three to six years old in municipal childcare centers. Since the administrative unit in the early seventies was different for municipalities classified as urban and rural, these data do not include urban municipalities for the years 1973-1975 (See Figure 1b).

Background characteristics. We link this municipality data to individual data from Statistics Finland's FOLK database (from Statistics Finland, 2021c) detailing each individual's gender as well as their year and municipality of birth. We then merge this data to a register containing parent-child links to identify the fathers and mothers of all individuals, and create measures of family composition (Statistics Finland, 2021c). Population-wide censuses from 1970-1985 contain data on parental education and income (Statistics Finland, 2021b). We form measures of family income rank based on cohorts from their childrens' birth year based on the full (not estimation) sample.[6] Since these data contain detailed information on occupations and places of work, we are able to identify childcare professionals in the data – providing us some information detailing the treatment and counterfactual. Unfortunately, these professionals are only identified for the year 1980, and we cannot differentiate between professional childcare providers working in public and private childcare centers or less formal family-run childcare centers.

Educational and labor market outcomes. National degree registries (Statistics Finland, 2021a) provide us information on educational attainment for everyone in our sample. We construct simple binary measures of secondary school dropout, upper-secondary general track graduation, upper-secondary vocational certification, and tertiary completion. We also aggregate these measures to

[6] Note that as opposed to some other measures of family income rank which average family income through several years of childhood (Pekkarinen et al., 2017), we only include income measured prior to the reform, meaning that our measure likely contains more measurement error. Still, as noted by Kitagawa et al. (2018), rank based measures of mobility are subject to less problems than, for example, log-based measures.

measure years of education. We use the FOLK databases to then generate annual measures of cohort income rank and employment. To measure income, we form a measure of mean cohort income rank (ranging from 0-1) of incomes between the ages of 35 and 40 – typically a good proxy for lifetime income (Bhuller et al., 2017). To measure employment, we take the mean years of employment between thirty and forty (0-10). We create a measure for whether each person in our data is observed married at any point by the time they reach forty.

Measures of adult skills. An exceptionally large and detailed data-set from the Finnish Defense Forces documents various dimensions of skills for everyone entering the military (Finnish Defence Forces, 2021). Due to national conscription for all male citizens, these measures—collected at age nineteen—are available for upwards of eighty percent of males from the cohorts we study. Our binary measure – "Military service" – measures whether such skill data exists for each individual. These data were collected upon conscription using testing instruments designed by psychologists that remained the same for all cohorts we study. This test includes three dimensions of cognitive skills—arithmetic, verbal reasoning, and visual-spatial skills—as well as several dimensions of socio-emotional skills including activity energy, achievement striving, deliberation, dutifulness, leadership motivation, self-confidence, and sociability. We standardize all test scores to have a mean of zero and a standard deviation of one for the cohort born in 1967, anchoring all other cohorts to this year. See Appendix Section 3 for more details Nyman et al. (2007) or Jokela et al. (2017) for an extensive overview of this data. We report results for each of these outcomes in the appendix of the paper, but focus on a set of constructs motivated by the literature on child development and economics. Although these measures are taken at age nineteen – still early in adulthood – these types of skills are understood to be relatively stable through later adulthood (Cobb-Clark and Schurer, 2012).

Occupational task shares. We follow Silliman and Virtanen (2022), linking occupational task share measures from Acemoglu and Autor (2011a) to four-digit ISCO occupational codes measured between the ages of 35-40. These allow us to measure the extent to which individuals in our study end up in jobs using cognitive and social skills.[7]

[7] Acemoglu and Autor (2011a) defined the labels of these occupational tasks within a context where they study the effects of technological change – and thereby emphasized the routine and non-routine content of tasks. To simplify the interpretation of these measures in our context we re-label some of the measures to use more consistent terminology. "Non-routine cognitive personal" is renamed "Social non-routine analytic", "Non-routine manual personal" is renamed "Social non-routine manual", "Non-routine cognitive analytic" is renamed "Cognitive non-routine", "Non-routine cognitive analytic" is renamed "Cognitive non-routine", and "Non-routine manual physical" is renamed "Manual non-routine".

2.2 Concepts and measurement

We organize our paper to test hypotheses from the literature on child development in economics and psychology.

Economists have argued that early childhood programs shape long term outcomes primary through social – as opposed to cognitive – skills (Deming, 2009; Chetty et al., 2011; Heckman et al., 2013). Empirical studies report a pattern of results where childcare programs have positive effects on early measures of both learning outcomes and behavioral skills, exhibit no effects on later measures of achievement, but improve long-term economic outcomes (Heckman and Rubinstein, 2001; Gibbs et al., 2011).

Understanding how early childhood programs shape people's later behavior has also been a central goal of research in psychology. Psychologists understand a child's socialization both at home and in childcare, to play an important role in this process (Clausen, 1966; Baumrind, 1967). Waters and Sroufe (1983) argue that *social competence* – the ability to recruit personal and interpersonal resources in the context of goal achievement – is the central organizing construct of early childhood. Since then, social competence has played an important organizing role in early childhood research (Dodge et al., 1986; Rose-Krasnor, 1997; Bost et al., 1998; Campbell et al., 2000; Denham et al., 2003; Ladd, 2005; Vaughn et al., 2009). Vaughn et al. (2009) describe that social competence consists of three parts: i) behavioral and cognitive skills for successful goal achievement in social contexts; ii) the ability to discover the goals of interactive peers; iii) the understanding of a child's relative value as a preferred playmate. Gunderson et al. (2013) describe one nice example of how such skills might develop, focusing on how parental praise can lead to persistent improvements in the self-confidence and motivation of young children.[8]

To root our empirical analysis in theory, we aggregate our measures of skills to three constructs: one which measures social competence, a second which measures academic skills typically developed in school, and a third which measures visual-spatial skills – a major component of fluid intelligence. While these are the primary constructs we study, we complement these by reporting results for raw measures from the Finnish Defence Forces data in the Appendix.

Social competence. The measures from the Finnish Defence Forces that map most closely to the concept of social competence are achievement striving, leadership motivation, and self-confidence.[9] We take the average of each child's standardized score across these measures to define their social

[8]In its emphasis on achievement striving and motivation, social competence has strong conceptual links to well known psychological concepts outside child development per se, including growth mindset (Dweck, 2006) and grit or perseverance and passion for long-term goals (Duckworth et al., 2007).

[9]Sociability, another measure collected by the Finnish Defence Forces, measures a person's gregariousness and preference for socialization. This measure has little information on how well a person navigates social situations in the context of goal achievement. As such, it is not included in our measure of social competence. However, adding it to the measure of social skills does not affect our results, and we report estimates for all individual concepts separately in the Appendix.

competence. To ease interpretation, we standardize this measure to have a mean of zero and a standard deviation of one. This measure is intended to gauge the hypothesis from developmental psychology and economics that the social competences developed in early childhood may explain the effects of public childcare on long-term outcomes (Waters and Sroufe, 1983; Deming, 2009).

Academic skills. Similarly, we create a blanket measure of academic skills by taking the mean of each child's arithmetic and verbal scores also standardized to have a mean of zero and a standard deviation of one. These skills map closely to the concept of school readiness often discussed in the literature on early childhood educcation (Duncan et al., 2022). This skill is included to test the hypothesis that, through dynamic complementarity, public childcare may shape long-term outcomes by facilitating academic learning (Heckman et al., 2013; García et al., 2021).

Visual-spatial skills. To test the hypothesis that childcare might not affect intelligence unrelated to academic learning, we include the measure of visual-spatial skills from the Finnish Defence Forces, again standardized to have a mean of zero and standard deviation of one. This measures fluid intelligence similar to Raven's matrices. We use this measure to see if the fadeout of cognitive skills affected by childcare may be explained by the fact that the effects on fluid intelligence remain small and potentially unrelated to long term outcomes (Deming, 2009; Chetty et al., 2011; Heckman et al., 2013).

The conceptual framework for our paper is laid out more thoroughly in Appendix Section 4. We also report results for all disaggregated outcomes available in the Appendix.

2.3 Descriptive statistics

Merging these data together provides us with a data set covering full cohorts born in the years 1962-1976 which includes information on family background, birth, educational attainment, and labor market outcomes through age forty. Additionally, the data-set includes measures of adult skills for eighty percent of the male population. Altogether, this data set spans 463 municipalities, and covers 928,500 individuals.

The analysis in this paper, described in more detail in Section 3, will be based on comparing the first set of municipalities that receive public childcare spots following the *Childcare Law of 1973* (treatment) to municipalities that only come to receive public childcare spots in later years (comparison). Since we can only estimate our full set of results for males (few women are included in the Finnish Defence Forces data), the primary sample used in our estimates only includes males in the above-mentioned set of municipalities. However, female siblings are included in our counts of siblings, and we report labor market estimates for females separately in the Appendix.

Family background. Table 1 presents the mean background characteristics of full Finnish cohorts (Column 1) and our estimation sample (Column 2). Given that military data on skill outcomes is only available for males, we also present mean characteristics separately for males in both the full and estimation sample (Columns 3 and 4). Since the expansion of public childcare took place outside of major urban areas, the estimation sample differs markedly from the full sample. Compared to the full sample, families in our estimation sample tend to be larger, poorer, and have less educated parents.

Outcomes and family background. Appendix Table 3 shows how Family income is related to long-term outcomes in our estimation sample. The first column presents mean outcomes for children from the poorest fifth of families in our sample, while the second column presents mean outcomes for children from the richest fifth of families in our sample. These data suggest large family-income based gaps in long-term outcomes for all outcomes in our data: children from the poorest fifth of families score 0.4 SD lower on academic achievement, and end up with incomes that rank 12 percentiles lower in the adult income rank between the ages of 35 and 40. Appendix Table 1 presents the mean outcomes for the full and estimation samples. Likely due to the exclusion of large cities from our estimation sample, individuals in our sample earn less and are slightly less educated than individuals in the full sample.

Skills and long-term outcomes. Both cognitive and socio-emotional skills measured in adulthood are strongly correlated with long-term outcomes such as income and educational attainment. Correlations between skills and income for our estimation sample are shown in Table 2.

Table 1: Estimation sample versus full sample: Family background

			Males	
	Full sample	Estimation	Full	Estimation
	(1)	(2)	(3)	(4)
Mother's education	10.50	10.23	10.52	10.24
	(2.34)	(2.12)	(2.35)	(2.12)
Father's education	10.73	10.25	10.74	10.26
	(2.57)	(2.22)	(2.58)	(2.22)
Mother's age at first birth	23.71	23.64	23.72	23.66
	(4.28)	(4.35)	(4.28)	(4.35)
Family size	2.00	2.11	1.99	2.10
	(1.04)	(1.12)	(1.04)	(1.11)
Family income percentile	49.94	43.12	50.02	43.10
	(28.95)	(27.86)	(28.98)	(27.88)
Lowest income decile	0.10	0.13	0.10	0.13
	(0.30)	(0.34)	(0.30)	(0.34)
Highest income decile	0.10	0.06	0.10	0.06
	(0.30)	(0.24)	(0.30)	(0.24)
Grandparent present	0.48	0.61	0.48	0.61
	(0.50)	(0.49)	(0.50)	(0.49)
Municipalities	463	229	463	229
Individuals	928,500	177,808	472,591	90,434

Notes: This table reports the means and standard deviations of the background characteristics for the full and estimation samples in this paper (Columns 1 and 2) and males (Columns 3 and 4).

Table 2: Correlations between skills (age 19) and adult income rank (ages 35-40)

	Income rank	Visual-spatial	Academic	Social competence
Income rank	1.000			
Visual-spatial	0.283	1.000		
Academic	0.300	0.693	1.000	
Social competence	0.256	0.365	0.422	1.000

Notes: This table is based on the estimation sample, and reports the correlations of our three primary skill outcomes with adult income rank. N= 90,434.

3 Evaluating the effects of access to public childcare

3.1 Empirical approach

The primary challenge in estimating the effects of access to childcare on later life outcomes is that municipalities that offer access to childcare may be different from municipalities that do not (childcare investments are *endogenously determined*). For example, while urban areas tend to have much greater access to childcare than rural areas, families living in urban areas differ from rural families in numerous ways, and growing up in a densely populated city might affect a child's trajectories through life through more channels than simply access to childcare.

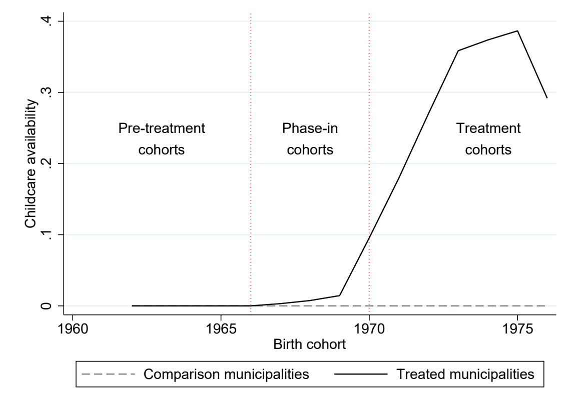
We focus on changes in the geography of public childcare availability in the years immediately after the *Childcare Law of 1973*, a law that passed suddenly after decades of political gridlock. As a result of the *Childcare Law of 1973*, the government provided resources to fund 5,000 new childcare spots a year, so that by the 1990's, all children in Finland aged 3-6 would have access to public childcare—irrespective of the municipality in which they were born. However, for the initial years after 1973, while children growing up in some areas could access public childcare, children from similar families in other municipalities had no access to public childcare. This change in the geographic availability of public childcare based on the swift passage of the *Childcare Law* provides us with the basis for our empirical strategy. To estimate the effects of access to public childcare, we compare the adult outcomes of cohorts who are differentially exposed to public childcare access based on their cohort and municipality of birth.

In our main analysis, we consider the first municipalities to receive access to public childcare after the policy to be our treatment group, and compare their outcomes to the set of municipalities that remains untreated for the entire duration we study. This simple binary two-by-two differences-in-differences approach alleviates potential concerns arising from staggered designs (Goodman-Bacon, 2018) and complications arising from continuous treatment measures (Callaway et al., 2021). To take advantage of the full set of rural municipalities as well as variation in the intensity of treatment across municipalities, we complement our main analysis with stacked binary and continuous differences-in-differences designs. Our approach to these designs is described at the end of this section, and results from these designs are reported in the Appendix.

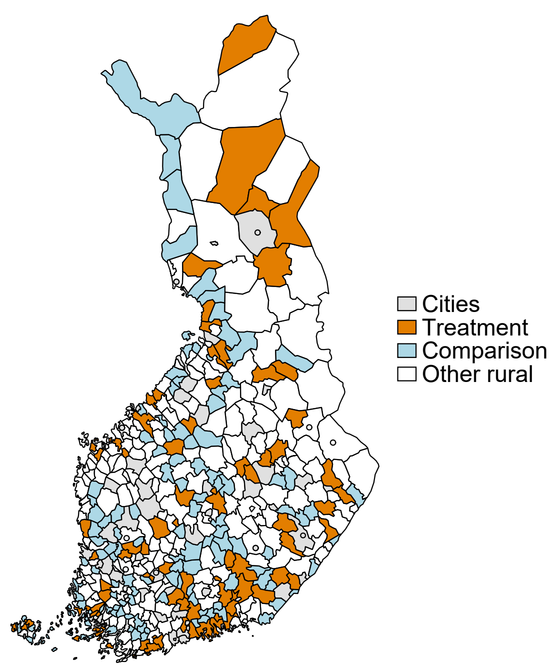
Regardless of the rural municipality they were born in, members of the 1962 cohort had no access to public childcare. As a result of the 1973 *Childcare Law*, children born in the set of treatment municipalities in 1970 or later could access public childcare for the full period between the ages of 3-6. Those born between the years 1967-1970 might have been able to attend public childcare for at most a portion of this period (phase-in period). Figure 2a shows childcare availability by birth cohort in treatment and comparison municipalities. As shown in Figure 2b, the roll-out of childcare

Figure 2: Availability of public childcare in treatment and comparison municipalities

(a) Places per cohort



(b) Geography of public childcare



Notes: Figure (a) reports the availability of childcare spots compared to the number of three to six year olds in birth-cohorts in treated (N=89) and comparison (N=134) municipalities. Figure (b) shows the geographic distribution of treatment (orange) and comparison (blue) municipalities. In addition to the municipalities in our estimation sample, rural municipalities that were in the process of expanding childcare amidst our period of study are shown in white and urban municipalities are shown in gray.

spots does not follow simple regional geography – both treated and comparison municipalities are distributed across the country. Just a few years after the introduction of the policy, upwards of thirty-five percent of children aged 3-6 were attending public childcare in treated municipalities, while no-one was attending public childcare in comparison municipalities.

Average treatment effects. In the most simple empirical operationalization of this approach, we estimate the effects of access to public childcare using the following specification:

$$Y_{imc} = \beta(FIRST_m \times POST_c) + \delta(FIRST_m \times PHASEIN_c) + \pi_m + \gamma_c + e_i \tag{1}$$

In the above equation, we regress individual (i) outcomes (Y) in municipality m and cohort c on an indicator variable for whether or not the municipality belonged to the first set of municipalities covered by the 1973 policy ($FIRST$), and whether the child was aged 3 years old in the period after the policy was implemented ($POST$) (cohorts are born between 1970 and 1976). Since some children are already four, five or six when the policy was implemented (cohorts born between 1967 and 1969) and may have also enrolled in childcare, we remove any effect on these cohorts from our primary coefficient of interest by adding an interaction between $FIRST$ and $PHASEIN$.[10] We account for consistent differences between children born in different municipalities (π_m) and cohorts (γ_c). Standard errors are clustered by municipality in all our analysis (Bertrand et al., 2004).

The coefficient of interest, β, is our difference-in-differences estimate of the effects of access to public childcare on outcome Y. The first difference measures the extent to which the outcomes of post-period cohorts vary from prior cohorts within their own municipalities. The second difference measures the extent that this within municipality variation differs between treated and comparison municipalities.[11]

We also adapt the above specification to produce annual estimates of any differences in outcomes between the first set of municipalities to receive access to public childcare and our set of comparison municipalities. This event-study specification is estimated by the following equation:

$$Y_{imc} = \sum_{c=1962}^{1976} \beta_c \Big(\mathbf{1}[c_i = c] \times FIRST_m \Big) + \pi_m + \gamma_c + \epsilon_i \tag{2}$$

The term β_c measures the extent to which the outcomes between the treatment and comparison sets of municipalities differ in outcomes in each year before and after the policy, taking into account

[10]Including these cohorts is likely to bias our TE downward, since they would have been exposed to childcare for a shorter period of time, if at all. This approach is similar to that taken by prior work on the gradual implementation of policies (Havnes and Mogstad, 2011b, 2015).

[11]Relating our empirical approach to the conceptual framework from the Appendix Section 4, we can imagine the policy (the term, $FIRST_m \times POST_c$) to result in a shock to public investment in childcare (D). The parameter, β, relating outcomes to shocks to D, may reflect endogenous changes in household provision of childcare that result from increases in public provision.

initial differences in outcomes as well as annual variation in outcomes affecting both treatment and comparison municipalities.

Treatment effect heterogeneity. Building from prior research suggesting that the effects of public childcare may vary significantly by the type of family that a child is from – for example by family income (Havnes and Mogstad, 2015), we modify our main specification to allow us to study heterogeneity in any effects of public childcare. For our main estimates of heterogeneous treatment effects we focus exclusively on heterogeneity by family income percentile and assume a linear relationship between family income percentile and the magnitude of the treatment effect. In the second half of the paper, we use this same equation to study heterogeneity predicted by our full set of background characteristics.

$$Y_{imc} = \beta_1(FIRST_m \times POST_c) + \beta_2(FIRST_m \times POST_c \times HET_i) + \tag{3}$$

$$\lambda HET_i + \delta(FIRST_m \times PHASEIN_c) + \pi_m + \gamma_c + e_i$$

The above equation is identical to Equation 1, but includes an additional term (HET_i) for a measure of heterogeneity both alone and interacted with treatment status.

Equation 3 provides us information about how access to public childcare shifted the relationship between family characteristics (family income) and childrens' outcomes. An assumption underlying this model is that there was a linear relationship between family characteristics and childrens' outcomes, and that any effects of public childcare on childrens' outcomes shifted the slope of that relationship. However, it is entirely possible that the relationship, and particularly the change in the relationship may be non-linear.

Both to test for whether treatment effects are indeed linear to the measure of heterogeneity as well as to form granular heterogeneity-group (g) specific estimates of treatment effects, we also estimate our model separately at for children for different types of families:

$$Y_{imcg} = \sum_{g=1}^{n} \beta_g(FIRST_m \times POST_c \times \mathbf{1}[g_i = g]) + \sum_{g=1}^{n} (\mathbf{1}[g_i = g]) + \tag{4}$$

$$\delta(FIRST_m \times PHASEIN_c) + \pi_m + \gamma_c + e_i$$

A benefit of this approach is that it relaxes the assumption of linear treatment effect heterogeneity (Løken et al., 2012), and allows for unique treatment effects for each heterogeneity group (g). These granular estimates of subgroup treatment effects will also provide a key component for our analysis of the associations between treatment effects across various outcomes (ex. skills and income). In our later analysis, we will estimate this equation with the number of groups ranging from ten to a hundred.

To test for whether parallel trends hold for children of different types of families, we also modify our event-study model (Equation 2 as follows:

$$Y_{imc} = \sum_{g=1}^{G} \sum_{c=1962}^{1976} \beta_{cg}\Big(\mathbf{1}[g_i = g] \times \mathbf{1}[c_i = c] \times FIRST_m\Big) + \pi_m + \gamma_c + \epsilon_i \tag{5}$$

The above equation produces one coefficient each cohort (c) for each group (g). This allows us to plot event-study pictures for different types of families (ex. rich vs. poor) separately.

Internal validity. The interpretation of the coefficients of interest as reflecting the causal effects of public childcare access rests on the assumption that, in the absence of public childcare, the outcomes of treated and comparison municipalities would have developed in a parallel manner. While we cannot observe what would have occurred in the absence of the policy – the identifying assumption does provide testable implications.

To probe the validity of our approach empirically, we study whether the outcomes of treatment and comparison municipalities are parallel prior to treatment (Figure 3) and examine whether changes in observable characteristics of families in treatment and comparison municipalities coincide with the introduction of public childcare (Table 3). Figure 3 suggests that there no are changes in the difference in mean income rank between treated and comparison municipalities before treatment municipalities receive access to public childcare. Appendix Figures 1-3 suggest a similar story holds when we examine educational, labor market, and skill outcomes (see also Appendix Figures 5-7, which present event-study diagrams based on a richer set of covariates). While we do see that the first municipalities to receive public childcare were slightly larger and more affluent than those receiving public childcare in later years, Table 3 suggests that changes in the observable characteristics of families in treatment versus comparison municipalities do not coincide with the introduction of public childcare.

Additionally, to test for the sensitivity of our results to pre-existing regional trends, we modify our estimating equations to include both parametric and non-parametric measures of regional trends. Suggesting that our estimates are not a result of regional trends, our estimates are insensitive to these modifications. To ensure that any heterogeneity we are picking up is due to heterogeneity in effects between family rather than municipality characteristics, we also modify our main estimating equation measuring heterogeneity to include year-by-municipality fixed effects. This triple-differences design compares changes in between family but within municipality variation between treatment and comparison municipalities.

Alternative estimators. While we use a simple two-by-two differences and differences approach for our main estimates, we complement these estimates of both average treatment effects and treatment effect heterogeneity with estimates from alternative estimation strategies that take advantage of richer variation in the extent childcare adoption as well as a much larger set of municipalities. We do this with both binary measures of treatment and continuous measures of treatment intensity — the portion of 3-6 year old children covered by public childcare. For both of these approaches, we avoid some issues regarding negative weighting in staggered differences-in-differences designs (Goodman-Bacon, 2018) by stacking our data so that the comparison group is always the set of municipalities that never receives treatment in our follow-up window. For these estimates we only consider children aged 3-6 for the entire post-reform period as treated and do not include the phase-in dummy, which we find to be irrelevant for our main estimates as well.

By including a broader range of municipalities, these estimators help alleviate concerns that the set of municipalities included in our main estimates might explain the results. And, by taking advantage of the extent of childcare access available within municipalities, we are able to assess the natural idea that the effects of childcare access should vary by the portion of children covered by public childcare.

Table 3: Descriptive data and covariate balance

	Pre-period mean		
	Treatment	Comparison	DiD
	(1)	(2)	(3)
Mother's education	10.01	9.92	0.08
	(2.01)	(1.92)	(0.07)
Father's education	10.11	9.96	0.13
	(2.19)	(2.07)	(0.09)
Mother's age at first birth	23.92	24.02	0.03
	(4.46)	(4.60)	(0.10)
Family size	2.05	2.12	-0.02
	(1.06)	(1.13)	(0.04)
Family income percentile	44.64	38.63	-1.19
	(27.81)	(26.92)	(0.71)
Lowest income decile	0.09	0.12	-0.00
	(0.29)	(0.32)	(0.01)
Highest income decile	0.11	0.08	-0.01 *
	(0.31)	(0.27)	(0.01)
Grandparent present	0.61	0.70	-0.03
	(0.49)	(0.46)	(0.02)
Cohort size	147.37	67.70	-5.22
	(85.05)	(48.83)	(3.30)
Municipalities	89	140	223
Observations	21,581	14,752	90,434

Notes: This table reports the pre-period (cohorts 1962-1966) means and standard deviations of background characteristics for the first group of municipalities that receive public childcare after the 1973 *Childcare Law* (Column 1, Treatment) and the group of municipalities that only receive public childcare in later years (Column 2, Comparison). The difference-in-differences estimate of the difference before and after the 1973 *Childcare Law* for treated and comparison municipalities is shown in Column 3. *= p<0.05, **=p<0.01,***<p<0.001.

3.2 The effects of public childcare, on average and by family income

We begin by reporting our estimates of the average effects of public childcare access. Figure 3a plots the average difference in outcomes between treated and comparison municipalities by cohort (Equation 2). These event-study plots do not show any signs of a discernible change in outcomes of cohorts eligible for public childcare in treated versus comparison municipalities. This suggests that public childcare access did not shift the mean outcomes of men in cohorts eligible for childcare in the first set of municipalities exposed to the *Childcare Law of 1973*. Difference-in-differences estimates of the average treatment effects of access to public childcare suggest a similar story (Table 4, Column 3). Across almost all outcome measures, these results suggest that public childcare access did little to shift the average long-term outcomes of our estimation sample. By and large, our estimates of average treatment effects are economically and statistically indistinguishable from zero. The results are similar for females (see Appendix Tables 5).

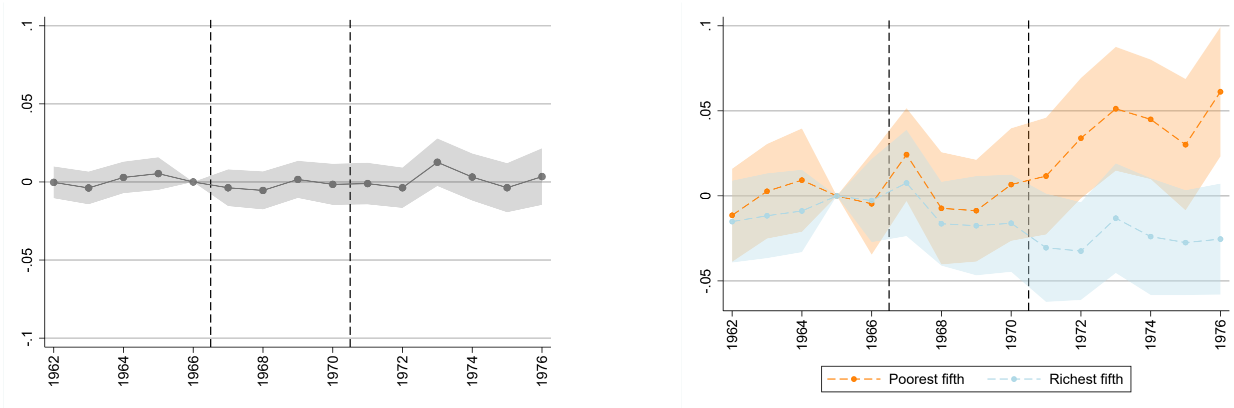
As public childcare could affect the relationship between parental income and children's outcomes without generating average effects, we examine any potential effects across the family income distribution (Figure 3b and Table 5). The coefficients displayed in the first column of Table 5 – β_2 from Equation 3 – measure how public childcare access changes the slope between parent income and child outcomes. The negative sign on these coefficients suggests that access to public childcare levels the playing field, reducing the association between family income and childrens' later outcomes. Since public childcare access could affect not just the slope of this relationship but also the intercept, we evaluate the magnitude of the treatment effect at three points in the family income distribution to help interpret the magnitude of these effects (Columns 2-4). These show that access to public childcare improves the long-term outcomes of children born to poor families substantially (Column 2), while children of the most affluent families (Column 4) experience worse outcomes from access to public childcare. Children from poor families are 3.6 percentage points less likely to drop out of secondary school, 3.3 percentage points more likely to gain a higher educational degree – and, as adults (ages 35-40) their incomes place them 2.3 percentage points higher in their cohort income rank. Children from rich families experience negative effects of almost comparable magnitudes.

Given the differential effects of public childcare access by family income, the association between family and son income rank falls by 5.1 points – reducing the intergenerational persistence of income rank by between twenty and forty-five percent from a baseline of 0.15.[12] These effects are comparable in magnitude to those produced by the Finnish comprehensive school reform (Pekkarinen et al., 2009). The combination of positive effects for poor children and negative effects for children from affluent families is consistent with prior work on universal childcare programs in Norway, Canada, Germany, and Italy (Havnes and Mogstad, 2011a; Kottelenberg and Lehrer, 2017; Cornelissen and Dustmann, 2019; Ichino et al., 2019).

[12]Since we only use pre-reform income to measure family income, this estimate of the baseline rank-rank relationship is slightly attenuated – but, in the same ballpark as in Norway during the same time-period (see Pekkarinen et al. (2017)).

Figure 3: Event-study comparisons of adult income rank

(a) Average treatment effects (b) Treatment effects by family income



Notes: These figures plot the estimates of treatment effects on adult income rank, following Equation 2 and Equation 5. The x-axis in this and all subsequent event-study figures is the birth cohort, rather than year. Corresponding graphs for other outcomes are shown in Appendix Figures 1-3.

Next, we study the effects of public childcare access on skills, measured in adulthood (age 19), upon conscription to the Finnish Defence Forces (Table 4, Panel B). While skills are only measured for eighty percent of the estimation sample, the coefficient measuring effects on military service suggests that public childcare access does not affect selection into measurement. In Panel B, we see that public childcare access levels the playing field not just in terms of adult economic outcomes, but also skills – with the largest effects on social competence, and the smallest effects on visual-spatial skills. Moreover, the confidence intervals imply that the effects on social competence and academic skills are greater than those on visual-spatial skills.[13]

The pattern of effects on skills – emphasizing social competence rather than visual-spatial skills – provides preliminary evidence that the effects on social skills may underlie the long-term effects of public childcare on adult outcomes. We delve more deeply into the relationship between skills and economic outcomes in the following section.

Robustness. Our results are robust to the inclusion of controls (Appendix Table 7, Columns 1-3) as well as various forms of regional trends (Appendix Table 8). The stability of the point estimates in the presence of controls suggests that changes in the demographic structure in municipalities are unlikely to explain our results. The inclusion of regional trends helps dissuade any fear that our estimates might be explained by simultaneous regional policies or patterns of industrialization. Importantly, results do not change when we use a triple-differences strategy that compares outcomes by family income within municipalities; this suggests that the pattern of heterogeneity by family

[13] While visual-spatial skills are sometimes considered to be innate, recent work has shown that working memory – an important dimension of these skills – is malleable (Berger et al., 2020).

income is not driven by differences in characteristics between treated municipalities (Appendix Table 8, Column 4).

Since skills are only measured for 80 percent of males, we pay particular attention to the sensitivity of these results. Reassuringly, we see that there are small and insignificant effects on military service, suggesting that we may not expect the skill estimates to be particularly affected by any missingness in that data (Tables 4-5). Further, we show that the effects on outcomes from administrative data do not budge when we restrict the sample to males with skill data available (Appendix Table 7, Column 4). Garlick and Hyman (2021) suggest that the inclusion of covariates tends to provide efficient and valid estimates in the presence of missingness. An exception to this is that estimates do shrink a little when we add education as a covariate; however, this is likely because parental education and family income-rank – the basis of heterogeneity – are so correlated. Further, we bound our estimates using Horowitz and Manski (2000) bounds (Columns 5 and 7). Although they are typically very conservative, given the negative selection to missing skill data, the bound generated when missing skill data is replaced with extremely low measures may be particularly informative in our case – since people for whom skills are missing tend to come from the bottom of the academic distribution. The estimates from this bound are similar to the skill estimates using controls (Column 5), and again suggest larger effects on social competence than on visual-spatial skills.

Further, estimates from staggered (stacked) binary and continuous difference-in-difference specifications complement our main estimates (Appendix Table 9). While 89 of the 229 municipalities in our main estimates receive treatment within our window, the number of municipalities in our treatment group nearly triples in our staggered design, where 248 of the 388 municipalities receive treatment. The fact that our estimates hold despite being based largely from a different set of municipalities helps alleviate the fear that our estimates might be a product of how we define our sample. Moreover, with the additional power from this design, we now have statistical power to observe that childcare access reduces dropout. Additionally, we use a continuous measure of treatment intensity which again corroborates our results. While it is possible that the precise number of spots in a municipality relates to local demand and it is more tricky to compare magnitudes between binary and continuous treatments, these results corroborate our results using a binary treatment definition.

Table 4: Descriptive data and average treatment effects

	Pre-period mean		Average treatment
	Treatment	Comparison	effect (ATE)
	(1)	(2)	(3)
	Panel A: Education, marriage, and the labor market		
Dropout	0.19	0.18	-0.016
	(0.39)	(0.39)	(0.010)
HS graduate	0.25	0.23	0.006
	(0.43)	(0.42)	(0.010)
Tertiary education	0.27	0.25	-0.009
	(0.44)	(0.43)	(0.010)
Years of education	12.40	12.33	0.008
	(2.30)	(2.20)	(0.060)
Income rank	0.46	0.44	0.003
	(0.28)	(0.28)	(0.008)
Years employed in 30's	8.08	8.16	0.013
	(2.99)	(2.91)	(0.074)
Ever married	0.59	0.59	-0.001
	(0.49)	(0.49)	(0.009)
Military service	0.81	0.82	0.011
	(0.39)	(0.38)	(0.033)
Municipalities	89	140	229
Individuals	55,730	34,704	90,434
		Panel B: Adult skills	
Visual-spatial	-0.20	-0.19	0.011
	(1.02)	(1.01)	(0.018)
Academic	-0.12	-0.08	0.015
	(1.02)	(1.01)	(0.019)
Social competence	-0.18	-0.19	-0.007
	(1.01)	(1.00)	(0.019)
Municipalities	89	134	223
Individuals	45,747	28,365	74,112

Notes: Column 1 and 2 report the means and standard deviations of all key outcome variables for cohorts who were childcare age prior to the introduction of the *Childcare Law* of 1973. Column 3 reports the average treatment effect estimates along with their standard errors following Equation 1. *= p<0.05, **=p<0.01,***<p<0.001.

Table 5: Treatment effects by family income

	Treat X family inc. percentile (1)	Effect for poorest fifth (2)	Effect at the median (3)	Effect for richest fifth (4)
	Panel A: Effects on education, marriage, and the labor market			
Dropout	0.051***	-0.036***	-0.016	0.004
	(0.011)	(0.010)	(0.010)	(0.013)
HS graduate	-0.102***	0.048***	0.007	-0.034**
	(0.016)	(0.010)	(0.009)	(0.013)
Tertiary education	-0.106***	0.033***	-0.009	-0.052***
	(0.015)	(0.009)	(0.009)	(0.013)
Years of education	-0.564***	0.234***	0.008	-0.217**
	(0.076)	(0.053)	(0.058)	(0.076)
Income rank	-0.051***	0.023**	0.002	-0.018
	(0.010)	(0.008)	(0.008)	(0.010)
Years employed in 30's	-0.378***	0.161	0.009	-0.142
	(0.089)	(0.082)	(0.073)	(0.080)
Ever married	-0.054***	0.020	-0.002	-0.024*
	(0.015)	(0.010)	(0.009)	(0.010)
Military service	0.022	0.003	0.012	0.020
	(0.013)	(0.034)	(0.033)	(0.033)
Municipalities	229			
Individuals	90,434			
		Panel B: Effects on skills		
Visual-spatial	-0.190***	0.088***	0.012	-0.064**
	(0.032)	(0.021)	(0.018)	(0.022)
Academic	-0.261***	0.120***	0.015	-0.089***
	(0.032)	(0.022)	(0.019)	(0.024)
Social competence	-0.269***	0.100***	-0.008	-0.116***
	(0.035)	(0.021)	(0.019)	(0.025)
Municipalities	223			
Individuals	75,996			

Notes: Column 1 of this table reports the coefficient β_2 from Equation 3. This coefficient measures the difference in effect of public childcare access between a child at the very bottom of the family income distribution compared to a child at the very top of the family income distribution. Column 2 (4) evaluates this expected treatment effect for the fifth of children from the poorest (richest) families. Column 3 evaluates the treatment effect for families at the middle of the family income distribution. *= p<0.05, **=p<0.01,***<p<0.001.

4 How does public childcare shape long-term outcomes?

4.1 Social skills as a potential driver of long-term outcomes

A prominent explanation for the effects of early childhood programs on adult economic outcomes is that they are driven by lasting effects on social skills (Deming, 2009; Heckman et al., 2013; Bailey et al., 2017; Pages et al., 2022). However, since linking the effects of childcare to measures of social skills from adulthood has been challenging, this hypothesis lacks strong empirical support.

To examine this empirically, we study the correlations between the treatment effects of childcare on long-term outcomes and the treatment effects of childcare on skills.[14] If the effects of childcare on economic outcomes are linked to social skills, we should expect that people who experience effects on economic outcomes experience lasting effects on social skills, but not necessarily on visual-spatial skills. This has a close parallel in the teacher value-added literature, where, for example, researchers study whether teachers who raise test-scores or improve behavior also improve long-term outcomes (Chetty et al., 2011; Jackson, 2018).

Since treatment effect correlations cannot be readily estimated in typical reduced form contexts like our own, where only one estimate (or a handful of estimates) per outcome is produced, we follow an insight from Angrist et al. (2022) and study the relationships between shorter and longer-term treatment effects across subgroups. We take several precautions to avoid bias as we estimate treatment effect correlations.[15] To avoid small sample bias in estimates of correlations, we generate a large number of granular subgroups based on predicted treatment effect heterogeneity across all our background covariates. These are estimated with an elastic net, using a machine learning framework from Chernozhukov et al. (2021).[16] This process is useful because it reduces researcher degrees of freedom by tying our hands in the construction of subgroups while maximizing the variation

[14] Bailey et al. (2020) argue that for long-term effects of childcare to be driven by some skill, that skill must be: i) malleable in early childhood; ii) relevant for long-term outcomes; and iii) would develop differently in the absence of childcare. Our reduced form results help to assess the malleability of various skills in the presence of childcare (points i and iii). Additionally, descriptive results provide evidence suggesting that the skills we measure are relevant for long-term outcomes (Table 2). This present test presents further empirical evidence supporting a mediation hypothesis. If treatment effects on social competence rather than visual-spatial skills drive the long-term effects of public childcare, we should expect treatment effects on long-term outcomes to be correlated with treatment effects on social competence, but not with those on visual-spatial skills. This strategy tests a basic assumption of mediation hypotheses – that people who experience effects on the main outcome should also experience effects on the mediator. Still, the empirical results it generates remain insufficient to show that effects on the mediator necessarily drive the effects on the outcome: effects on the outcome could be driven by any combination of both observed and unobserved mediators. Importantly, however, comparing the magnitudes of the treatment effect correlations across potential mediating hypotheses can be particularly informative if there are large differences in these coefficients, or if one goes to zero. The same approach to mediation can also be used to study the relation between moderators and treatment effects – for recent examples, see Gendron-Carrier et al. (2018) or Terrier et al. (2020).

[15] For an overview of our approach, see Appendix Figure 4, and for a more formal discussion of our approach and its assumptions is found in Appendix Section 5.1.

[16] For a full discussion of how we implement this, see Appendix Section 5.2.

in predicted treatment effects between subgroups. While our linear estimates of treatment effect heterogeneity by family income suggest a broad overlap between effects on economic outcomes and skills, these linear estimates may mask differences in the contours of the effects across the distribution. These subgroup estimates extend our linear estimates of treatment effect heterogeneity and reveal distinct non-linearities in the pattern of treatment effects across different outcomes (Appendix Figures 8-10).

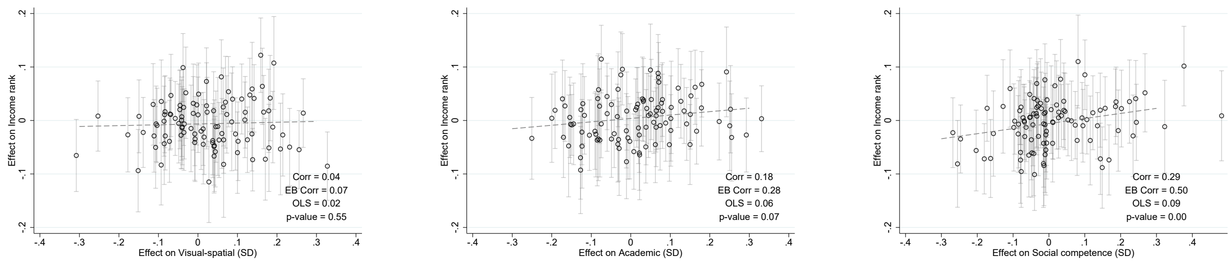
Then, to avoid upward bias caused by a mechanical correlation between treatment effects on different outcomes estimated using the same sample, we use a split-sample approach where we estimate the treatment effects on skills in one sample and the treatment effects on long-term outcomes in another. We randomize these samples several times and report the median correlation across these randomizations as our main result, and show the full distribution of split sample correlations in Appendix Figure 14. Finally, we avoid attenuation bias stemming from imprecision in the estimation of a large number of treatment effects by adjusting our estimates of the treatment effect correlation using empirical Bayes.

We find that treatment effects on income are most highly correlated with treatment effects on social competence ($r = 0.50$), less correlated with those on academic skills ($r = 0.28$), and least correlated with those on visual-spatial skills ($r = 0.07$). The relationship between treatment effects on years of education and skills displays a similar pattern of results (Figure 5b). The lack of a correlation in treatment effects on visual-spatial skills and long-term outcomes makes it unlikely that effects on fluid-intelligence explain long-run effects, while the relatively strong correlation between treatment effects on long-term outcomes and treatment effects on social skills provides evidence consistent with the idea that social skills may explain some of the long-term effects of childcare. When we put the different dimensions of skills in a horse-race with one another, the estimates produce a similar pattern, with social competence winning the horse-race in terms of both income and education effects (Appendix Table 11).

We complement our analysis of the association between treatment effects on long-term outcomes and treatment effects on skills with decomposition exercises that estimate the extent that income effects might be explained by skills (Imai et al., 2010; Heckman et al., 2013). Appendix Table 12 shows that when we add skills to the right side of the estimating equation (see Appendix Section 5.3, the treatment effect coefficient is reduced by about fifty percent – suggesting that effects of childcare on skills may be capable of explaining a sizable portion of the effects of childcare on labor market outcomes. Social competence alone reduces the treatment effect by about thirty percent. We also see that the addition of visual-spatial skills has no effect above and beyond academic skills and social competence (Appendix Table 12). Skills appear to have an even larger role in explaining the effects on years of education. The effects are reported separately for each skill in Appendix Table 13.

Figure 4: Treatment effects on income rank and treatment effects on measures of adult skills

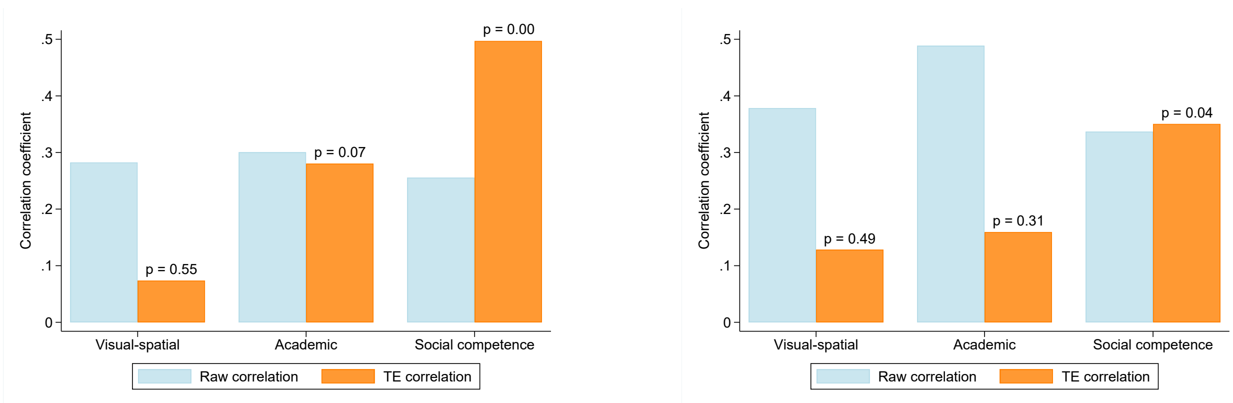
(a) Visual-spatial (b) Academic (c) Social competence



Notes: This figure plots the split-sample relationships between treatment effects on adult income rank (y-axis) and the treatment effects on measures of adult skills (x-axis), for a hundred subgroups based on the predicted treatment effect heterogeneity index. The correlation, raw and adjusted using empirical Bayes, as well as regression coefficients (along with their statistical significance) are reported in the bottom right of each figure.

Figure 5: Correlations between long-term outcome treatment effects and skill treatment effects

(a) Skills and adult income rank (b) Skills and years of education



Notes: This figure presents the correlation between treatment effects on long-term outcomes and treatment effects on measures of adult skills alongside the raw correlation between skills and long-term outcomes. This treatment effect correlation is the median split-sample estimate of treatment effect correlation, scaled using empirical Bayes, across fifty-one subgroups. The p-values are from the regression analogs of the correlations.

Robustness. To study the sensitivity of our estimates of treatment effect correlations to exactly how we split the data into subgroups, we re-run our estimates with a range of ten to a hundred subgroups (see Appendix Figure 15). The results from this exercise suggest that while the actual correlation coefficient does appear to attenuate up through around fifty splits, the relative ranks of the correlations across skills appears relatively stable. This decrease in the magnitude of our estimates could be due to mechanical small sample bias in the estimation of correlations as well as differences in the non-parametric pattern of effects across outcomes. This exercise highlights the importance of gauging the sensitivity of correlations in treatment effects across outcomes for various numbers of subgroups. While an insufficient number of subgroups may bias our estimates upward, the increasing imprecision of estimates based on more granular subgroups may bias our estimates downward. In our main estimates of the covariance, we correct our estimates for (some of) such attenuation bias using empirical Bayes. To ensure that these adjustments are not contributing to the qualitative nature of our findings, we report our estimates with and without empirical Bayes in Appendix Figure 15. As expected, although scaling our estimates by their reliability produces slightly higher treatment effect correlations it does not change the results in a substantive way.

Even if these treatment effect correlations are estimated without bias, differential measurement error in skills may make it difficult to compare treatment effects across outcomes. Unfortunately, without access to the particular items underlying the measures from the Finnish Defence Forces, we cannot estimate the reliabilities of these measures directly. However, the Finnish Defence Forces report that the Cronbach alphas for their measures of cognitive skills range between 0.76 and 0.88, while the Cronbach alphas for their measures of socio-emotional skills range from 0.6 and 0.9 (Nyman et al., 2007) – suggesting that, if anything, the measures of socio-emotional skills may be less reliable than the measures of cognitive skills. In our data, we see that the raw correlations between these skills and long-term outcomes are in a similar range to each other, though long-term outcomes are slightly more correlated with academic skills than with social competence (left bars in Figure 5). Further, as another way to interrogate the reliability of our skill measures, we study sibling correlations in the different skill outcomes (Appendix Table 10). In the case that each skill is equally shaped by genetic and environmental factors shared by siblings, we should expect the magnitudes of the sibling correlations of skills to reflect the reliability with which each skill is measured (a form of test-retest reliability). Both of these empirical exercises suggests that the pattern we see in our treatment effect correlations is not due to social skills being measured more reliably than cognitive skills.

Re-estimating our results using raw measures of skills from the Finnish Defence Force data helps to ensure that our main takeaways are not sensitive to how we construct our measures of skills (Appendix Figure 16). Just as with the composite measures, long-term effects of public childcare are most correlated with effects on social skills tied to social motivation and self-confidence, less correlated with effects on academic skills (arithmetic and verbal), and least correlated with visual-spatial skills. Interestingly, compared to the high treatment effect correlation between long-term outcomes and skills linked to motivation and sociability, in our context, effects on skills linked to self-regulation (deliberation and dutifulness) show a weaker correlation with long-term outcomes. This goes against the idea that childcare shifts long-term outcomes by improving self-regulation skills (Duncan et al., 2022). In line with research from psychology on childhood language exposure and language acquisition (Weisleder and Fernald, 2013; Romeo et al., 2018), effects on verbal skills appear slightly more correlated with effects on long-term outcomes than are effects on arithmetic, but we lack adequate statistical power to discern whether these differences are real.

Finally, the alignment in the results between the treatment effect correlation and the decomposition exercise suggest that, of candidate mechanisms, effects on visual-spatial effects are least likely to explain the long-term effects of childcare, while effects on social skills – and potentially the academic skills they facilitate – are most likely to explain the long-run effects of childcare.

Nonetheless, these exercises measure different things. While the treatment effect correlation tests the extent that changes in one long-term outcomes are associated with changes in skill outcomes, the decomposition exercise is based on regression – which scales this covariance by the magnitude of changes in the skill outcome. Moreover, the decomposition approach to mediation rests on stronger assumptions than the treatment effect correlations. First, given that in the decomposition exercise the effects we care about are modeled parametrically across the family income distribution, this analysis relies on parametric assumptions regarding the linearity in treatment effects across the family income distribution. As we can see in the non-linearities exhibited in Appendix Figures 8, there are reasons to think that this may not quite be the case – however, the direction of any bias from such non-linearities is ambiguous. Second, the decomposition exercise assumes that access to public childcare does not shift the adult income production function – the same inputs should matter in the same types of ways both in the presence of and absence of treatment. If the treatment affects only one dimension of a certain skill, and the different skills have distinct effects on labor market outcomes, changes in the composite skill may relate to labor market outcomes in ways that are distinct from the observed production in the absence of treatment. Third, this exercise may be more sensitive to exactly what is being measured, and how well. If skills are measured with error (see discussion in above paragraphs), the observed role of skills will be attenuated, leading us to understate the role of skills. While we might be able to scale the effects of skills by their reliabilities – we refrain from doing this, since these reliabilities apply when each skill is measured independently.

For both exercises we should be wary of ascribing a causal interpretation to the role of these specific skills. For example, it could well be that some unobserved factor, correlated with both long-term outcomes as well as the measures of skills is the causal driver – rather than the skills themselves. Still, these exercises do help provide evidence for the relative likelihood of the skills we focus on explaining the labor market effects of childcare.

4.2 Social skills, education, and occupational tasks

Social skills could affect economic outcomes in two primary ways. First, social skills may have a direct effect on productivity, for example by facilitating teamwork (Weidmann and Deming, 2021). Alternatively, social skills may make it possible to learn academic skills or make educational choices, thereby affecting productivity indirectly through education (see Heckman et al. (2013), Johnson and Jackson (2019), or Appendix Section 4.2).

Since childcare in 1970's Finland had little emphasis on academic learning, the fact that we see effects on academic skills suggests a possibility for some degree of indirect effects, whereby social skills enable academic learning. This is further highlighted by the effects we see on educational choices (Table 5). Together, shifts in academic and social skills explain the majority of the effects we see on educational choices (Table 12). Interestingly, however, even effects on education are more strongly correlated with effects on social skills than effects on academic skills (Figure 5).

To see if there might be direct effects of social skills on income, we perform two decomposition exercises (Appendix Table 14). In one, we add education to the right-side of the estimating equation with and without our measures of skills (Columns 2 and 3). Similarly, we add academic skills to the right-side of the estimating equation with and without our measure of social skills (Columns 4 and 5). Supporting the idea of direct effects, results from both exercises suggest that social skills play a role in explaining effects on income above and beyond through education or academic skills. At the same time, there appears to be significant overlap in effects on academic and social skills, making it hard to parse out the extent that either is responsible for effects on income.

Speer (2017) suggests that lasting effects on education and skills may affect people's choices of jobs, and thereby their incomes. Descriptively, our data show that skills are correlated with occupational tasks, and that tasks are correlated with income (Appendix Tables 15-16).[17] Moreover, using our main differences-in-differences approach, we estimate that access to public childcare shifts the types of jobs people end up in rather than changing their productivity in jobs that they would have done anyway (Table 6, Panel A). Similar estimates on the task content of their work (Acemoglu and Autor, 2011a) suggest that people who shift their occupations and earn more (less) as a result of childcare access are more (less) likely to work in jobs requiring social and non-routine cognitive tasks, and less (more) likely to work in manual jobs (Table 6, Panel B). There are no effects on the likelihood of working in a job requiring cognitive routine tasks.

Further decomposition results suggest that a large portion of the shifts in tasks can be explained by shifts in skills (Appendix Table 17). Likewise, we show that the shifts in occupational tasks can explain a large portion of the shift in incomes people experience as a result of expanded public childcare access (Appendix Table 18). The explanatory power of both skills and tasks is even larger

[17] For links between grades, education, and tasks in Finland, see Figure A4 from Silliman and Virtanen (2022).

– nearly sixty percent (Appendix Table 19).

Together, these results provide encouraging empirical evidence that high-quality childcare may be able to prepare people for jobs of the future that increasingly demand high levels of interpersonal skills (Acemoglu and Autor, 2011b; Deming, 2017).

Table 6: Treatment effects by family income percentile: Occupational task shares

	Treat X family inc. percentile (1)	Effect for poorest fifth (2)	Effect at the median (3)	Effect for richest fifth (4)
	Panel A: Effects on occupation and position			
Mean occupational income rank	-0.041***	0.013*	-0.004	-0.020***
	(0.007)	(0.005)	(0.004)	(0.005)
Income rank within occupation	0.011	0.007	0.011	0.015*
	(0.008)	(0.007)	(0.006)	(0.007)
Municipalities	229			
Individuals	90,434			
	Panel B: Effects on occupational task content			
Social non-routine analytic	-0.152***	0.072***	0.011	-0.050*
	(0.028)	(0.017)	(0.015)	(0.021)
Social non-routine manual	-0.195***	0.087***	0.009	-0.069***
	(0.030)	(0.019)	(0.015)	(0.020)
Cognitive non-routine	-0.151***	0.072***	0.012	-0.048*
	(0.026)	(0.017)	(0.015)	(0.020)
Cognitive routine	-0.014	0.013	0.008	0.002
	(0.024)	(0.016)	(0.012)	(0.015)
Manual non-routine	0.229***	-0.092***	-0.001	0.091***
	(0.036)	(0.020)	(0.016)	(0.024)
Manual routine	0.207***	-0.089***	-0.006	0.076**
	(0.033)	(0.019)	(0.018)	(0.024)
Municipalities	229			
Individuals	77,154			

Notes: Column 1 of this table reports the coefficient β_2 from Equation 3. This coefficient measures the difference in effect of public childcare access between a child at the very bottom of the family income distribution compared to a child at the very top of the family income distribution. Column 2(4) evaluates this expected treatment effect for the fifth of children from the poorest(richest) families. Column 3 evaluates the treatment effect for families at the middle of the family income distribution. *= p<0.05, **=p<0.01,***<p<0.001.

4.3 Early childhood socialization and shifts in social skills

We build on prior literature to assess how both between and within family differences in counterfactual early childhood environments relate to the effects of public childcare access. While the treatment in our study – access to public childcare – is relatively constant across children from different families, the way that public childcare changes each person's early childhood environment varies based on the quality of the counterfactual care environment children are exposed to (See Appendix Figure 18). In particular, theory from psychology highlights the potential socializing role of early childhood environment (Black et al., 2017) and the importance of interactions between young children and adults in shaping social skills (Clarke-Stewart et al., 1994; Csibra and Gergely, 2009).[18]

First, we test for the extent that public childcare might substitute for maternal care. In a companion paper we show that public childcare access substantially increases maternal employment Mäkinen and Silliman, 2022 and Appendix Figure 19a), which is consistent with the idea that public childcare may serve as a substitute for home-care. To test for the substitution between maternal care and childrens outcomes more directly, we correlate maternal treatment effects on maternal labor market participation with changes in childrens' outcomes. These correlations suggest that increases with maternal labor force participation are associated with shifts in childrens outcomes (Appendix Figure 19b).[19] That said, it is also possible that public childcare access substitutes for other forms of childcare, such as care by grandparents, (unsubsidized) family-run childcare centers, or informal care arrangements between neighbors. For more affluent families, it is also possible that public childcare substituted for private care by a nanny. Unfortunately, we cannot observe these counterfactuals directly in the data.

Recognizing the potential substitution between public childcare and other forms of care, other papers have studied the differential effects of public childcare on children born to families with different levels of resources (Havnes and Mogstad, 2015). A potential reason for this is that more affluent parents are shown to invest more time in their children (Guryan et al., 2008). Our results are consistent with this pattern and suggest that children born to poor families may benefit from public childcare, while children born to higher income families can be hurt by public childcare. Such heterogeneity may also explain why the effects of Head Start for early cohorts may be greater than for later cohorts (Deming, 2009; Pages et al., 2019). However, results from Cornelissen et al. (2018)

[18] A large literature in child development also finds that socialization in the home – parenting – is enormously important in shaping children's development of behavior and personality (Baumrind, 1967; Grolnick and Ryan, 1989; Darling and Steinberg, 1993; Kochanska, 1993; Coleman and Karraker, 1998; Aunola et al., 2000; Aunola and Nurmi, 2005; Pomerantz et al., 2005).

[19] Prior work studying a similar policy in Norway find no effects for mothers employment (Havnes and Mogstad, 2011a), suggesting that the effects they find might not be explained by substitution between maternal care and public childcare. It is possible that the different results between our context and theirs stem from measurement error. When we focus on a binary measure of employment we do not find any effects either, but when we use maternal earnings we find substantial effects.

suggest that heterogeneity in the benefits of public childcare may follow several other – though in their case unobserved – dimensions.

Next, we test for heterogeneity in the effects of public childcare by a dimension known to be tied to the quality of early life socialization – first-born status (Price, 2008; Black et al., 2018; Lehmann et al., 2018) – but not tied to between-family variation in resources. Black et al. (2018) show that first born children attain higher levels of skills – notably non-cognitive skills. Empirical evidence suggests that this may be because first-born children receive greater attention from their parents (Price, 2008; Lehmann et al., 2018). We corroborate results from Black et al. (2018) and show that first-born male children perform significantly more highly than their siblings in terms of our measures of visual-spatial, academic, and social skills, and then study heterogeneity in the effects of public childcare by first-born status. Though somewhat underpowered, our results are in line with theory, suggesting that public childcare may equalize differences in parental attention, resulting in some of the advantage (about 25%) of first-born children disappearing in terms of social competence (Appendix Table 20).[20]

Heterogeneity in the treatment effects of public childcare may also extend to further dimensions beyond just maternal employment, family income, and first-born status. The machine learning predictions of treatment effect heterogeneity we use to divide our data into subgroups offer us a way to examine heterogeneity across a number of measures of family characteristics (Appendix Figure 20 and Appendix Table 23).[21] These corroborate our prior results, showing that children from poor families and those with older mothers – particularly those likely to work (Heiland et al., 2017) (those who are older and more educated) – benefit from public childcare. Additionally, we see some evidence that other dimensions of early childhood socialization might matter. Children with few older siblings benefit from public childcare access, as do children without grandparents nearby.

Although we cannot directly observe the counterfactual form of childcare, each of the patterns of heterogeneity we study – by maternal employment, family income, first-born status, and our predictions of treatment effect heterogeneity – emphasizes different dimensions of quality in childhood interactions with adults. Together these provide evidence consistent with the idea that public

[20]We may detect effects on social competence but not other skills because these skills are taught by parents with or without public childcare, because the development of social competences is a greater focus for public childcare programs, or because social competence is more sensitive to change between the ages 3-6 than are visual-spatial or academic skills.

[21]Appendix Figure 20 and Appendix Table 23 describe this heterogeneity. In Appendix Figure 20, our sample is split into ten equal size bins based on predicted treatment effect rank are plotted along the x-axis. Those in the rightmost bins are expected to benefit the most from public childcare access, while those in the leftmost bins are expected to benefit the least from public childcare access. The extent to which the bin-mean differs from the overall mean for a variable determines the color assigned to that bin for that variable. The deepest red squares indicate that the bin-mean is greater than 0.15 standard deviations larger than the overall mean; conversely, the darkest-blue squares indicate that the bin-mean is at least 0.15 standard deviations less than the overall mean. Appendix Table 23 shows the mean values of the covariates for individuals in the bottom and top quintiles of predicted treatment effects (these correspond to the two leftmost (rightmost) bins in Figure 20).

childcare shifts long-term outcomes by changing the quality of early childhood socialization. In contrast, the mix of both positive and negative effects we find is unlikely to be explained by income effects (Black et al., 2014), since no families experience negative income shocks as a result of public childcare access. This pattern of results is in line with recent work by Yum (2022), who argues that public investments in early childhood can equalize differences in parental time investments.

5 Conclusion

We study how public childcare shifts long-term outcomes, linking together data on public childcare access, measures of skills from adulthood, labor market outcomes, and occupational task content. First, we find that Finland's first national public childcare program markedly reduces the intergenerational persistence between parent and child earnings. Next we provide a series of empirical results that suggest that the lasting effects on social skills rather than fluid intelligence explain the long-run effects of childcare. Apart from bringing new evidence on the effects of a national public childcare program (Duncan et al., 2022), these results provide information on how early childhood programs might better be designed to ensure that they provide their participants sustained benefits (Rege et al., 2021).

Beyond linking the effects of public childcare on earnings to social skills, we show that by shaping people's skills, childcare shifts the types of jobs where people work as adults. Given recent evidence that there is increasing demand for jobs requiring social skills (Deming, 2017), these results suggest that high quality childcare may help provide people the skills they need to meet this demand. Additionally, our results provide a basis for optimism regarding the ability of a new generation of educational interventions targeting social skills to generate long-term gains (Alan et al., 2019; Berger et al., 2020; Cappelen et al., 2020; Sorrenti et al., 2020; Algan et al., 2022).

More broadly, these results add new evidence highlighting the importance of accounting for the multidimensional effects of educational programs (Heckman et al., 2013; Jackson, 2018; Jackson et al., 2020; Rose et al., 2022). As labor markets continue to shift to reward different types of skills – non-routine analytic skills (Autor et al., 2003; Acemoglu and Autor, 2011b), cognitive endurance (Brown et al., 2022), decision-making (Deming, 2021) or social skills (Deming, 2017) – a key area of research will be to understand how to design educational programs that are capable of producing skills that meet the demands of the future of work.

References

Acemoglu, D. and Autor, D. (2011a). Data for: Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages Task measure construction, onet–task–occ2000.dta in http://economics.mit.edu/faculty/dautor/data/acemoglu. Accessed June 15th, 2018. Elsevier.

Acemoglu, D. and Autor, D. (2011b). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier.

Alan, S., Boneva, T., and Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*, 134(3):1121–1162.

Algan, Y., Beasley, E., Côté, S., Vitaro, F., Tremblay, R. E., and Park, J. (2022). The impact of childhood social skills and self-control training on economic and non-economic outcomes: Evidence from a randomized experiment using administrative data. *American Economic Review*.

Alila, K., Eskelinen, M., Estola, E., Kahiluoto, T., Kinos, J., et al. (2014). Varhaiskasvatuksen historia, nykytila ja kehittamisen suuntalinjat: Tausta-aineisto varhaiskasvatusta koskevaa lainsaadantoaa valmistelevan tyoryhman tueksi.

Almond, D., Currie, J., and Duque, V. (2017). Childhood circumstances and adult outcomes: Act ii. Technical report, National Bureau of Economic Research.

Angrist, J., Autor, D., and Pallais, A. (2022). Marginal effects of merit aid for low-income students. *The Quarterly Journal of Economics*, 137(2):1039–1090.

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11:685–725.

Aunola, K. and Nurmi, J.-E. (2005). The role of parenting styles in childrens problem behavior. *Child Development*, 76(6):1144–1159.

Aunola, K., Stattin, H., and Nurmi, J.-E. (2000). Parenting styles and adolescents' achievement strategies. *Journal of adolescence*, 23(2):205–222.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333.

Bailey, D., Duncan, G. J., Odgers, C. L., and Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of research on educational effectiveness*, 10(1):7–39.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., and 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, M. (2011). Innis lecture: Universal early childhood interventions: what is the evidence base? *Canadian Journal of Economics/Revue canadienne d'économique*, 44(4):1069–1105.

Barr, A. and Gibbs, C. (2022). Breaking the cycle? intergenerational effects of an anti-poverty program in early childhood.

Baumrind, D. (1967). Child care practices anteceding three patterns of preschool behavior. *Genetic psychology monographs*.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, Part 2):9–49.

Becker, G. S. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of labor economics*, 4(3, Part 2):S1–S39.

Belloni, A., Chernozhukov, V., Fernández-Val, I., and Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

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

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Bhuller, M., Mogstad, M., and Salvanes, K. G. (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, 35(4):993–1030.

Biden, J. (2021). Fact sheet: The american families plan. *The White House. https://www. whitehouse. gov/briefing-room/statements-releases/2021/04/28/fact-sheet-theamerican-families-plan*.

Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., McCoy, D. C., Fink, G., Shawar, Y. R., Shiffman, J., et al. (2017). Early childhood development coming of age: science through the life course. *The Lancet*, 389(10064):77–90.

Black, S. E., Devereux, P. J., Løken, K. V., and Salvanes, K. G. (2014). Care or cash? the effect of child care subsidies on student performance. *Review of Economics and Statistics*, 96(5):824–837.

Black, S. E., Grönqvist, E., and Öckert, B. (2018). Born to lead? the effect of birth order on noncognitive abilities. *Review of Economics and Statistics*, 100(2):274–286.

Bost, K. K., Vaughn, B. E., Washington, W. N., Cielinski, K. L., and Bradbard, M. R. (1998). Social competence, social support, and attachment: Demarcation of construct domains, measurement, and paths of influence for preschool children attending head start. *Child Development*, 69(1):192–218.

Brown, C. L., Kaur, S., Kingdon, G., and Schofield, H. (2022). Cognitive endurance as human capital. Technical report, National Bureau of Economic Research.

Busch-Rossnagel, N. A. and Knauf-Jensen, D. E. (1995). Mothers and others: The role of the socializing environment. *Mastery motivation: Origins, conceptualizations, and applications*, 12:117.

Callaway, B., Goodman-Bacon, A., and Sant'Anna, P. H. (2021). Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*.

Campbell, J. J., Lamb, M. E., and Hwang, C. P. (2000). Early child-care experiences and children's social competence between 11/2 and 15 years of age. *Applied Developmental Science*, 4(3):166–176.

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

Carneiro, P. and Ginja, R. (2014). Long-term impacts of compensatory preschool on health and behavior: Evidence from head start. *American Economic Journal: Economic Policy*, 6(4):135–73.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernandez-Val, I. (2021). Generic machine learning inference on heterogenous treatment effects in randomized experiments. Technical report.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Clarke-Stewart, A., Gruber, C. P., and Fitzgerald, L. M. (1994). *Children at home and in day care*. Psychology Press.

Clausen, J. A. (1966). Family structure, socialization, and personality. *Review of child development research*, 2:1–53.

Cobb-Clark, D. A. and Schurer, S. (2012). The stability of big-five personality traits. 115(1):11–15.

Coleman, P. K. and Karraker, K. H. (1998). Self-efficacy and parenting quality: Findings and future applications. *Developmental review*, 18(1):47–85.

Coolahan, K., Fantuzzo, J., Mendez, J., and McDermott, P. (2000). Preschool peer interactions and readiness to learn: Relationships between classroom peer play and learning behaviors and conduct. *Journal of Educational Psychology*, 92(3):458.

Cornelissen, T. and Dustmann, C. (2019). Early school exposure, test scores, and noncognitive outcomes. *American Economic Journal: Economic Policy*, 11(2):35–63.

Cornelissen, T., Dustmann, C., Raute, A., and 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.

Csibra, G. and Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, 13(4):148–153.

Cunha, F. and Heckman, J. J. (2010). Investing in our young people. Technical report, National Bureau of Economic Research.

Currie, J. and Almond, D. (2011). Human capital development before age five. *Handbook of labor economics*, 4:1315–1486.

Darling, N. and Steinberg, L. (1993). Parenting style as context: An integrative model. *Psychological bulletin*, 113(3):487.

Davis, J. M. and Heller, S. B. (2020). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *Review of Economics and Statistics*, 102(4):664–677.

Decker, P. and Kelly, K. (2022). Should childcare subsidies be universal or targeted? *Journal of Policy Analysis and Management*, 41(3):921–943.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, 1(3):111–34.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640.

Deming, D. J. (2021). The growing importance of decision-making on the job. Technical report, National Bureau of Economic Research.

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S., and Queenan, P. (2003). Preschool emotional competence: Pathway to social competence? *Child Development*, 74(1):238–256.

Dodge, K. A., Pettit, G. S., McClaskey, C. L., Brown, M. M., and Gottman, J. M. (1986). Social competence in children. *Monographs of the society for research in child development*, pages i–85.

Drange, N. and Havnes, T. (2019). Early childcare and cognitive development: Evidence from an assignment lottery. *Journal of Labor Economics*, 37(2):581–620.

Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, 92(6):1087.

Duncan, G., Kalil, A., Mogstad, M., and Rege, M. (2022). Investing in early childhood development in preschool and at home. Technical report, National Bureau of Economic Research.

Duncan, G. J., Brooks-Gunn, J., and Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development*, 65(2):296–318.

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

Dweck, C. S. (1986). Motivational processes affecting learning. *American psychologist*, 41(10):1040.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Elango, S., García, J. L., Heckman, J. J., and Hojman, A. (2016). *4. Early Childhood Education*. University of Chicago Press.

Erikson, E. H. (1950). *Childhood and society*. WW Norton & Company.

Erola, J. (2018). Varhaisen päivähoidon pitkän aikavälin vaikutukset: Virheitä ja vahvistuneita havaintoja.

Erola, J., Hiilamo, H., and Laaninen, M. (2020). Lasten päivähoitovalintojen yhteys pisamenestykseen: Ristiriitaisia tuloksia ja tulkintoja.

European Union (2019). Council recommendation of 22 may 2019 on high-quality early childhood education and care systems. *Official Journal of the European Union*, 2019/C 189/02.

Falk, A., Kosse, F., Pinger, P., Schildberg-Hörisch, H., and Deckers, T. (2021). Socioeconomic status and inequalities in children's iq and economic preferences. *Journal of Political Economy*, 129(9):2504–2545.

Finnish Defense Forces (2021). Test data, cohorts 1962-1976.

García, J. L., Heckman, J. J., and Ronda, V. (2021). The lasting effects of early childhood education on promoting the skills and social mobility of disadvantaged african americans. Technical report, National Bureau of Economic Research.

Gardner, J. (2021). Two-stage differences-in-differences. *Unpublished manuscript*.

Garlick, R. and Hyman, J. (2021). Quasi-experimental evaluation of alternative sample selection corrections. *Journal of Business & Economic Statistics*, pages 1–15.

Gendron-Carrier, N., Gonzalez-Navarro, M., Polloni, S., and Turner, M. A. (2018). Subways and urban air pollution. Technical report, National Bureau of Economic Research.

Gibbs, C., Ludwig, J., and Miller, D. L. (2011). Does head start do any lasting good? Technical report, National Bureau of Economic Research.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Technical report, National Bureau of Economic Research.

Gray-Lobe, G., Pathak, P., and Walters, C. (2021). The long-term effects of universal pre-k in boston. *NBER WP Series*.

Grolnick, W. S. and Ryan, R. M. (1989). Parent styles associated with children's self-regulation and competence in school. *Journal of Educational Psychology*, 81(2):143.

Gunderson, E. A., Gripshover, S. J., Romero, C., Dweck, C. S., Goldin-Meadow, S., and Levine, S. C. (2013). Parent praise to 1-to 3-year-olds predicts children's motivational frameworks 5 years later. *Child Development*, 84(5):1526–1541.

Gupta, N. D. and Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics*, 94(1):30–43.

Guryan, J., Hurst, E., and Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic Perspectives*, 22(3):23–46.

Hart, B. and Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing.

Havnes, T. and Mogstad, M. (2011a). Money for nothing? universal child care and maternal employment. *Journal of Public Economics*, 95(11-12):1455–1465.

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

Havnes, T. and Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of Public Economics*, 127:100–114.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900–1902.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and 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. and Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, 91(2):145–149.

Heiland, F., Price, J., and Wilson, R. (2017). Maternal employment and time investments in children. *Review of Economics of the Household*, 15(1):53–67.

Hiilamo, H., Merikukka, M., and Haataja, A. (2018). Long-term educational outcomes of child care arrangements in finland. *SAGE Open*, 8(2):2158244018774823.

Horowitz, J. L. and Manski, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American statistical Association*, 95(449):77–84.

Hulkko, J. (1971). Children's day-care in finland. *Finnish Yearbook of Population Research*, pages 81–85.

Ichino, A., Fort, M., and Zanella, G. (2019). Cognitive and non-cognitive costs of daycare 0–2 for children in advantaged families. *Journal of Political Economy*, 704075.

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, 15(4):309.

Imai, K., Ratkovic, M., et al. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470.

Izadi, R. and Tuhkuri, J. (2021). School vs. action-oriented personalities in the labor market.

Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, 126(5):2072–2107.

Jackson, C. K., Porter, S. C., Easton, J. Q., Blanchard, A., and Kiguel, S. (2020). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights*, 2(4):491–508.

Johnson, R. C. and Jackson, C. K. (2019). Reducing inequality through dynamic complementarity: Evidence from head start and public school spending. *American Economic Journal: Economic Policy*, 11(4):310–49.

Jokela, M., Pekkarinen, T., Sarvimäki, M., Terviö, M., and Uusitalo, R. (2017). Secular rise in economically valuable personality traits. *Proceedings of the National Academy of Sciences*, 114(25):6527–6532.

Karhula, A., Erola, J., and Kilpi-Jakonen, E. (2017). *Childcare, early education, and social inequality*, chapter "Home sweet home? Long-term educational outcomes of childcare arrangements in Finland", pages 268–286. Edward Elgar Publishing.

Kitagawa, T., Nybom, M., and Stuhler, J. (2018). Measurement error and rank correlations. Technical report, cemmap working paper.

Kline, P. and 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. (1993). Toward a synthesis of parental socialization and child temperament in early development of conscience. *Child Development*, 64(2):325–347.

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

Kottelenberg, M. J. and Lehrer, S. F. (2017). Targeted or universal coverage? assessing heterogeneity in the effects of universal child care. *Journal of Labor Economics*, 35(3):609–653.

Ladd, G. W. (1990). Having friends, keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development*, 61(4):1081–1100.

Ladd, G. W. (2005). *Children's peer relations and social competence: A century of progress*. Yale University Press.

Law 145/1922 (1922). Suomen asetuskokoelma. Koyhainhuoltolaki 145/1922.

Law 296/1927 (1927). Suomen asetuskokoelma. Laki lastentarhain valtioavusta 296/1927.

Law 36/1973 (1973). Suomen asetuskokoelma. Paivahoito laki 36/1973.

Law 80/1936 (1936). Suomen asetuskokoelma. Asetus lastentarhoista, jotka saavat valtioapua 80/1936.

Lehmann, J.-Y. K., Nuevo-Chiquero, A., and Vidal-Fernandez, M. (2018). The early origins of birth order differences in children's outcomes and parental behavior. *Journal of Human Resources*, 53(1):123–156.

Lenard, M. and Silliman, M. (2021). Informal social interactions, behavior, and academic achievement: Evidence from bus-peers. *EdWorkingPapers. Annenberg Institute.*

Li, W., Duncan, G. J., Magnuson, K., Schindler, H. S., Yoshikawa, H., and 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*.

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

Ludwig, J. and 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.

Maccoby, E. E. and Lewis, C. C. (2003). Less day care or different day care? *Child Development*, 74(4):1069–1075.

Mäkinen, J. and Silliman, M. (2022). Achieving career and family: The effects of public childcare on maternal work through retirement. *Unpublished manuscript*.

Mincer, J. (1974). Schooling, experience, and earnings. human behavior & social institutions no. 2.

Nyman, K. et al. (2007). Varusmiesten johtajavalintojen luotettavuus. *Julkaisusarja 1 Nro 1/2007*.

Pages, R., Bailey, D. H., and Duncan, G. J. (2022). Exploring the "dark matter" of early childhood educational programs: A pattern-of-indirect-effect approach.

Pages, R. J.-C., Lukes, D. J., Bailey, D. H., Duncan, G. J., et al. (2019). Elusive longer-run impacts of head start: Replications within and across cohorts. *arXiv preprint arXiv:1903.01954*.

Papageorge, N. W., Ronda, V., and Zheng, Y. (2019). The economic value of breaking bad: Misbehavior, schooling and the labor market. Technical report, National Bureau of Economic Research.

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

Pekkarinen, T., Uusitalo, R., and Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the finnish comprehensive school reform. *Journal of Public Economics*, 93(7-8):965–973.

Phillips, D., McCartney, K., and Scarr, S. (1987). Child-care quality and children's social development. *Developmental psychology*, 23(4):537.

Piaget, J. (1954). *The construction of reality in the child*. Routledge.

Pomerantz, E. M., Grolnick, W. S., and Price, C. E. (2005). The role of parents in how children approach achievement: A dynamic process perspective.

Price, J. (2008). Parent-child quality time does birth order matter? *Journal of human resources*, 43(1):240–265.

Rege, M., Størksen, I., Solli, I. F., Kalil, A., McClelland, M. M., Ten Braak, D., Lenes, R., Lunde, S., Breive, S., Carlsen, M., et al. (2021). The effects of a structured curriculum on preschool effectiveness: A field experiment. *Journal of Human Resources*, pages 0220–10749R3.

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., and Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological science*, 29(5):700–710.

Rose, E. K., Schellenberg, J. T., and Shem-Tov, Y. (2022). The effects of teacher quality on adult criminal justice contact. Technical report, National Bureau of Economic Research.

Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social development*, 6(1):111–135.

Saarinen, A., Lipsanen, J., Huotilainen, M., Hintsanen, M., Keltikangas-Järvinen, L., et al. (2019). The association of early childhood education and care with cognitive learning outcomes at 15 years of age in finland. *Psychology*.

Silliman, M. and Virtanen, H. (2022). Labor market returns to vocational secondary education. *American Economic Journal: Applied Economics*.

Sorrenti, G., Zölitz, U., Ribeaud, D., and Eisner, M. (2020). The causal impact of socio-emotional skills training on educational success. *University of Zurich, Department of Economics, Working Paper*, (343).

Speer, J. D. (2017). Pre-market skills, occupational choice, and career progression. *Journal of Human Resources*, 52(1):187–246.

Statistics Finland (2021a). Degree registry [tutkintorekisteri]. *Statistics Finland Research Services*.

Statistics Finland (2021b). Five-year census 1970-1985. *Statistics Finland Research Services*.

Statistics Finland (2021c). Folk basic module [perustietomoduuli]. *Statistics Finland Research Services*.

Suomen Kunnallisliitto (1974). *Huoltolaitosten taloustilasto 1973*. Suomen Kunnallisliitto.

Suomen Kunnallisliitto (1975). *Huoltolaitosten taloustilasto 1974*. Suomen Kunnallisliitto.

Suomen Kunnallisliitto (1976). *Huoltolaitosten taloustilasto 1975*. Suomen Kunnallisliitto.

Suomen Kunnallisliitto (1977). *Huoltolaitosten taloustilasto 1976*. Suomen Kunnallisliitto.

Suomen Kunnallisliitto (1978). *Huoltolaitosten taloustilasto 1977*. Suomen Kunnallisliitto.

Suomen Kunnallisliitto (1979). *Huoltolaitosten taloustilasto 1978*. Suomen Kunnallisliitto.

Suomen Kunnallisliitto (1980). *Huoltolaitosten taloustilasto 1979*. Suomen Kunnallisliitto.

Suomen Kunnallisliitto (1981). *Huoltolaitosten taloustilasto 1980, osa II*. Suomen Kunnallisliitto.

Suomen Kunnallisliitto (1982). *Huoltolaitosten taloustilasto 1981, osa II*. Suomen Kunnallisliitto.

Terrier, C., Machin, S. J., McNally, S., and Ventura, G. (2020). Closing the gap between vocational and general education? evidence from university technical colleges in england.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Valtiopaivat 1972 (1973). *Asiakirjat, Ensimmainen osa, 2*. Valtion Painatuskeskus, Helsinki.

Vaughn, B. E., Shin, N., Kim, M., Coppola, G., Krzysik, L., Santos, A. J., Peceguina, I., Daniel, J. R., Veríssimo, M., DeVries, A., et al. (2009). Hierarchical models of social competence in preschool children: A multisite, multinational study. *Child Development*, 80(6):1775–1796.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Waters, E. and Sroufe, L. A. (1983). Social competence as a developmental construct. *Developmental review*, 3(1):79–97.

Weidmann, B. and Deming, D. J. (2021). Team players: How social skills improve team performance. *Econometrica*, 89(6):2637–2657.

Weiland, C. and 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.

Weisleder, A. and Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological science*, 24(11):2143–2152.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., and Zaslow, M. J. (2013). Investing in our future: The evidence base on preschool education.

Yum, M. (2022). Parental time investment and intergenerational mobility. *Available at SSRN 3862378*.

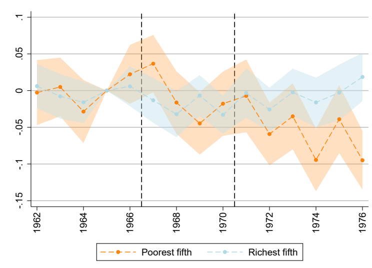
Zhao, Q., Small, D. S., and Ertefaie, A. (2018). Selective inference for effect modification via the lasso. *arXiv preprint arXiv:1705.08020*.

APPENDIX

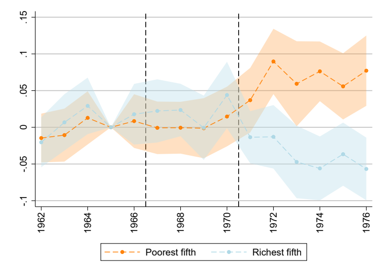
1 Figures

Figure 1: Event-study plots of long-run outcomes based on family income

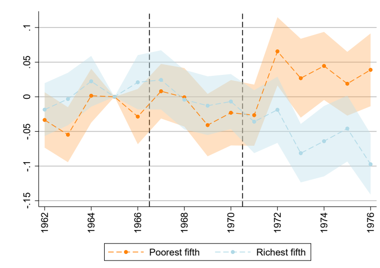
(a) Dropout



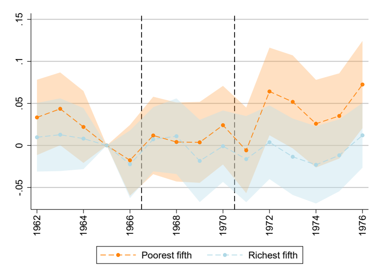
(b) General track secondary degree



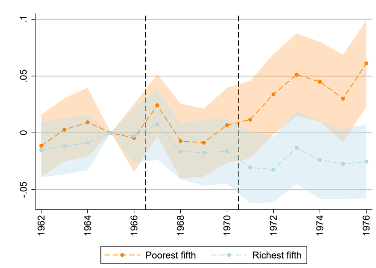
(c) Tertiary completion



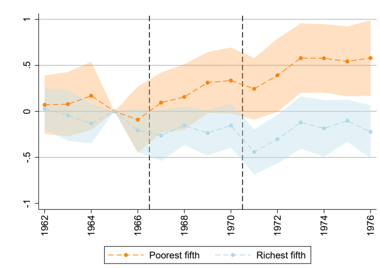
(d) Ever married



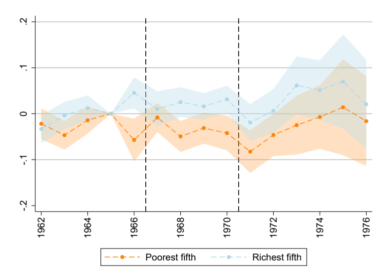
(e) Adult income rank, 35-40



(f) Years employed, 30-40



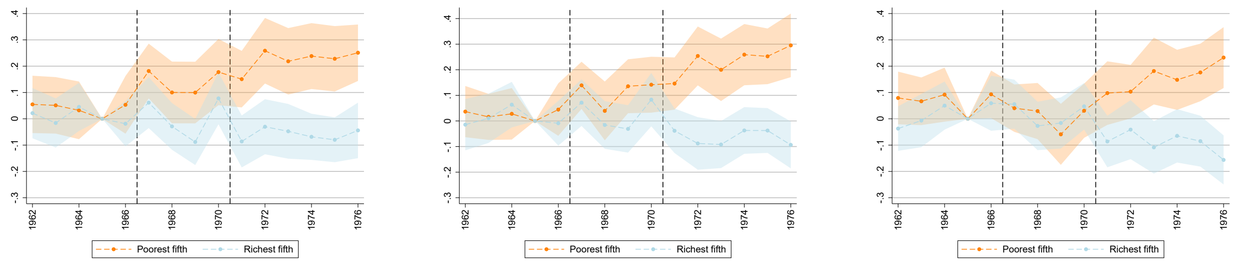
(g) Military service



Notes: This figure shows event-study plots for measures long-term outcomes for children from the richest and poorest fifth of families in our estimation sample using the specification from Equation 5.

Figure 2: Event-study plots of adult skills based on family income

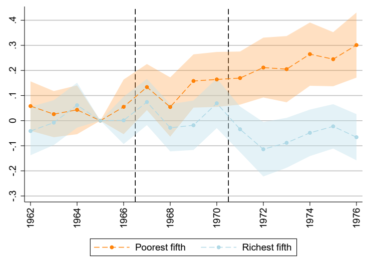
(a) Visual-spatial skills (b) Academic skills (c) Social competence



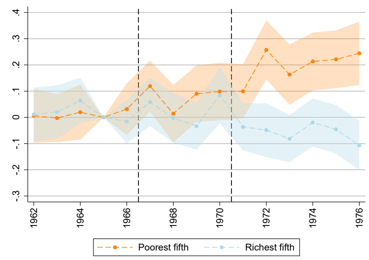
Notes: This figure shows event-study plots for measures of adult skills for children from the richest and poorest fifth of families in our estimation sample using the specification from Equation 5.

Figure 3: Event-study plots for measures of adult skills based on family income

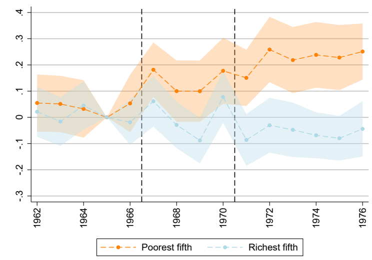
(a) Arithmetic



(b) Verbal

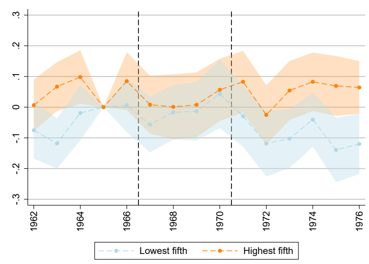


(c) Visual-spatial

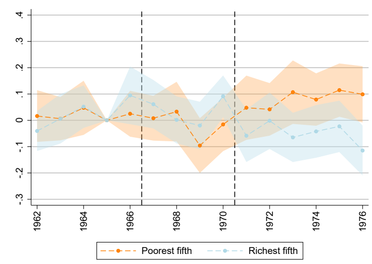


(d) Deliberation

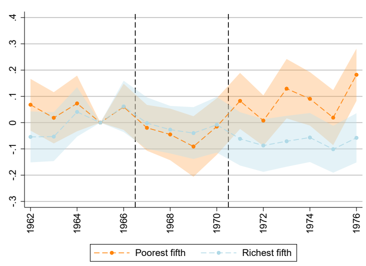
(e) Dutifulness



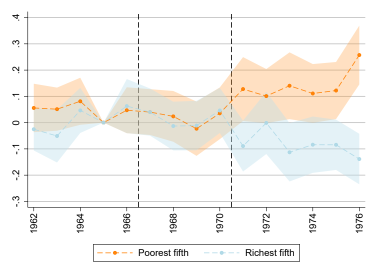
(f) Achievement-striving



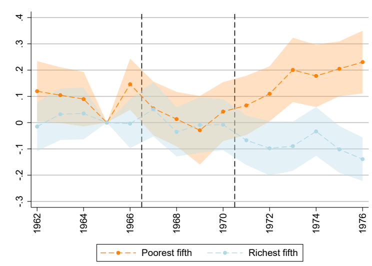
(g) Activity-energy



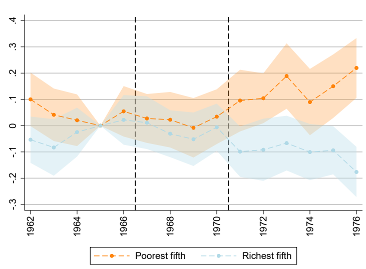
(h) Leadership motivation



(i) Self-confidence

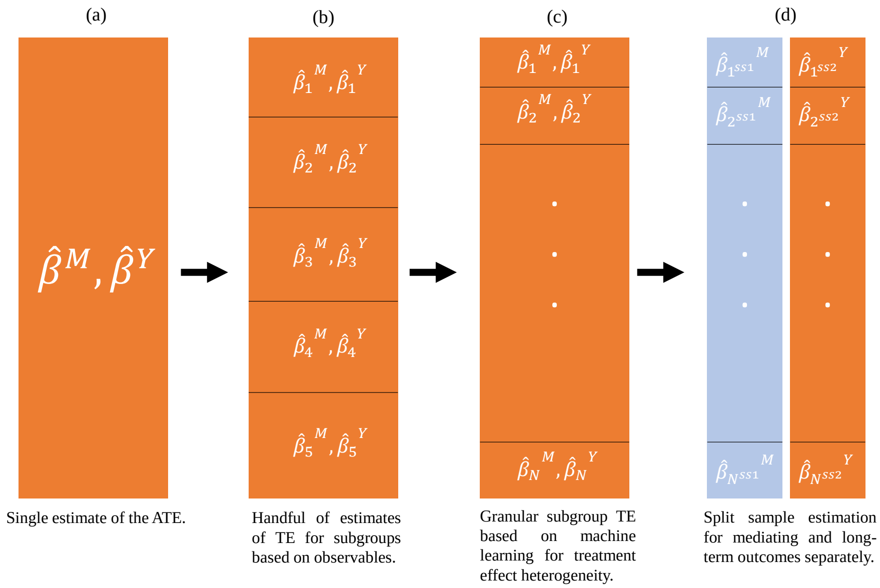


(j) Sociability



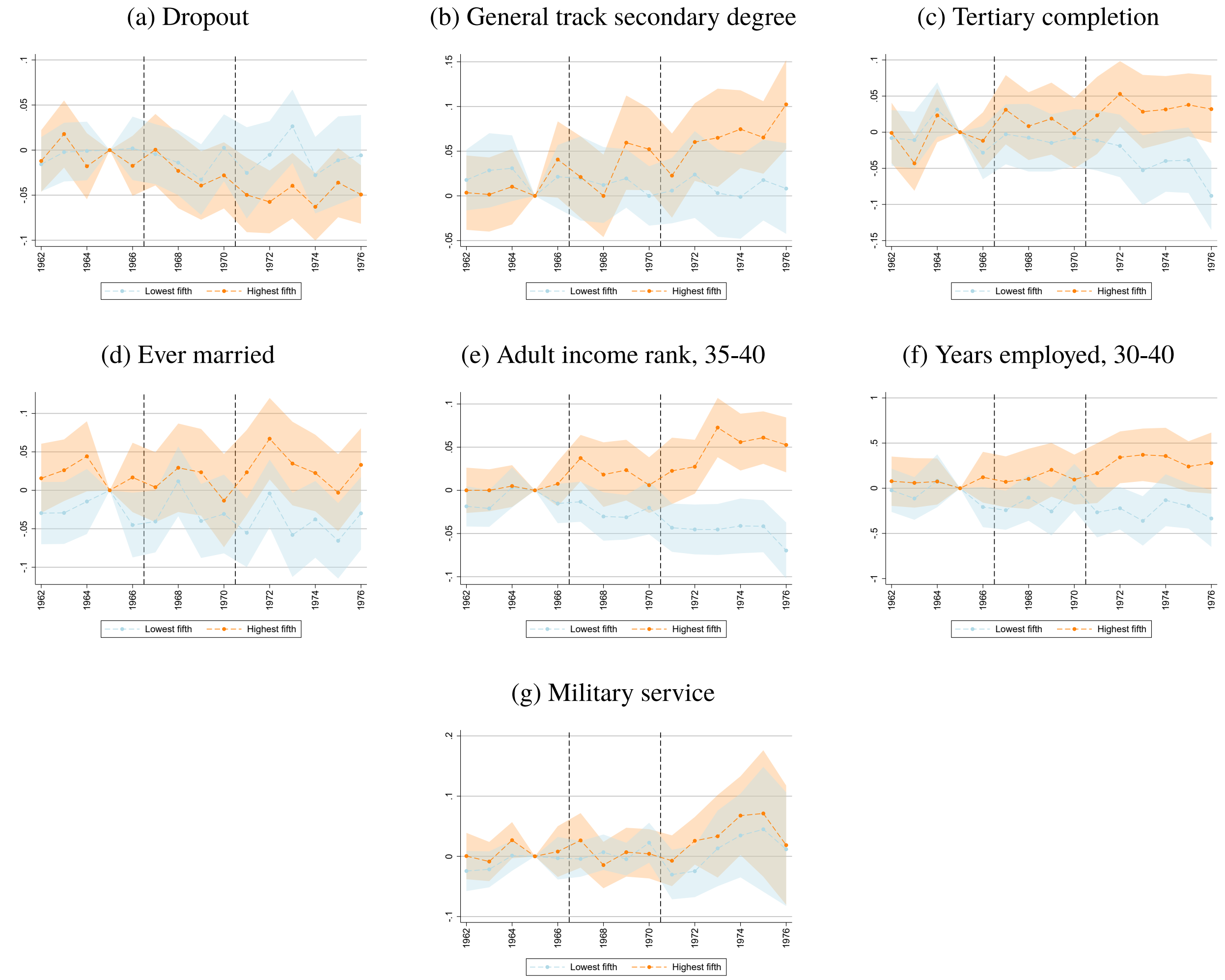
Notes: This figure shows event-study plots for measures of adult skills for children from the richest and poorest fifth of families in our estimation sample using the specification from Equation 5.

Figure 4: Our approach to estimating the covariance in treatment effects between medium-term and long-term outcomes



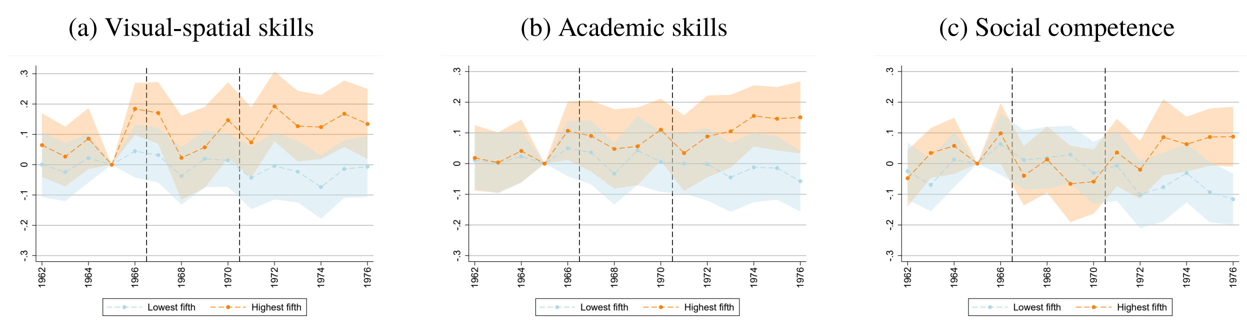
Notes: This diagram depicts the thinking behind our approach to estimating the covariance between medium-term (skills) and long-term (income, educations) treatment effects. As is common in the context of teacher effects, one way to assess the relationship between mediating variables and long-term outcomes is by estimating the covariance in their treatment effects. While insufficient for causal mediation, a relationship between the two is typically assumed for a mediation hypothesis to be true: For some mediator (M) to drive the effects of D on Y, individuals who experience effects on Y must also experience effects on M. As shown in Figure (a), in typical reduced form contexts like our own, researchers often only have one main estimate of average treatment effects for each outcome. In this case, the covariance between medium-term and long-term treatment effects cannot be estimated. One way to overcome this is by dividing the full sample of data into subgroups, and then estimating the covariance across a handful of subgroups – such as, in our case, family income quintile (Figure b). These estimates are, however, likely to be upward biased for three reasons: i) small sample bias in the estimation of the sample covariance; ii) if the same general – but not granular – groups of individuals respond in terms of both outcomes; and, iii) any bias in one estimates of one outcome is likely to exist also in estimates of the other outcome. Generating a large number of subgroups will help overcome the biases in points i) and ii) – but, when done manually, may suffer from insufficient variance in treatment effects across groups or the garden of forking paths in how groups are chosen to maximize variation in these treatment effects. One way to both tie our hands in the creation of these groups as well as maximize the predicted variation in treatment effects across groups is to use machine learning for treatment effect heterogeneity (Figure c). Still, machine learning will not alleviate the problem regarding bias being correlated across outcomes. To mitigate this concern, we divide each machine-learning based subgroup into split samples – one of which we use to estimate treatment effects on medium term outcomes, and the other of which we use to estimate treatment effects for longer-term outcomes (Figure d). Not shown in this figure: To reduce attenuation bias that increases as we estimate an increasingly large number of treatment effects we use empirical Bayes to adjust our estimates of treatment effect correlations; to increase precision, we repeat the split-sample estimates in a number of randomly chosen splits, and then choose the median of these covariances; to check for sensitivity in our estimates of the covariance, we repeat this process for a range of numbers of subgroups, ranging from ten to a hundred.

Figure 5: Event-study plots of long-run outcomes based on predicted treatment effect heterogeneity



Notes: This figure shows event-study plots for measures of adult skills based on the predicted treatment effect heterogeneity ranking, using the specification from Equation 5.

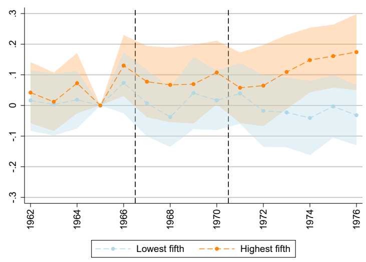
Figure 6: Event-study plots of adult skills based on predicted treatment effect heterogeneity



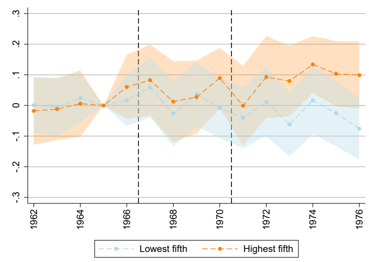
Notes: This figure shows event-study plots for measures of adult skills based on the predicted treatment effect heterogeneity ranking, using the specification from Equation 5.

Figure 7: Event-study plots for measures of adult skills based on predicted treatment effect heterogeneity

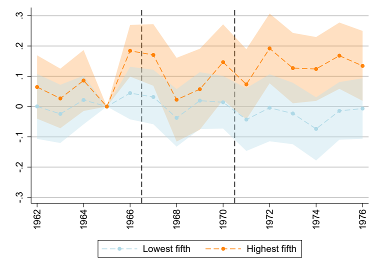
(a) Arithmetic



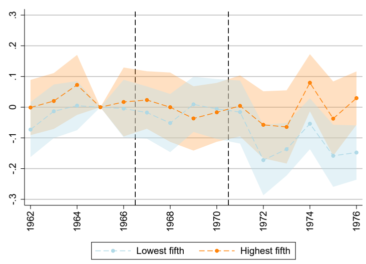
(b) Verbal



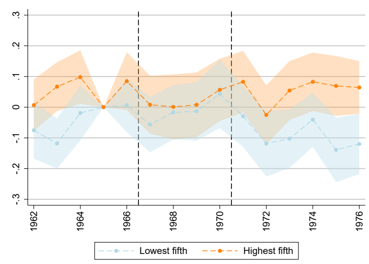
(c) Visual-spatial



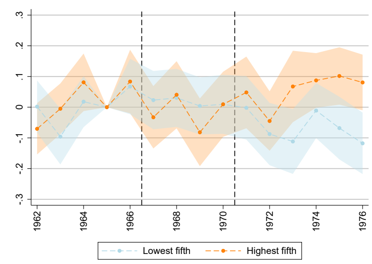
(d) Deliberation



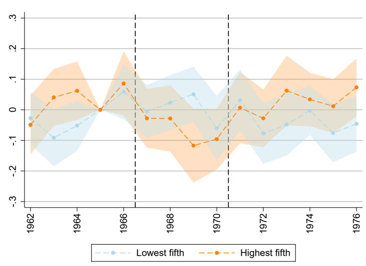
(e) Dutifulness



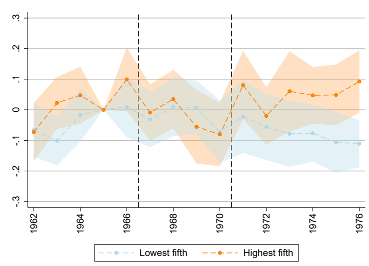
(f) Achievement-striving



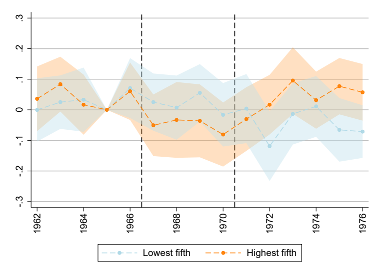
(g) Activity-energy



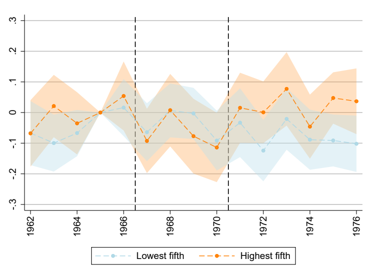
(h) Leadership motivation



(i) Self-confidence

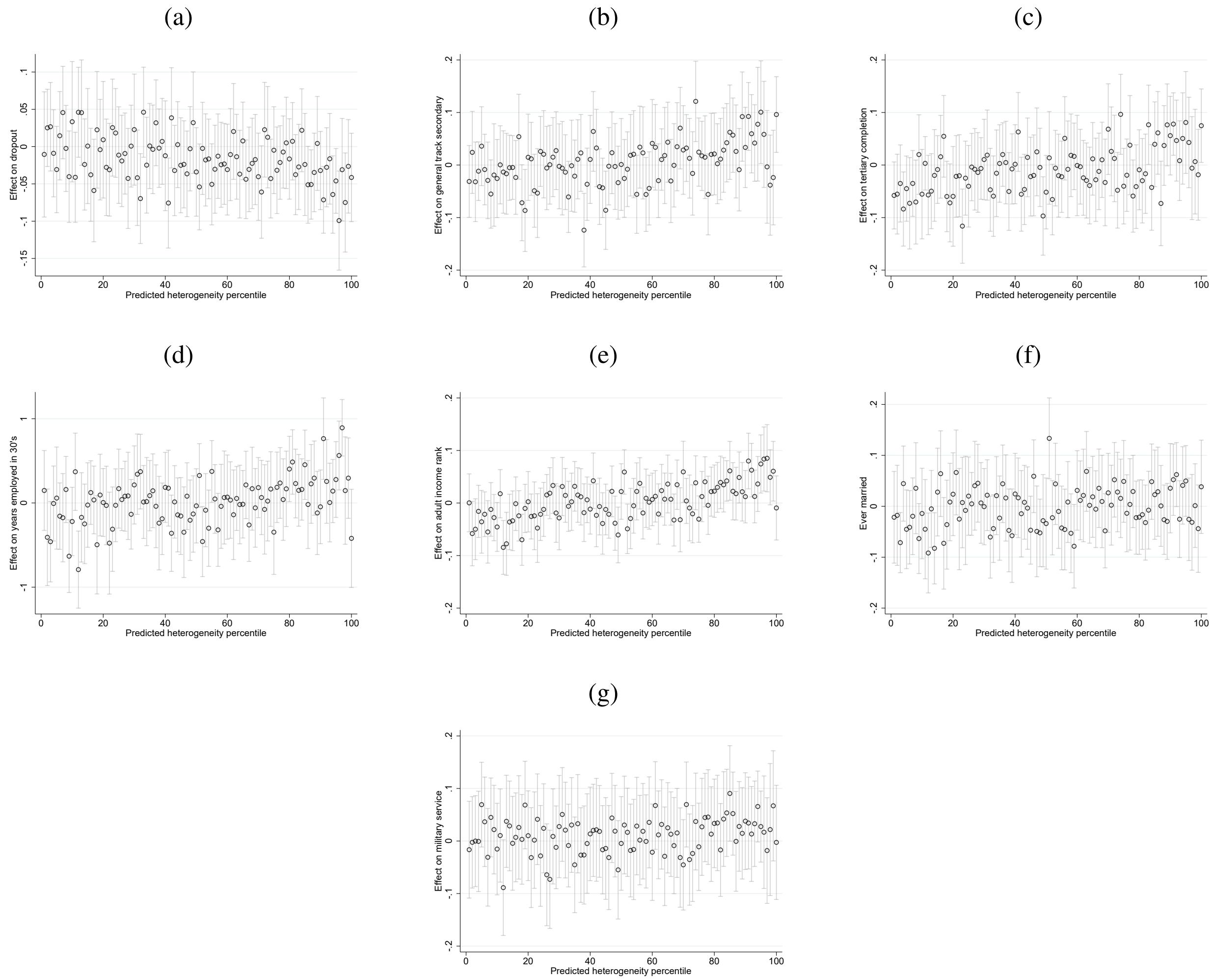


(j) Sociability



Notes: This figure shows event-study plots for measures of adult skills based on the predicted treatment effect heterogeneity ranking, using the specification from Equation 5.

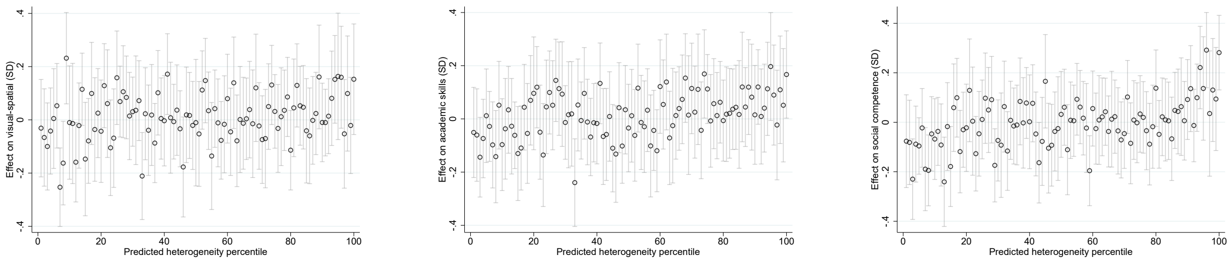
Figure 8: Subgroup treatment effects on long-term outcomes



Notes: This figure plots the granular subgroup treatment effect estimates based on the predicted treatment effect heterogeneity ranking, using the specification from Equation 4.

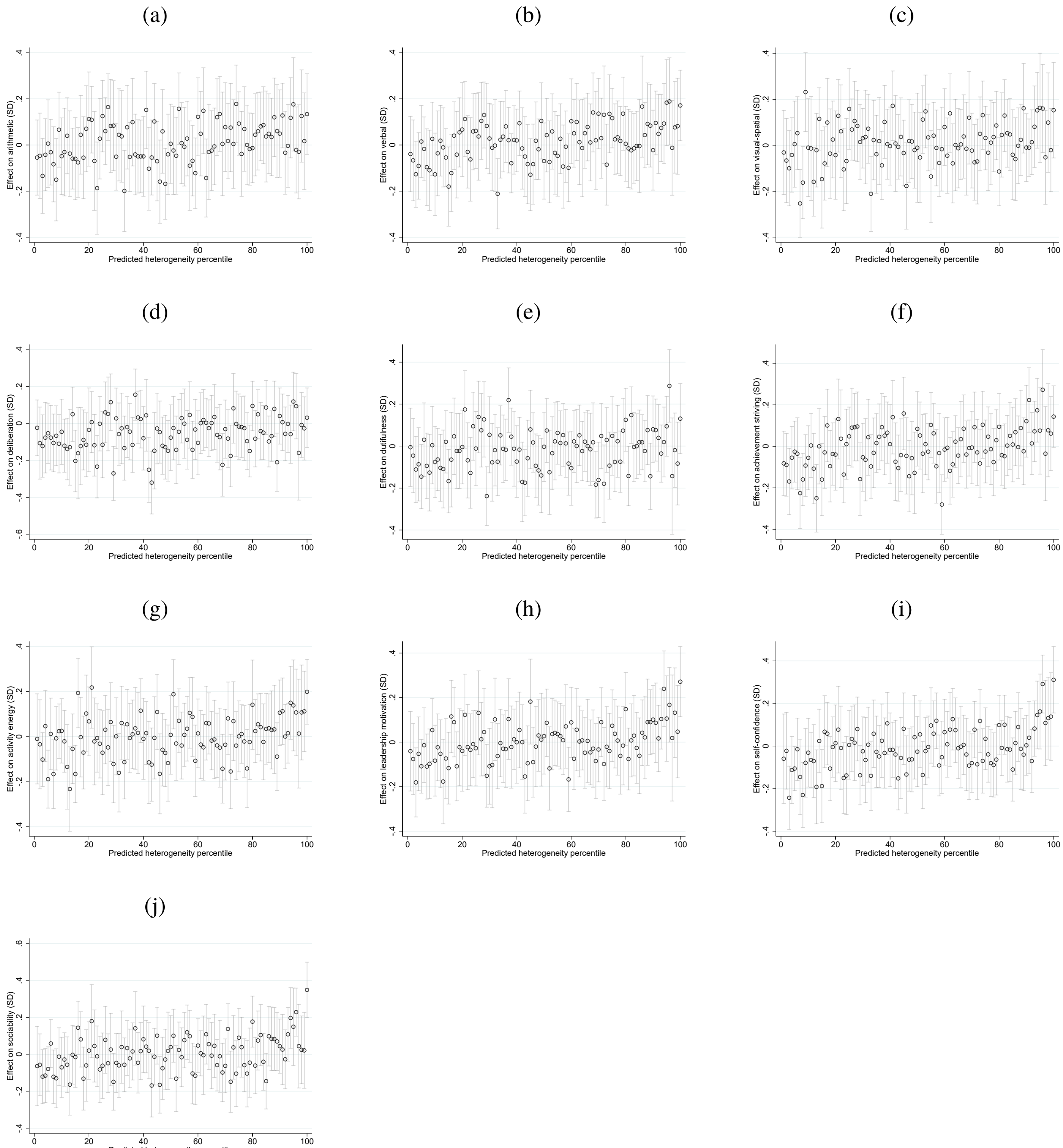
Figure 9: Subgroup treatment effects on main skill outcomes

(a) (b) (c)



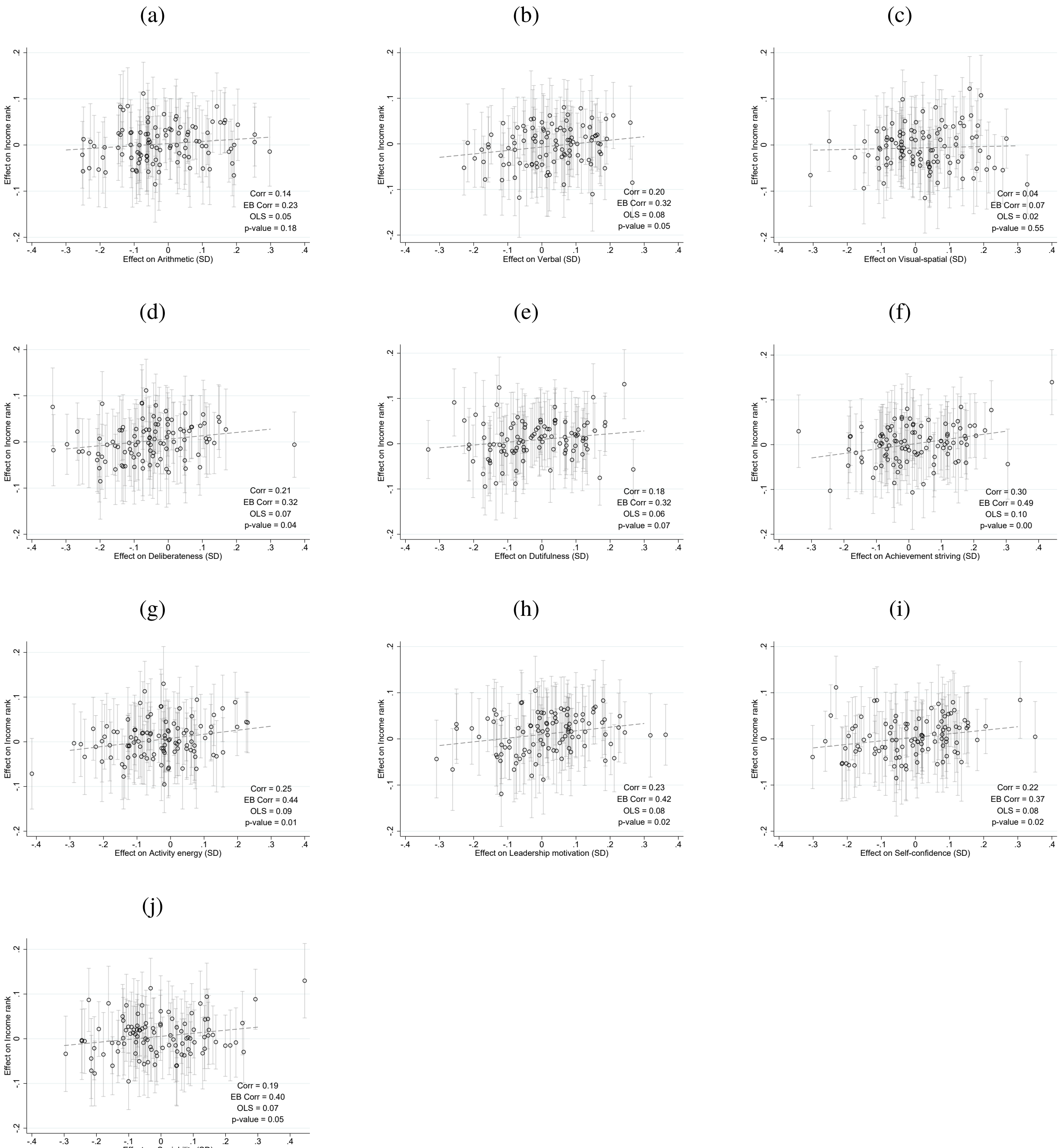
Notes: This figure plots the granular subgroup treatment effect estimates based on the predicted treatment effect heterogeneity ranking, using the specification from Equation 4.

Figure 10: Subgroup treatment effects on measures of adult skills



Notes: This figure plots the granular subgroup treatment effect estimates based on the predicted treatment effect heterogeneity ranking, using the specification from Equation 4.

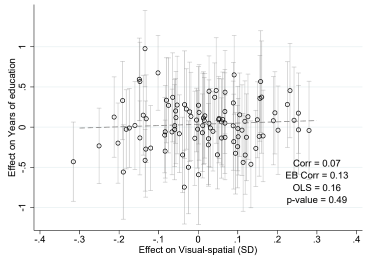
Figure 11: Treatment effects on income rank and treatment effects on measures of adult skills



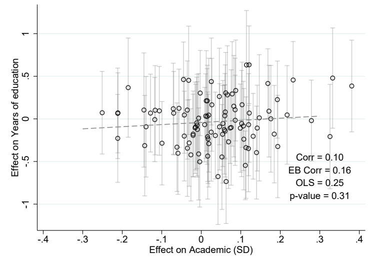
Notes: This figure plots the split-sample relationships between treatment effects on adult income rank (y-axis) and the treatment effects on measures of adult skills (x-axis), for a hundred subgroups based on the predicted treatment effect heterogeneity index. The correlation, raw and adjusted using empirical Bayes, as well as regression coefficients (along with their statistical significance) are reported in the bottom right of each figure.

Figure 12: Treatment effects on education and treatment effects on measures of adult skills

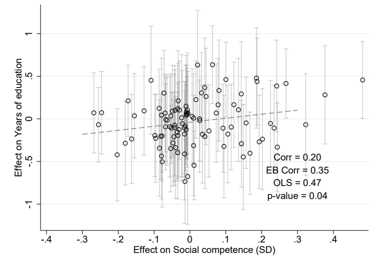
(a) Visual-spatial



(b) Academic

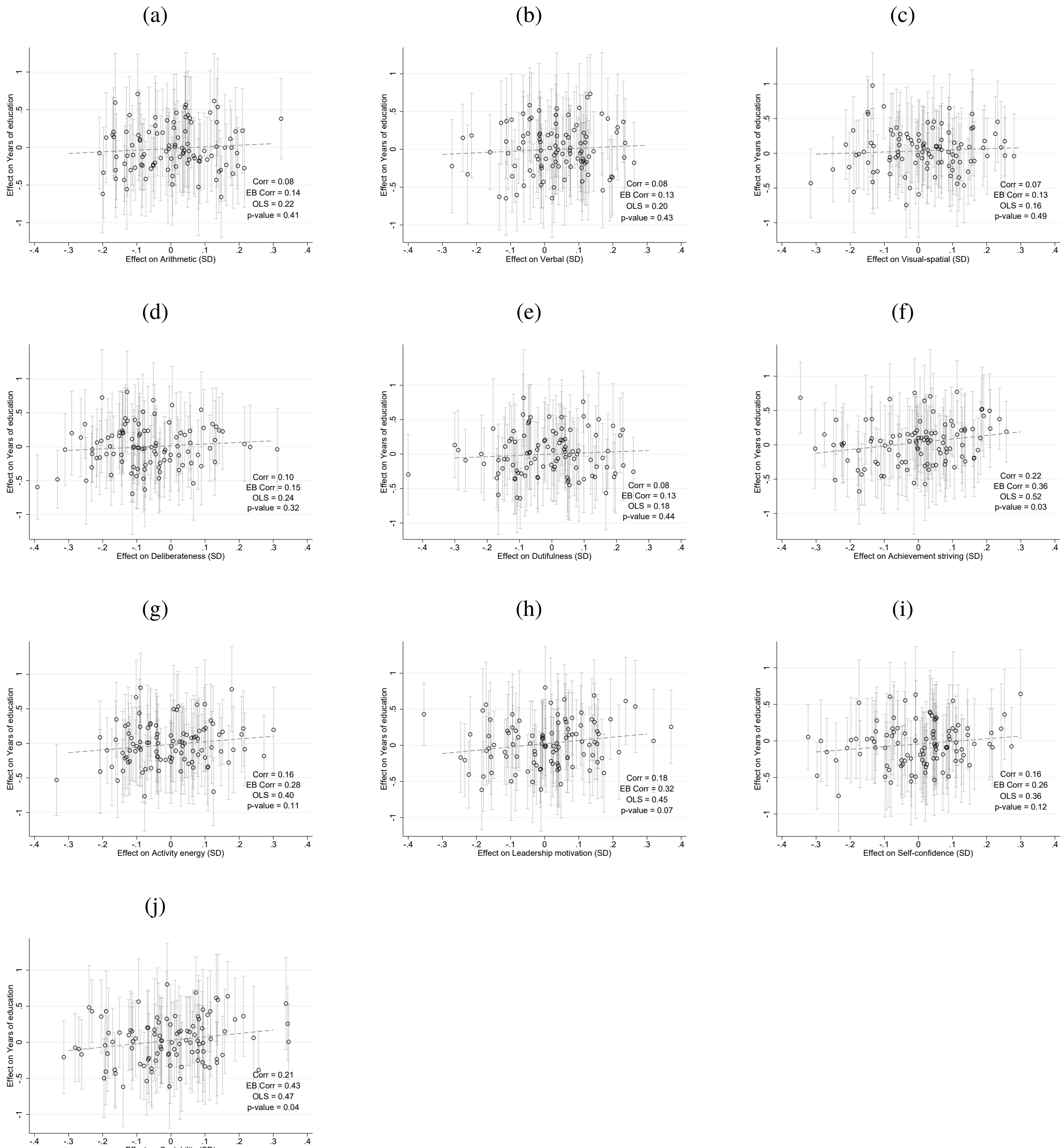


(c) Social competence



Notes: This figure plots the split-sample relationships between treatment effects on adult income rank (y-axis) and the treatment effects on measures of adult skills (x-axis), for a hundred subgroups based on the predicted treatment effect heterogeneity index. The correlation, raw and adjusted using empirical Bayes, as well as regression coefficients (along with their statistical significance) are reported in the bottom right of each figure.

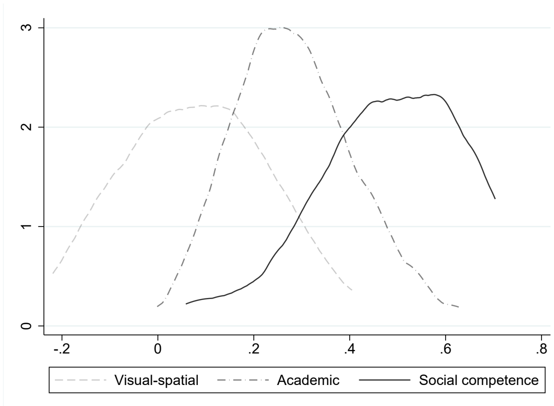
Figure 13: Treatment effects on years of education and treatment effects on measures of adult skills



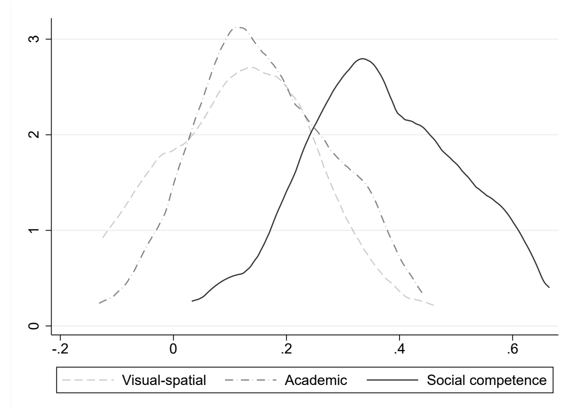
Notes: This figure plots the split-sample relationships between treatment effects on adult income rank (y-axis) and the treatment effects on measures of adult skills (x-axis), for a hundred subgroups based on the predicted treatment effect heterogeneity index. The correlation, raw and adjusted using empirical Bayes, as well as regression coefficients (along with their statistical significance) are reported in the bottom right of each figure.

Figure 14: Split-sample treatment effect correlations

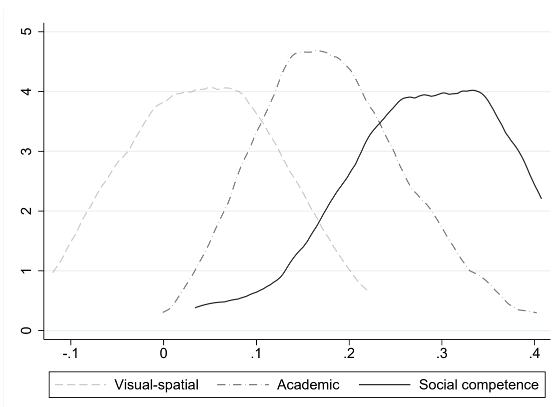
(a) Treatment effects on income, scaled by reliability



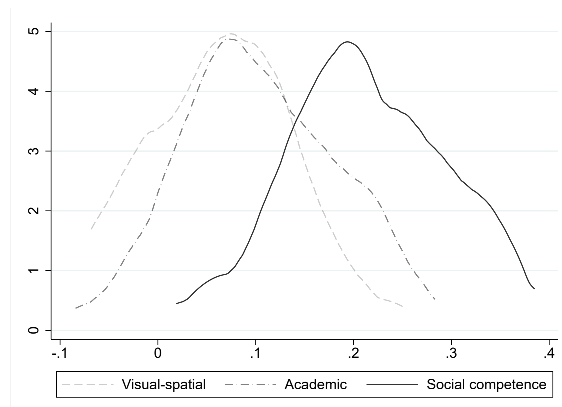
(b) Treatment effects on years of education, scaled by reliability



(c) Treatment effects on income, raw



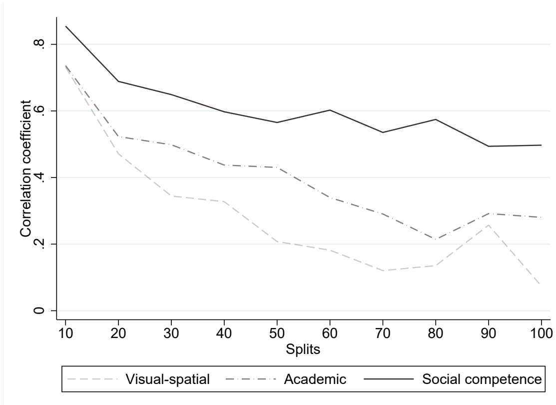
(d) Treatment effects on years of education, raw



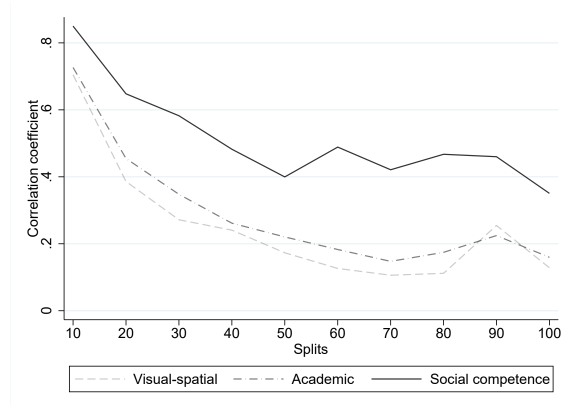
Notes: This figure plots the split-sample correlation coefficient between treatment effects on long-term outcomes and treatment effects on measures of adult skills by the number of subgroups that treatment effects are estimated for. The uppermost figures report the treatment effect correlations adjusted using empirical Bayes; the lower figures report the raw correlations.

Figure 15: Sensitivity to number of splits and empirical Bayes scaling

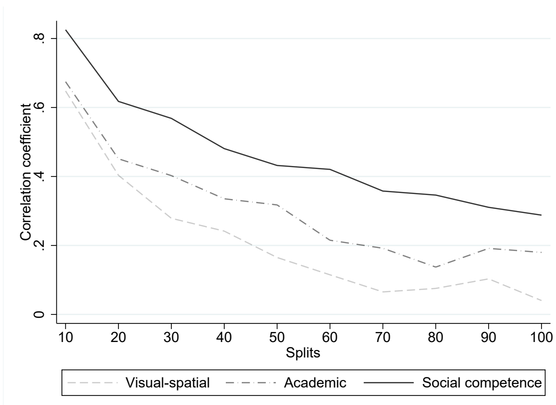
(a) Treatment effects on income, scaled by reliability



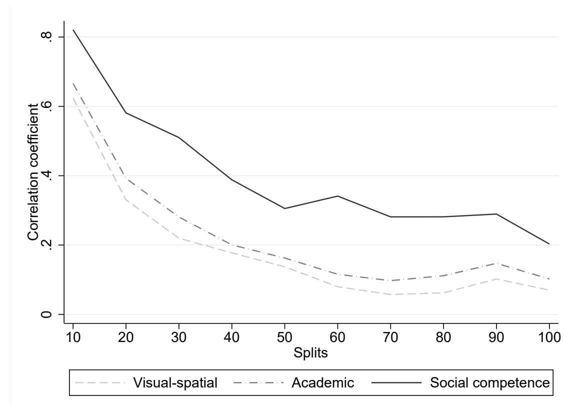
(b) Treatment effects on years of education, scaled by reliability



(c) Treatment effects on income, raw



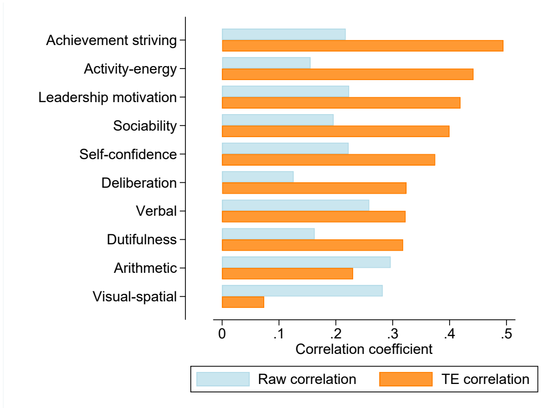
(d) Treatment effects on years of education, raw



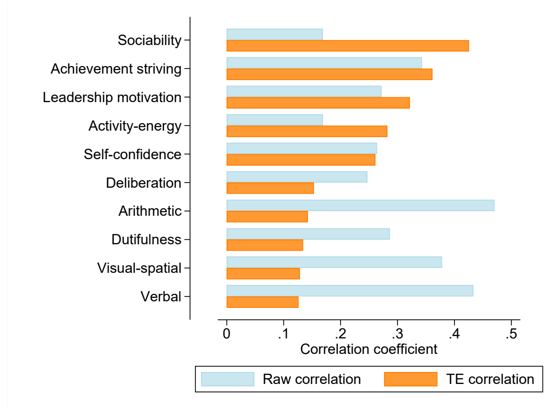
Notes: This figure plots the split-sample correlation coefficient between treatment effects on long-term outcomes and treatment effects on measures of adult skills by the number of subgroups that treatment effects are estimated for. The uppermost figures report the correlations when the estimates of treatment effects on adult skills are adjusted using empirical Bayes; the lower figures report the raw correlations.

Figure 16: Correlations between long-term outcome treatment effects and skill treatment effects compared to raw correlations between outcomes and skills

(a) Adult income rank



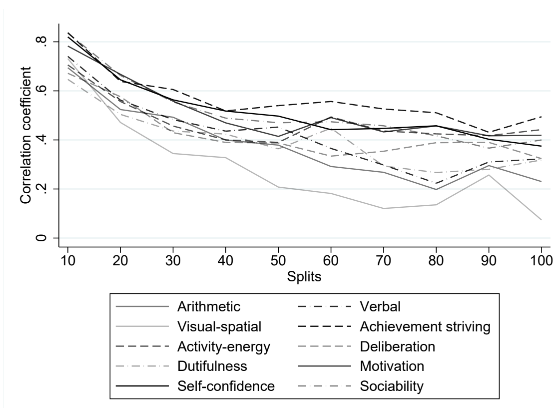
(b) Years of education



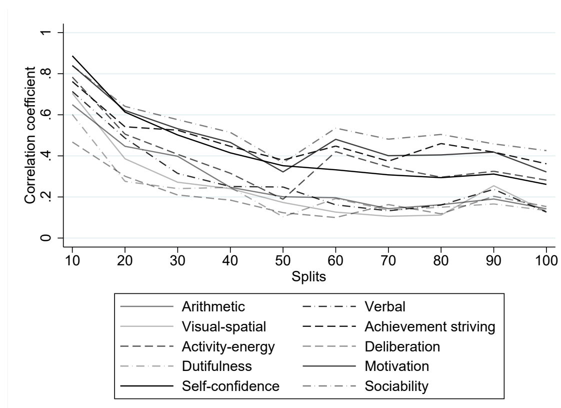
Notes: This figure plots the split-sample correlation coefficient between treatment effects on long-term outcomes (adult income rank and years of education) and treatment effects on measures of adult skills beside raw correlations of the particular skills and long-term outcomes.

Figure 17: Sensitivity to number of splits

(a) Treatment effects on income, scaled by reliability

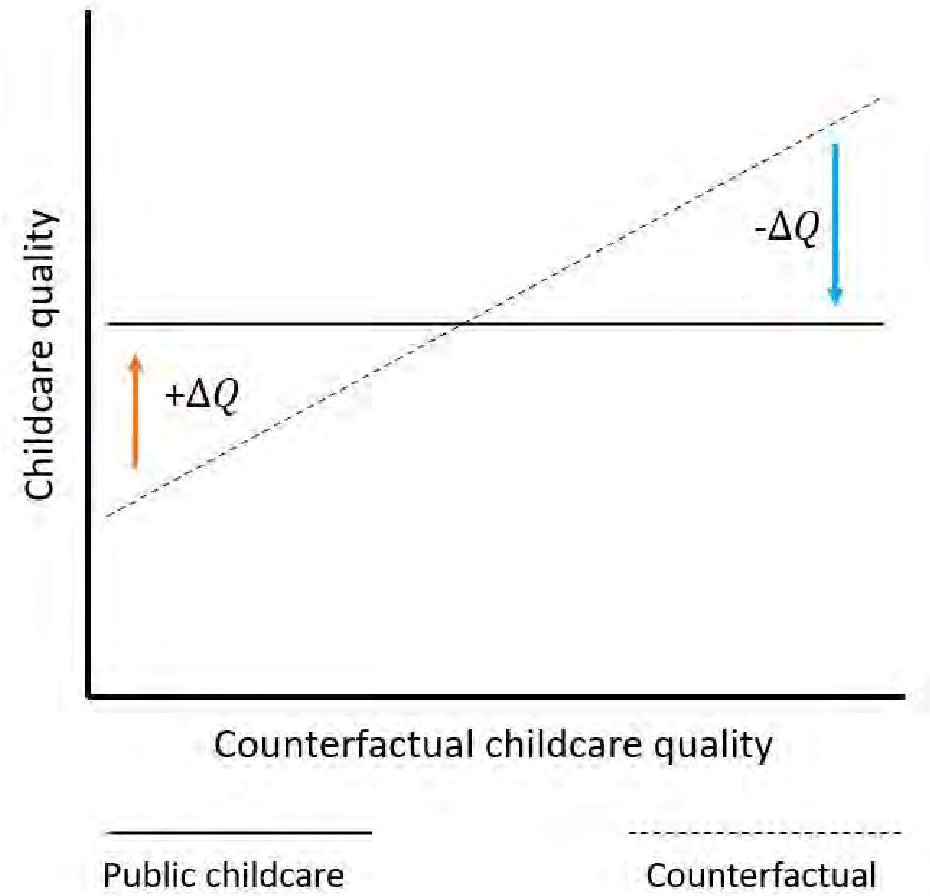


(b) Treatment effects on years of education, scaled by reliability



Notes: This figure plots the split-sample correlation coefficient between treatment effects on long-term outcomes and treatment effects on measures of adult skills by the number of subgroups that treatment effects are estimated for. Treatment effect correlations are adjusted using empirical Bayes to account for attenuation bias.

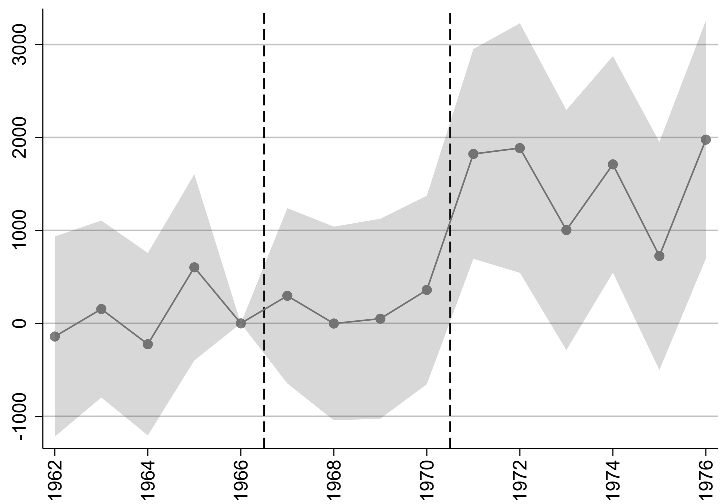
Figure 18: Counterfactual early childhood environments



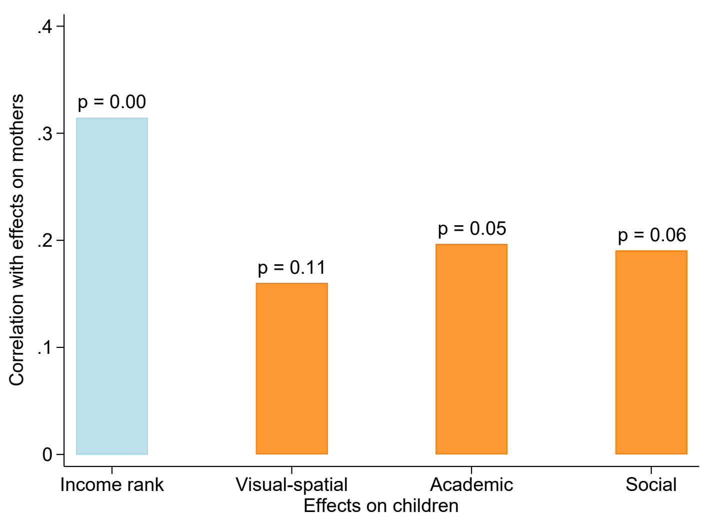
Notes: This diagram presents a conceptual framework for how public childcare may affect children in different ways. The dashed line represents the quality of childcare experienced by children growing up in different early childhood environments, while the solid line represents the quality of early childhood environment if the child attends public childcare. As the framework shows, public childcare can improve outcomes for children with a relatively low quality counterfactual childcare environment, and reduce outcomes for those with a high quality counterfactual environment.

Figure 19: Public childcare access, maternal labor market outcomes, and child development

(a) TE on maternal labor market outcomes

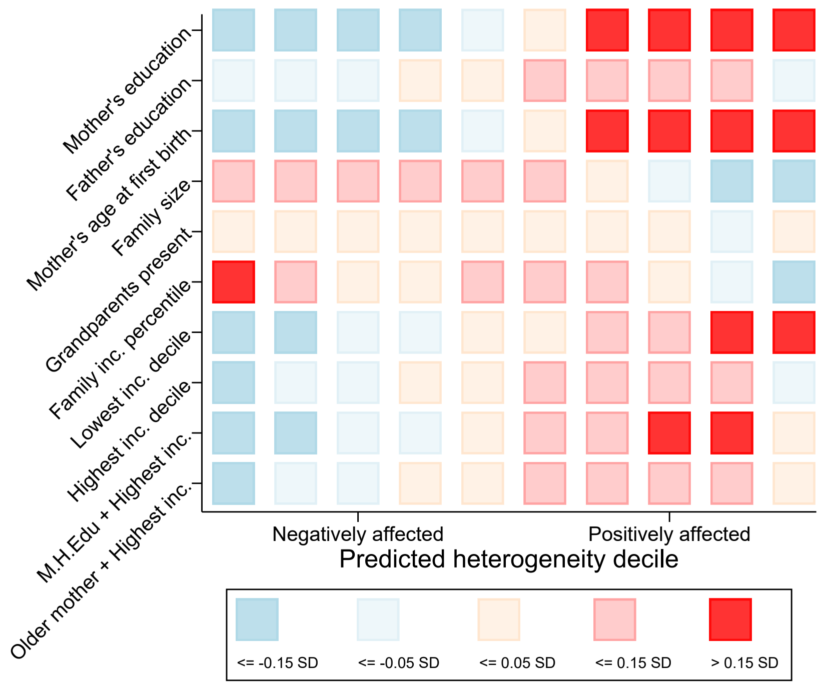


(b) TE correlation with children's TE



Notes: Figures (a) depicts the effects of childcare for mothers, based on the birth-year of their youngest child. Figure (b) correlates the TE for mothers with TE on children's income rank as well as skills. For more on mothers, see Mäkinen and Silliman (2022).

Figure 20: Background characteristics of predicted heterogeneity deciles



Notes: This figure describes the background characteristics of individuals by predicted treatment effect heterogeneity decile. Colors are assigned to each square based on the extent that the mean values of the background covariates differ from the estimation sample mean in terms of standard deviations. Red squares denote larger values in terms of the covariates shown on the left.

2 Tables

Table 1: Estimation sample versus full sample: Outcomes

			Males	
	Full sample	Estimation	Full	Estimation
	(1)	(2)	(3)	(4)
Dropout	0.15	0.14	0.18	0.17
	(0.36)	(0.35)	(0.38)	(0.38)
HS graduate	0.43	0.38	0.34	0.28
	(0.49)	(0.49)	(0.47)	(0.45)
Tertiary education	0.39	0.37	0.31	0.29
	(0.49)	(0.48)	(0.46)	(0.45)
Years of education	12.96	12.91	12.63	12.52
	(2.44)	(2.34)	(2.43)	(2.33)
Income rank at age 35-40	0.50	0.47	0.56	0.53
	(0.29)	(0.29)	(0.31)	(0.31)
Years employed in 30s	7.97	8.01	8.27	8.34
	(2.96)	(2.92)	(2.92)	(2.85)
Ever married	0.63	0.63	0.60	0.59
	(0.48)	(0.48)	(0.49)	(0.49)
Military service	0.81	0.82	0.81	0.82
	(0.39)	(0.39)	(0.39)	(0.39)
Municipalities	463	229	463	229
Individuals	928,500	177,808	174,126	90,434

Notes: This table reports the means and standard deviations of the outcomes for the full and estimation samples in this paper (Columns 1 and 2) and males (Columns 3 and 4).

Table 2: Estimation sample outcomes by the availability of skill data

	Skill data (1)	No skill data (2)
Dropout	0.16	0.26
	(0.36)	(0.44)
HS graduate	0.28	0.28
	(0.45)	(0.45)
Tertiary education	0.30	0.27
	(0.46)	(0.44)
Years of education	12.58	12.24
	(2.27)	(2.53)
Income rank at age 35-40	0.54	0.46
	(0.30)	(0.33)
Years employed in 30s	8.58	7.28
	(2.57)	(3.69)
Ever married	0.60	0.52
	(0.49)	(0.50)
Skill data exists	1.00	0.00
	(0.00)	(0.00)
Municipalities	223	229
Individuals	73,999	16,435

Notes: This table estimates the mean outcomes for the individuals in our sample with (Column 1) and without (Column 2) skill data.

Table 3: Gaps in outcomes between children from rich and poor families

	Poorest fifth of families (1)	Richest fifth of families (2)
	Panel A: Education, marriage, and the labor market	
Dropout	0.22	0.13
	(0.42)	(0.34)
HS graduate	0.20	0.44
	(0.40)	(0.50)
Tertiary education	0.22	0.41
	(0.42)	(0.49)
Years of education	12.11	13.17
	(2.22)	(2.52)
Income rank at age 35-40	0.47	0.59
	(0.31)	(0.31)
Years employed in 30s	7.88	8.63
	(3.19)	(2.60)
Ever married	0.54	0.63
	(0.50)	(0.48)
Skill data exists	0.81	0.81
	(0.39)	(0.39)
	Panel B: Adult skills	
Visual-spatial	-0.21	0.20
	(1.03)	(0.96)
Academic	-0.31	0.15
	(1.02)	(1.00)
Social competence	-0.31	0.10
	(0.98)	(1.00)
Individuals with skill data	15,024	15,114
Individuals	18,133	18,156

Notes: This table presents the mean outcomes for males born to the poorest and richest fifth of families in our estimation sample.

Table 4: Information on childcare teachers in 1980

	Treatment	Control
Age	27.11	29.29
	(11.42)	(9.54)
Male	0.11	0.07
	(0.31)	(0.25)
Married	0.46	0.56
	(0.50)	(0.50)
Kids at home	1.07	0.78
	(1.10)	(1.00)
Childcare-age kids	0.47	0.28
	(0.75)	(0.53)
At least high school	0.88	0.87
	(0.33)	(0.34)
Post-secondary degree	0.80	0.79
	(0.40)	(0.41)
Under the age of 19	0.11	0.06
	(0.31)	(0.24)
Income rank	61.87	64.19
	(15.51)	(14.96)
Months employed	10.02	10.09
	(3.32)	(3.28)
Childcare teachers	256	131
Municipalities	89	140

Notes: This table reports data on childcare teachers for the year 1980, the first year that childcare teachers are included as an occupation distinct from kindergarten teachers.

Table 5: Treatment effects for females

	ATE (1)	Treat X family inc. percentile (2)	Effect for poorest fifth (3)	Effect at the median (4)	Effect for richest fifth (5)
		Panel A: Effects on education and the labor market			
Dropout	0.005	0.030**	-0.003	0.006	0.015*
	(0.006)	(0.010)	(0.007)	(0.006)	(0.007)
HS graduate	-0.001	-0.166***	0.042***	-0.008	-0.058***
	(0.009)	(0.017)	(0.010)	(0.009)	(0.011)
Tertiary education	-0.013	-0.134***	0.021*	-0.019*	-0.060***
	(0.010)	(0.017)	(0.011)	(0.010)	(0.011)
Years of education	-0.053	-0.610***	0.102*	-0.081	-0.264***
	(0.045)	(0.071)	(0.048)	(0.045)	(0.051)
Income rank	0.004	-0.044***	0.015**	0.002	-0.011*
	(0.005)	(0.008)	(0.005)	(0.005)	(0.005)
Years employed in 30's	0.015	-0.285**	0.091	0.006	-0.080
	(0.049)	(0.086)	(0.056)	(0.049)	(0.054)
Ever married	-0.003	-0.041**	0.008	-0.004	-0.017
	(0.007)	(0.015)	(0.008)	(0.007)	(0.009)
Municipalities	229	229			
Individuals	87,374	87,374			

Notes: Column 1 of this table reports the coefficient β_2 from Equation 3. This coefficient measures the difference in effect of public childcare access between a child at the very bottom of the family income distribution compared to a child at the very top of the family income distribution. Column 2(4) evaluates this expected treatment effect for the fifth of children from the poorest(richest) families. Column 3 evaluates the treatment effect for families at the middle of the family income distribution. *= p<0.05, **=p<0.01,***<p<0.001.

Table 6: Treatment effects by family income (raw measures of skills)

	Treat X family inc. percentile (1)	Effect for poorest fifth (2)	Effect at the median (3)	Effect for richest fifth (4)
Arithmetic	-0.221***	0.101***	0.013	-0.075**
	(0.032)	(0.022)	(0.019)	(0.023)
Verbal	-0.255***	0.118***	0.016	-0.086***
	(0.030)	(0.021)	(0.018)	(0.023)
Visual-spatial	-0.190***	0.088***	0.012	-0.064**
	(0.032)	(0.021)	(0.018)	(0.022)
Achievement striving	-0.194***	0.071**	-0.006	-0.084***
	(0.036)	(0.022)	(0.018)	(0.024)
Activity energy	-0.167***	0.064**	-0.003	-0.069**
	(0.032)	(0.020)	(0.018)	(0.024)
Deliberateness	-0.055	-0.031	-0.053***	-0.075***
	(0.032)	(0.019)	(0.016)	(0.021)
Dutifulness	-0.149***	0.044*	-0.016	-0.075**
	(0.032)	(0.019)	(0.017)	(0.023)
Leadership motivation	-0.254***	0.101***	-0.001	-0.102***
	(0.034)	(0.020)	(0.019)	(0.026)
Self-confidence	-0.238***	0.081***	-0.014	-0.109***
	(0.030)	(0.020)	(0.017)	(0.022)
Sociability	-0.242***	0.097***	0.000	-0.097***
	(0.030)	(0.020)	(0.018)	(0.022)
Municipalities	223			
Individuals	75,996			

Notes: Column 1 of this table reports the coefficient β_2 from Equation 3. This coefficient measures the difference in effect of public childcare access between a child at the very bottom of the family income distribution compared to a child at the very top of the family income distribution. Column 2(4) evaluates this expected treatment effect for the fifth of children from the poorest(richest) families. Column 3 evaluates the treatment effect for families at the middle of the family income distribution. *= p<0.05, **=p<0.01,***<p<0.001.

Table 7: Treatment effects by family income percentile: Robustness and bounds

	Family inc. only (1)	Demographic covariates (2)	Dem. + Edu. covariates (3)	Restricted sample (4)	Missings as low (5)	Missings imputed (6)	Missings as high (7)
	Panel A: Effects on education, marriage, and the labor market						
Dropout	0.051***	0.052***	0.032**	0.049***			
	(0.011)	(0.011)	(0.011)	(0.013)			
HS graduate	-0.102***	-0.095***	-0.040**	-0.101***			
	(0.016)	(0.015)	(0.013)	(0.017)			
Tertiary education	-0.106***	-0.104***	-0.060***	-0.104***			
	(0.015)	(0.014)	(0.013)	(0.015)			
Years of education	-0.564***	-0.554***	-0.301***	-0.532***			
	(0.076)	(0.072)	(0.066)	(0.078)			
Income rank	-0.051***	-0.046***	-0.032**	-0.054***			
	(0.010)	(0.010)	(0.010)	(0.010)			
Years employed in 30's	-0.378***	-0.412***	-0.366***	-0.374***			
	(0.089)	(0.087)	(0.085)	(0.080)			
Ever married	-0.054***	-0.052***	-0.038*	-0.064***			
	(0.015)	(0.015)	(0.015)	(0.016)			
Military service	0.022	0.017	0.014				
	(0.013)	(0.012)	(0.012)				
Municipalities	229			223			
Individuals	90,434			75,996			
	Panel B: Effects on skills						
Visual-spatial	-0.190***	-0.171***	-0.095**	-0.190***	-0.109**	-0.192***	-0.202***
	(0.032)	(0.032)	(0.031)	(0.032)	(0.039)	(0.029)	(0.036)
Academic	-0.261***	-0.241***	-0.146***	-0.261***	-0.171***	-0.258***	-0.265***
	(0.032)	(0.030)	(0.028)	(0.032)	(0.037)	(0.030)	(0.039)
Social competence	-0.269***	-0.246***	-0.173***	-0.269***	-0.160***	-0.253***	-0.292***
	(0.035)	(0.034)	(0.032)	(0.035)	(0.039)	(0.031)	(0.039)
Municipalities	223				223	223	223
Individuals	75,996				90,434	90,434	90,434

Notes: Column 1 of this table reports the coefficient β_2 from Equation 3. Columns 2-4 include successively more background covariates to the specification: demographic controls, and parental education controls. Column 5 reports effects on registry data outcomes using a sample restricted to individuals for whom there exists data on skills. Columns 5 and 7 report estimates where missing skill measures are imputed as extremely low or high outcomes, and column 6 reports estimates where missing skill measures are imputed using later measures from registry data. *= p<0.05, **=p<0.01,***<p<0.001.

Table 8: Robustness to regional trends

	Original specification (1)	Regional trends (2)	Within region (3)	Within municipality (4)
	Panel A: Effects on education, marriage, and the labor market			
Dropout	0.051***	0.058***	0.058***	0.057***
	(0.011)	(0.011)	(0.011)	(0.012)
HS graduate	-0.102***	-0.105***	-0.107***	-0.109***
	(0.016)	(0.016)	(0.016)	(0.017)
Tertiary education	-0.106***	-0.108***	-0.110***	-0.110***
	(0.015)	(0.014)	(0.014)	(0.017)
Years of education	-0.564***	-0.595***	-0.604***	-0.598***
	(0.076)	(0.068)	(0.067)	(0.080)
Income rank	-0.051***	-0.048***	-0.049***	-0.048***
	(0.010)	(0.010)	(0.010)	(0.011)
Years employed in 30's	-0.378***	-0.338***	-0.340***	-0.348***
	(0.089)	(0.087)	(0.087)	(0.088)
Ever married	-0.054***	-0.059***	-0.059***	-0.058***
	(0.015)	(0.015)	(0.015)	(0.015)
Military service	0.022	0.028*	0.029*	0.029*
	(0.013)	(0.012)	(0.012)	(0.012)
Municipalities	229			
Individuals	90,434			
	Panel B: Effects on skills			
Visual-spatial	-0.190***	-0.186***	-0.190***	-0.185***
	(0.032)	(0.032)	(0.032)	(0.033)
Academic	-0.261***	-0.260***	-0.264***	-0.261***
	(0.032)	(0.031)	(0.031)	(0.032)
Social competence	-0.269***	-0.262***	-0.265***	-0.263***
	(0.035)	(0.033)	(0.034)	(0.035)
Municipalities	223			
Individuals	75,996			

Notes: This table reports the coefficient β_2 from Equation 3. Column one reports results from the original specification. Column two adds a parametric measure of regional trends as controls. Column three controls for regional trends non-parametrically, through an interaction term between year and region. Column three controls for municipality specific trends non-parametrically, through a triple-differences design using an interaction term between year and municipality.

Table 9: Results using alternative DiD estimators

	ATE		T X Fam. Inc. Pctile.	
	Binary	Continuous	Binary	Continuous
	(1)	(2)	(3)	(4)
	Panel A: Effects on education, marriage, and the labor market			
Dropout	-0.011*	-0.071**	0.043***	0.178***
	(0.005)	(0.023)	(0.008)	(0.037)
HS graduate	0.005	0.021	-0.107***	-0.411***
	(0.005)	(0.022)	(0.010)	(0.053)
Tertiary education	-0.003	-0.010	-0.119***	-0.456***
	(0.005)	(0.021)	(0.010)	(0.048)
Years of education	0.017	0.163	-0.573***	-2.234***
	(0.032)	(0.124)	(0.051)	(0.252)
Income rank	-0.000	-0.008	-0.055***	-0.204***
	(0.005)	(0.017)	(0.007)	(0.032)
Years employed (30's)	-0.022	-0.143	-0.405***	-1.331***
	(0.040)	(0.143)	(0.056)	(0.290)
Ever married	-0.001	-0.006	-0.044***	-0.178***
	(0.005)	(0.019)	(0.010)	(0.040)
Military service	0.012	0.000	0.010	0.065
	(0.018)	(0.064)	(0.008)	(0.035)
Municipalities	388	388	388	388
Individuals	199,328	199,328	199,328	199,328
	Panel B: Effects on skills			
Visual-spatial	0.017	0.075	-0.225***	-0.937***
	(0.011)	(0.043)	(0.021)	(0.097)
Academic	0.006	0.047	-0.282***	-1.130***
	(0.012)	(0.047)	(0.023)	(0.122)
Social competence	-0.004	-0.040	-0.247***	-1.040***
	(0.012)	(0.048)	(0.022)	(0.113)
Municipalities	382	382	382	382
Individuals	164,602	164,602	164,602	164,602

Notes: This table reports the main results using alternative DiD estimators. Columns 1 and 2 report estimates of ATE. Columns 3 and 4 report TE by family income percentile. Columns 1 and 3 differ from the main estimates in that they make use of all rural municipalities through a binary staggered (stacked) approach where the treatment indicator takes the value of 1 after the first childcare spots open in that municipality and 0 before then. Columns 2 and 4 report estimates that, again, take advantage of the whole sample through a staggered (stacked) approach, but where treatment is defined using a continuous variable that measures the extent of childcare coverage in the municipality for children aged three to six. *= p<0.05, **=p<0.01,***<p<0.001.

Table 10: Sibling correlations in skills

	Sibling correlation
	Panel A: Main outcomes
Visual-spatial	0.37
Academic	0.47
Social competence	0.34
	Panel B: Cognitive measures
Arithmetic	0.44
Verbal	0.42
Visual-spatial	0.37
	Panel C: Socio-emotional measures
Achievement striving	0.27
Activity energy	0.23
Deliberateness	0.19
Dutifulness	0.24
Leadership motivation	0.33
Self-confidence	0.26
Sociability	0.24
Sibling pairs	69,015

Notes: This table presents the correlations between skills across siblings, using data from the full sample.

Table 11: Horse race between subgroup skill effects

	Income rank (1)	Years of education (2)
Visual-spatial	-0.05	-0.08
	(0.04)	(0.31)
p-value	0.29	0.64
Academic	0.06	0.09
	(0.05)	(0.33)
p-value	0.23	0.57
Social competence	0.09	0.46
	(0.04)	(0.25)
p-value	0.01	0.06
Groups	100	100
Repetitions	51	51

Notes: This table presents the median coefficient estimates, standard errors and p-values of fifty-one split-sample regressions where one hundred granular subgroup effects of all three dimensions of skills are included on the right side of the equation at the same time.

Table 12: Decomposition results, skills, education, and income

	Treatment X family income (skill data sample)			
	(1)	(2)	(3)	(4)
Income rank	-0.054***	-0.037***	-0.027**	-0.027**
	(0.010)	(0.009)	(0.009)	(0.009)
Portion of effect explained		0.305	0.506	0.505
Years of education	-0.532***	-0.336***	-0.186**	-0.186**
	(0.078)	(0.076)	(0.068)	(0.067)
Portion of effect explained		0.367	0.651	0.650
Social competence		Yes	Yes	Yes
Academic skills			Yes	Yes
Visual-spatial skills				Yes
Municipalities	223	223	223	223
Individuals	73,999	73,999	73,999	73,999

Notes: This table presents results from a decomposition exercise where mediating outcomes (skills) are added to the right side of the main estimating equation. The first row of both panels reports the main coefficient when these variables are included, and the second row reports how the magnitude of this coefficient compares to the coefficient from the main estimate.

Table 13: Decomposition results, each skill separately

	Treatment X family income (skill data sample)			
	(1)	(2)	(3)	(4)
Income rank	-0.054***	-0.037***	-0.032**	-0.040***
	(0.010)	(0.009)	(0.009)	(0.009)
Portion of effect explained		0.305	0.396	0.247
Years of education	-0.532***	-0.336***	-0.247***	-0.377***
	(0.078)	(0.076)	(0.067)	(0.070)
Portion of effect explained		0.367	0.535	0.291
Social competence		Yes		
Academic skills			Yes	
Visual-spatial skills				Yes
Municipalities	223	223	223	223
Individuals	73,999	73,999	73,999	73,999

Notes: This table presents results from a decomposition exercise where mediating outcomes (skills) are added to the right side of the main estimating equation. The first row of both panels reports the main coefficient when these variables are included, and the second row reports how the magnitude of this coefficient compares to the coefficient from the main estimate.

Table 14: Decomposition results, direct effects vs. dynamic complementarity

	Treatment X family income (skill data sample)				
	(1)	(2)	(3)	(4)	(5)
Years of education	-0.054***	-0.028**	-0.019*	-0.032**	-0.027**
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Portion of effect explained		0.473	0.638	0.396	0.506
Years of education		Yes	Yes		
Academic skills				Yes	Yes
Social competence			Yes		Yes
Municipalities	223	223	223	223	223
Individuals	73,999	73,999	73,999	73,999	73,999

Notes: This table presents results from a decomposition exercise where educational and academic skills are added to the right-side of the main estimating equation with and without measures of social skills. The first row of both panels reports the main coefficient when these variables are included, and the second row reports how the magnitude of this coefficient compares to the coefficient from the main estimate.

Table 15: Associations between income rank and tasks

	Adult income rank	
	(1)	(2)
	Panel A: Raw correlations	
Social non-routine analytic	0.24	
Social non-routine manual	0.12	
Cognitive non-routine	0.35	
Cognitive routine	0.07	
Manual non-routine	-0.29	
Manual routine	-0.19	
	Panel B: OLS coefficients, individually and jointly	
Social non-routine analytic	0.077	0.020
	(0.000)	(0.001)
Social non-routine manual	0.035	-0.016
	(0.000)	(0.001)
Cognitive non-routine	0.120	0.100
	(0.001)	(0.001)
Cognitive routine	0.025	0.044
	(0.001)	(0.001)
Manual non-routine	-0.083	-0.091
	(0.000)	(0.001)
Manual routine	-0.054	0.056
	(0.000)	(0.001)
Adjusted R-squared		0.175
Observations		386,724

Notes: This table presents the raw correlations between income rank and occupational task content.

Table 16: Associations between skills and tasks

	Visual-spatial	Academic	Social Competence
	(1)	(2)	(3)
	Panel A: Raw correlations		
Social non-routine analytic	0.17	0.23	0.24
Social non-routine manual	0.17	0.25	0.25
Cognitive non-routine	0.29	0.35	0.30
Cognitive routine	-0.04	-0.07	-0.08
Manual non-routine	-0.27	-0.37	-0.31
Manual routine	-0.25	-0.34	-0.30
	Panel B: OLS coefficients, jointly regressed		
Social non-routine analytic	0.013	0.133	0.155
	(0.002)	(0.002)	(0.002)
Social non-routine manual	-0.035	0.197	0.175
	(0.002)	(0.002)	(0.002)
Cognitive non-routine	0.054	0.199	0.153
	(0.002)	(0.002)	(0.001)
Cognitive routine	0.012	-0.038	-0.051
	(0.002)	(0.002)	(0.001)
Manual non-routine	-0.020	-0.283	-0.188
	(0.002)	(0.002)	(0.002)
Manual routine	-0.019	-0.257	-0.202
	(0.002)	(0.002)	(0.002)
Observations			334,301

Notes: This table presents the raw correlations between skills and occupational task content.

Table 17: Decomposition results, skills and tasks

	Treatment X family income (skill and task data sample)				
	(1)	(2)	(3)	(4)	(5)
Social non-routine analytic	-0.155***	-0.108**	-0.090**	-0.090**	-0.075*
	(0.032)	(0.031)	(0.030)	(0.031)	(0.031)
Portion of effect explained		0.303	0.419	0.418	0.514
Social non-routine manual	-0.187***	-0.131***	-0.110**	-0.110**	-0.093**
	(0.033)	(0.031)	(0.031)	(0.031)	(0.032)
Portion of effect explained		0.295	0.412	0.411	0.500
Cognitive non-routine	-0.139***	-0.084**	-0.055*	-0.055*	-0.030
	(0.029)	(0.027)	(0.026)	(0.026)	(0.025)
Portion of effect explained		0.395	0.605	0.603	0.783
Manual non-routine	0.203***	0.135***	0.098**	0.098**	0.064
	(0.038)	(0.037)	(0.036)	(0.036)	(0.035)
Portion of effect explained		0.334	0.518	0.517	0.685
Manual routine	0.190***	0.119***	0.084*	0.085*	0.055
	(0.035)	(0.033)	(0.033)	(0.033)	(0.032)
Portion of effect explained		0.372	0.556	0.554	0.712
Social competence		Yes	Yes	Yes	Yes
Academic skills			Yes	Yes	Yes
Visual-spatial skills				Yes	Yes
Education					Yes
Municipalities	222	222	222	222	222
Individuals	65,545	65,545	65,545	65,545	65,545

Notes: This table presents results from a decomposition exercise where skill-outcomes are added to the right side of the main estimating equation. The first row of both panels reports the main coefficient when these variables are included, and the second row reports how the magnitude of this coefficient compares to the coefficient from the main estimate.

Table 18: Decomposition results, tasks and income

	Treatment X family income (task data sample)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income rank	-0.048***	-0.037***	-0.038***	-0.032***	-0.028**	-0.023**	-0.023**	-0.014
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Portion of effect explained		0.231	0.210	0.339	0.414	0.517	0.521	0.707
Social non-routine analytic		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Social non-routine manual			Yes	Yes	Yes	Yes	Yes	Yes
Cognitive non-routine				Yes	Yes	Yes	Yes	Yes
Cognitive routine					Yes	Yes	Yes	Yes
Manual non-routine						Yes	Yes	Yes
Manual routine							Yes	Yes
Education								Yes
Municipalities	229	229	229	229	229	229	229	229
Individuals	77,154	77,154	77,154	77,154	77,154	77,154	77,154	77,154

Notes: This table presents results from a decomposition exercise where task outcomes are added to the right side of the main estimating equation. The first row of both panels reports the main coefficient when these variables are included, and the second row reports how the magnitude of this coefficient compares to the coefficient from the main estimate.

Table 19: Decomposition results, skills and tasks together

	Treatment X family income (skill and task data sample)			
	(1)	(2)	(3)	(4)
Income rank	-0.047***	-0.025**	-0.027**	-0.019*
	(0.010)	(0.009)	(0.009)	(0.009)
Portion of effect explained		0.458	0.416	0.599
Skills		Yes		Yes
Tasks			Yes	Yes
Municipalities	222	222	222	222
Individuals	65,545	65,545	65,545	65,545

Notes: This table presents results from a decomposition exercise where skill-outcomes are added to the right side of the main estimating equation. The first row of both panels reports the main coefficient when these variables are included, and the second row reports how the magnitude of this coefficient compares to the coefficient from the main estimate.

Table 20: Treatment effects on skills by first-born status

	Visual-spatial	Academic	Social competence
	PanelA: Pre-period difference compared to other siblings		
Oldest child	0.144***	0.217***	0.153***
	(0.006)	(0.006)	(0.005)
Sibship size	Yes	Yes	Yes
Municipality and cohort FE	Yes	Yes	Yes
Observations	144,784	144,784	144,784
	Panel B: Childcare treatment effect heterogeneity		
DiD	0.024	0.027	0.013
	(0.020)	(0.020)	(0.021)
DiD X Oldest child	-0.025	-0.024	-0.040*
	(0.018)	(0.018)	(0.019)
Sibship size	Yes	Yes	Yes
Sibling rank	Yes	Yes	Yes
Municipality and cohort FE	Yes	Yes	Yes
Observations	76,020	76,020	76,020

Notes: Panel A replicates results from Black et al. (2018), suggesting first-born men have higher skills across each of the three dimensions we focus on. In panel B, we study how these skills are shaped by access to public childcare. *= p<0.05, **=p<0.01,***<p<0.001.

Table 21: Treatment effects by predicted heterogeneity percentile

	Treat X predicted het. percentile (1)	Effect for lowest fifth (2)	Effect at the median (3)	Effect for highest fifth (4)
	Panel A: Effects on education, marriage, and the labor market			
Dropout	-0.041***	0.000	-0.016	-0.032**
	(0.012)	(0.013)	(0.010)	(0.010)
HS graduate	0.071***	-0.022	0.006	0.034**
	(0.015)	(0.012)	(0.009)	(0.011)
Tertiary education	0.075***	-0.039***	-0.009	0.021
	(0.013)	(0.010)	(0.009)	(0.011)
Years of education	0.421***	-0.162*	0.007	0.175**
	(0.070)	(0.067)	(0.059)	(0.064)
Income rank	0.077***	-0.028**	0.002	0.033***
	(0.008)	(0.009)	(0.008)	(0.009)
Years employed in 30's	0.383***	-0.141	0.012	0.165*
	(0.075)	(0.075)	(0.073)	(0.083)
Ever married	0.033**	-0.014	-0.001	0.012
	(0.012)	(0.010)	(0.009)	(0.010)
Military service	0.027*	-0.000	0.011	0.022
	(0.011)	(0.033)	(0.033)	(0.034)
Municipalities	229			
Individuals	90,434			
		Panel B: Effects on skills		
Visual-spatial	0.082**	-0.022	0.011	0.044
	(0.030)	(0.020)	(0.018)	(0.023)
Academic	0.138***	-0.040	0.015	0.070**
	(0.035)	(0.023)	(0.019)	(0.024)
Social competence	0.171***	-0.076***	-0.007	0.061**
	(0.030)	(0.021)	(0.019)	(0.023)
Municipalities	223			
Individuals	75,996			

Notes: Column 1 of this table reports the coefficient β_2 from Equation 3. This coefficient measures the difference in effect of public childcare access between a child at the very bottom of the predicted treatment effect heterogeneity ranking compared to a child at the very top of the predicted treatment effect heterogeneity ranking. Column 2(4) evaluates this expected treatment effect for the fifth of children expected to be affected most negatively(positively) by public childcare access. Column 3 evaluates the treatment effect for families at the middle of the predicted treatment effect ranking. *= p<0.05, **=p<0.01,***<p<0.001.

Table 22: Treatment effects by predicted heterogeneity percentile (raw measures of skills)

	Treat X predicted het. percentile (1)	Effect for lowest fifth (2)	Effect at the median (3)	Effect for highest fifth (4)
Arithmetic	0.112**	-0.033	0.012	0.057*
	(0.037)	(0.022)	(0.018)	(0.025)
Verbal	0.138***	-0.039	0.016	0.071**
	(0.032)	(0.022)	(0.018)	(0.022)
Visual-spatial	0.082**	-0.022	0.011	0.044
	(0.030)	(0.020)	(0.018)	(0.023)
Achievement striving	0.145***	-0.065**	-0.006	0.052*
	(0.026)	(0.020)	(0.018)	(0.021)
Activity energy	0.094***	-0.039	-0.001	0.037
	(0.028)	(0.022)	(0.018)	(0.021)
Deliberateness	0.081**	-0.086***	-0.054***	-0.021
	(0.031)	(0.019)	(0.015)	(0.020)
Dutifulness	0.072**	-0.045*	-0.016	0.012
	(0.027)	(0.020)	(0.016)	(0.019)
Leadership motivation	0.134***	-0.054*	-0.000	0.053*
	(0.029)	(0.021)	(0.019)	(0.023)
Self-confidence	0.155***	-0.075***	-0.013	0.049*
	(0.032)	(0.022)	(0.017)	(0.021)
Sociability	0.120***	-0.046*	0.002	0.050*
	(0.034)	(0.021)	(0.018)	(0.023)
Municipalities	223			
Individuals	75,996			

Notes: Column 1 of this table reports the coefficient β_2 from Equation 3. This coefficient measures the difference in effect of public childcare access between a child at the very bottom of the predicted treatment effect heterogeneity ranking compared to a child at the very top of the predicted treatment effect heterogeneity ranking. Column 2(4) evaluates this expected treatment effect for the fifth of children expected to be affected most negatively(positively) by public childcare access. Column 3 evaluates the treatment effect for families at the middle of the predicted treatment effect ranking. *= p<0.05, **=p<0.01,***<p<0.001.

Table 23: Background characteristics of predicted heterogeneity quintiles

	Negatively affected quintile (1)	Positively affected quintile (2)	Pos. - Neg. affected (3)
Moether's education	9.53	10.93	1.40***
	(0.03)	(0.03)	
Father's education	10.08	10.31	0.24***
	(0.02)	(0.02)	
Mother's age at first birth	21.69	26.02	4.33***
	(0.07)	(0.07)	
Family size	2.24	1.76	-0.47***
	(0.02)	(0.02)	
Grandparents present	0.62	0.60	-0.02**
	(0.00)	(0.00)	
Family income percentile	47.37	35.88	-11.49***
	(0.37)	(0.39)	
Lowest income decile	0.03	0.14	0.11***
	(0.00)	(0.00)	
Highest income decile	0.05	0.11	0.06***
	(0.00)	(0.00)	
Older mother + Highest inc.	0.00	0.06	0.06***
	(0.00)	(0.00)	
M.H.Edu + Highest inc.	0.01	0.04	0.04***
	(0.00)	(0.00)	
Correlation of index w/ family income		0.15	

Notes: This table presents the mean background characteristics and standard errors of the twenty percent of families predicted to experience the most negative (Column 1) and most positive (Column 2) effects of access to public childcare. The quintiles in this table correspond to the two leftmost and rightmost deciles in Figure 20. The differences between these groups are plotted in Column 3. *= p<0.05, **=p<0.01,***<p<0.001.

3 Data details

Our measures of skills come from the The Finnish Defence Forces. These data include measures of cognitive skills (arithmetic, verbal skills, and visual-spatial skills) as well as socio-emotional skills (achievement striving, activity energy, deliberation, dutifulness, leadership motivation, self-confidence, and sociability) measured upon conscription at age 19 through a battery of tests and surveys. Researchers are only able to able to access the raw composite scores, not the items or item level data. The Finnish Defence Forces report that the Cronbach alphas for the set of cognitive skills ranges between 0.76 and 0.88 and socio-emotional skills ranges between 0.6 and 0.9, but do not allow researchers to see which skills are measured with which reliabilities. The following descriptions of the different dimensions measured can be found in Nyman et al. (2007) and Jokela et al. (2017).

Arithmetic reasoning. Arithmetic reasoning is measured through numeric pattern completion, solving verbal problems, simple arithmetic operations, and choosing relationships between pairs of numbers.

Verbal reasoning. This test measures verbal abilities, focusing on the definitions of words, as well as relationships between words.

Visual-spatial skills. This test measures pattern recognition and matrix completion in a manner similar to Raven's Progressive Matrices.

Achievement striving. 24 items measuring the extent that an individual wants to perform well and achieve socially valued life goals. This measure includes questions aimed at revealing the extent a person is ready to make sacrifices to achieve success.

Activity energy. 28 items measuring the way that individuals approach their day to day activities, including how fast or vigorously someone gets things done, as well as their preferences for fast-paced work.

Deliberation. 26 items measuring the extent that some plans ahead rather than acts in the moment, related to for example, a person's ability to save money rather than spend it right away.

Dutifulness. 18 items measuring the degree that someone follows social norms, for example if they would return incorrectly given change at the store.

Leadership motivation. 30 items measuring people's preferences for taking charge in group situations and abilities to influence others.

Self-confidence. 32 items measuring a person's self-esteem and beliefs regarding their own abilities. Two examples of underlying concepts are whether a person feels as if they are as good and able as others, and whether the person can meet other people's expectations

Sociability. 27 items measuring a persons's gregariousness and preference for socialization. These include measures such as a person's preference for hosting parties and not withdrawing from social events.

4 Conceptual framework

4.1 Social competence as an organizing concept

In this paper we are interested in how public childcare might shape the social and emotional skills of children aged between three and six years old. The literature in child development and psychology provide an important base from which to approach the potential effects such an intervention may have. We outline the relevant literature from child development here, and show how these concepts may be incorporated into an economics framework in the context of this paper.

The importance of early childhood is documented in prior research across a wide range of disciplines including economics, pscyhology and child development, as well as sociology (Duncan et al., 1994, 2010; Currie and Almond, 2011; Black et al., 2017). In a recent overview of the science of child development, Black et al. (2017) suggest that childhood is a period consisting of ordered stages in which perceptual, motor, cognitive, language, socio-emotional, self-regulation, and cultural skills develop through a rich series of interactions. They explain that several factors can affect the development of these skills, including play, socialization, responsive caregiving and early learning.

For children three to six, the literature on child development has long emphasized that how children are socialized shapes their behavior in later years (Erikson, 1950; Piaget, 1954; Baumrind, 1967). The treatment we study is exposure to public childcare between the ages of three and six. While we might expect public childcare itself to be relatively constant in our context, the family environment or other type of informal care which public childcare subsitutes for may vary drastically. And, understanding the role of public childcare involves understanding how it may potentially substitute for this informal childcare option, often in the family (Clarke-Stewart et al., 1994; Busch-Rossnagel and Knauf-Jensen, 1995; Maccoby and Lewis, 2003; Csibra and Gergely, 2009). Moreover, as has been long understood, the family presents not only the likely counterfactual for public childcare, but also the first place where young children are socialized (Clausen, 1966). As such, the actual treatment we study is likely to vary at the family level and be defined by the difference in early childhood environments between informal or home care and public childcare.

Waters and Sroufe (1983) argue that social competence – the ability to recruit personal and interpersonal resources in the context of goal achievement – is the central organizing construct of early childhood. Since then, social competence has played an important organizing role in early childhood research (Campbell et al., 2000; Denham et al., 2003; Vaughn et al., 2009). Vaughn et al. (2009) describe that social competence consists of three parts: i) behavioral and cognitive skills for successful goal achievement with social contexts; ii) the ability to discover the goals of interactive peers; iii) the understanding of a child's relative value as a preferred playmate. For example, focusing on parents, Pomerantz et al. (2005) highlight the role of parental socialization as a determinant of how children approach achievment, and Gunderson et al. (2013) describe one nice example of how such skills might develop, focusing on how parental praise can lead to persistent improvements the self-confidence and particularly motivation of young children still several years after treatment. Phillips et al. (1987) emphasize verbal interactions between caregivers and children more broadly in childcare settings.[22]

[22] Another potential mechanism behind the development of social skills in childcare is simply the informal interactions

In turn, social competence – through motivation in social contexts – may shape other learning outcomes (Dweck, 1986). Of course, in addition to shaping a child's social competence, early childhood socialization may directly affect other areas of learning such as verbal skills (Hart and Risley, 1995).

4.2 Life-cycle skill development, a framework

As has been noted in the prior literature in economics, the way early experiences may affect later outcomes is not necessarily obvious. We formalize key points using a multi-period model of childhood investment as laid out in the prior literature (Becker and Tomes 1986; Heckman 2006; Cunha and Heckman 2010; Heckman et al. 2013). People's skills (θ) across various dimensions (k) develop over multiple periods of childhood and adolescence ($t \in 1, 2, ..., T$)—shaping various adult (A) outcomes. Skill development in one period is a function of household investments (H)[23], public investments (D), and skills in the prior period such that,

$$\theta_{k,t+1} = f_{t,k}(H_{k,t}, D_{k,t}, \theta_{k,t}).$$

Self productivity. Higher levels of skills in one period may allow for more efficient learning of the same skill in later periods, suggesting that the possibility for effects of childhood investments measured at later stages to be larger than those measured initially.

$$\frac{\partial f_{k,t}(H_{k,t}, D_{k,t}, \theta_{k,t})}{\partial \theta_{k,t}} > 0$$

Dynamic complementarity. Individuals with greater early childhood skills in one domain may be more efficient in learning other types of skills later (say in elementary school). This idea highlights the potential for initial effects in one area to result in later effects in others, and stresses the highly interactive nature of skill investments across periods. This is referred to dynamic complementarity between investments in one skill (k) and the development of other skills (l) in later periods:

$$\frac{\partial^2 f_{k,t}(H_{k,t}, D_{k,t}, \theta_{k,t})}{\partial D_{k,t} \partial \theta_{l,t}} > 0$$

Endogenous investments and substitution. Additionally, we might imagine that household and

between children themselves. The role of peer interactions in formal and informal contexts in early childhood and elementary school has been a large area of research (see, for example, Ladd 1990; Coolahan et al. 2000; Lenard and Silliman 2021).

[23]While children themselves may be unlikely to make consequential investment decisions in early childhood, we consider the household to include the child themselves—whose own investments become more consequential in later years.

public investments are endogenously determined.[24] Accordingly, households may react to public investments in childcare by changing their own investment behavior - potentially substituting away from other forms of childcare:

$$H_{k,t+1} = f(H_{k,t}, D_{k,t}, \theta_{k,t})$$

$$D_{k,t+1} = g(H_{k,t}, D_{k,t}, \theta_{k,t})$$

Skills, education, and the labor market. Lastly, educational attainment is a function of skills as well as household and public investments. Following seminal models in education and labor economics, we consider labor market performance to be a function of education (Becker, 1962; Mincer, 1974) potentially in addition to the direct effect of skills on labor market outcomes (Deming, 2017; Papageorge et al., 2019; Izadi and Tuhkuri, 2021).

$$E_{k,t+1} = f(\theta_{k,t}, H_t, D_t)$$

$$Y_{k,t+1} = f(E_t, \theta_{k,t})$$

Empirical implications. Thus, an empirical implication of the above model is that if changes in some skill $\theta_{k,t}$ are part of the reason we see effects on a long-term outcome $Y_{k,t+1}$, it should be the case that the people who experience effects on the long-term outcome also experience effects in that particular skill:

$$corr[(\frac{\partial \theta_k}{\partial D})_i, (\frac{\partial Y}{\partial D})_i] \neq 0$$

An important note, here, is that this correlation can be different from zero even if some particular skill (θ_{kt}) does not causally drive the the effects on Y. It could, for example, be that some skill adjacent to k is driving the effect, and we simply happen to obsere the effect on k. Likewise, since it is not necessarily some particular skill (k) that shifts Y, it is possible for effects on multiple skills (k, l, m) to all be correlated with the effect on Y such that the sum of these correlations is greater than one.

In the context we study – where public investment in early childhood (D) changes – this framework suggests the following points: i) public investments in early childhood can shift skill development in specific domains, and affect the level of these skills at different points in time; ii) skills acquired in one domain (say social competence) can shape the productivity of later investments in other domains (say verbal skills); iii) changes in public investments in early childhood may affect household

[24] If households are more nimble to respond to public provision than the government is in responding to household provision, the function (f) may include an additional term for same-period public investment ($D_{k,t+1}$).

investments in skills; iv) skills and education may have distinct effects on labor market outcomes; v) since childcare investments are endogeneously determined (by both municipalities and households), the relationship between household or public investments in childcare and later outcomes is not identified by a cross-sectional comparison of households accessing public childcare with those that do not.

5 Methods

5.1 Estimating correlations between skill and labor market treatment effects

We are interested in understanding how the effects of public childcare on skills and long-term outcomes are related to each other. More concretely, we ask: Do children who benefit from access to public childcare in terms of particular skills also experience improvements in longer-run outcomes such as income? Answering this question probes a basic assumption underlying causal mediation.

In an ideal world, we might look at whether the specific individuals who experience an improvement in skills as a result of access to public childcare are the same people that earn more as a result of public childcare. Recapitulating the points from Section 4, one estimator for this relationship is the correlation:

$$\theta = corr(TE_i^M, TE_i^Y)$$

This framing takes a similar form as the study of the covariance between treatment effects on test scores and treatment effects on wage earnings in the context of teacher value added (Chetty et al., 2011).[25] Unfortunately, as opposed to the case of teacher value-added, in our context – as in many others – estimating individual level treatment effects rests on untenable assumptions. One way to overcome this challenge is to estimate a large number of subgroup treatment effects for individuals who are likely to respond to treatment in different ways, and correlate these. For example, Angrist et al. (2022) split up their sample in a number of ways based on a number of background characteristics such as race, gender, and prior academic performance.

We extend this work by using machine learning to group individuals by their predicted response to public childcare access, discussing potential biases in the estimation of treatment effect covariances, and outlining ways to test the robustness of the estimates.

For each group (g), we estimate treatment effects (β) on both measures of adult skills (M) as well as longer-run outcomes (Y).

[25] See also, for example, Jackson (2018) who studies the relationship between teacher value added based on academic outcomes compared to teacher value added based on behavioral outcomes, and their relationship to longer-term outcomes.

$$\hat{\theta} = corr(\hat{\beta}_g^M, \hat{\beta}_g^Y)$$

Assuming we can estimate the β's precisely and without bias, $\hat{\theta}$ is likely to be biased upwards as long as the number of groups is small. This may be because the extent that the same *general* group of people experiences effects on M also experience effects on Y may conceal differences in effects for more granular groups. Additionally, in small samples correlations are estimated with bias, since the sample covariance is divided by $n-1$ rather than just n. For both these reasons, as the number of groups grows, $\hat{\theta}$ should approach θ. We estimate $\hat{\theta}$ for ten to one-hundred splits, and show that $\hat{\theta}$ does indeed decrease as the number of splits increases, but plateaus after about fifty splits $\hat{\theta}$.

$$\theta = lim_{g\to\infty}\hat{\theta}$$

In the presence of any imprecision in our estimates, even if such estimation error is i.i.d., we face another challenge. So long as any error in the estimate of the TE on the mediator ($\tilde{\beta}_g^M$) is correlated with error in the estimate of the TE of the long-term outcome ($\hat{\beta}_g^Y$) for the same group (g), the two estimates will be mechanically correlated, and this estimator is likely to be upward biased. To avoid such mechanical correlation we use a split-sample approach, where we estimate $\hat{\beta}^M$ in using data from one half of each group (g^{ss1}) and estimate $\hat{\beta}^Y$ with data from the other half of each group (g^{ss2}).

$$\hat{\theta}^{ss} = corr(\hat{\beta}_{g^{ss1}}^M, \hat{\beta}_{g^{ss2}}^Y)$$

Additionally, we can run a OLS regression through $\hat{\beta}_{g^{ss1}}^M$ and $\hat{\beta}_{g^{ss2}}^Y$. The OLS estimate provides a simple and intuitive way to gauge the statistical significance of the correlation.

Since the splitting of each group in two will induce randomness, we estimate this split sample estimator $\hat{\theta}^{ss}$ fifty-one times and take the median of these estimates across splits ($\hat{\theta}^{\bar{ss}}$) to improve the reliability of these estimates. We consider this our main estimate of the raw correlation in treatment effects. Likewise, we consider the median p-value from the corresponding OLS regression an estimate of the statistical significance of this relationship.

At the same time, however, as the number of groups increases, however, it is likely that the β's will be estimated increasingly imprecisely, thereby inducing our estimate $\hat{\theta}$ to be biased downwards. To avoid some of this downward bias, we use empirical Bayes to produce adjusted estimates of the relationship between medium term and long-term treatment effects. We explain this below. For simplicity, the additional notation for split-sample subgroups (g^{ss1} and g^{ss2}) are removed from the below explanation.

For each outcome our split-sample subgroup estimates of treatment effects are estimated with error.

$$\hat{\beta} = \beta + e$$

This error carries over to the variances and standard deviations of these estimates.

$$var(\hat{\beta}) = var(\beta) + var(e^{\beta})$$

When we divide the covariance by the estimated standard deviations instead of the true standard deviations, the correlation becomes attenuated.

$$corr(\beta^M, \beta^Y) = \frac{cov(\beta^M, \beta^Y)}{sd(\beta^M)sd(\beta^Y)} = \frac{cov(\hat{\beta}^M, \hat{\beta}^Y)}{sd(\beta^M)sd(\beta^Y)}$$

The extent of the attenuation in our estimates of the correlation is determined by the reliability of the denominator, where the reliability is the portion of the true variance in relation to the extent of total variance:

$$\lambda^{MY} = \frac{sd(\beta^M)sd(\beta^Y)}{sd(\hat{\beta}^M)sd(\hat{\beta}^Y)}$$

We can attempt to correct our raw estimates of the correlation using empirical Bayes by scaling the denominator by its reliability (see Chetty et al. (2014)) :

$$\frac{cov(\hat{\beta}^M, \hat{\beta}^Y)}{[sd(\hat{\beta}^M)sd(\hat{\beta}^Y)] * [\frac{sd(\beta^M)sd(\beta^Y)}{sd(\hat{\beta}^M)sd(\hat{\beta}^Y)}]} = corr(\hat{\beta}^M, \hat{\beta}^Y) * [\frac{sd(\beta^M)sd(\beta^Y)}{sd(\hat{\beta}^M)sd(\hat{\beta}^Y)}]^{-1}$$

$$corr(\beta^M, \beta^Y) = corr(\hat{\beta}^M, \hat{\beta}^Y) * \lambda^{MY}$$

We interpret these scaled correlations as our main estimates, but report the unscaled versions as well. To assess the statistical significance of the correlations we simply take the p-values from the corresponding OLS regressions.

5.2 Predicting treatment effect heterogeneity

We can imagine that potential outcomes for individuals are Y_i^1 if they have access to childcare and Y_i^0 when they do not. Further, each individual (i) can be characterized by a vector of covariates, Z. We can imagine that the baseline potential outcome in the untreated state $b(Z)$ is defined as $E[Y^0|Z]$, and that treatment effects conditional on the vector of coefficients are defined as follows:

$$s(Z) := E[Y^1|Z] - E[Y^0|Z] \tag{6}$$

As described in the above section, we want to form granular subgroups so that we can estimate correlations in treatment effects across outcomes. Moreover, we want to do this in a way that both maximizes variation in treatment effects across subgroups while minimizing our degrees of freedom as researchers. With this goal in mind, we will use observable characteristics (Z) from our data to provide a measure of predicted treatment effect heterogeneity that can be used to split our data up into an arbitrarily large number of subgroups.

To predict variation in outcomes using information on observables we use a machine learning framework based on Chernozhukov et al. (2021). These authors acknowledge the near impossibility of consistently estimating the conditional average treatment effect (CATE) given a particular set of background variables. Instead, they accept the fact that different splits of training (auxiliary) and test (main) samples may produce different estimates of treatment effect heterogeneity – each time identifying different sets of variables predictive of heterogeneous treatment effects. This inconsistency in the particular set of characteristics predicting heterogeneity is likely to be all the more accentuated when variables are correlated with one another. Setting aside the goal identifying a single CATE, they suggest taking advantage of repeated data-splitting to avoid overfitting and provide valid estimates of feature of treatment effect heterogeneity.

The general approach to estimating treatment effect heterogeneity we implement (Chernozhukov et al., 2021) works by randomizing units into main and auxiliary samples hundreds of times, applying machine learning to the auxiliary samples, generating predictions in the main samples, and then picking median main sample parameter estimates across the splits. Given that adult income rank is the main long-term outcome that we study, we run our machine learning exercise for adult income rank. We operationalize this approach as follows.

Before implementing our machine learning procedure, we divide all of our background variables into categorical measures. Then, we interact all these measures with each other, generating several hundred variables. For investigating treatment effect heterogeneity, we then interact each of these variables with treatment status ($D_{imc} = FIRST_m \times POST_c$).

We begin by randomizing municipalities into auxiliary and main samples with equal probability 500 times (N).[26] We maximize power by ensuring that half the treated and municipality units end up in the main and auxiliary splits at each randomization (Chernozhukov et al., 2021). For each split, we then use elastic net regressions in the auxiliary (A) sample to generate predictions of adult income rank given a array of covariates (Z) for both treated units and untreated units ($D_{imc} = FIRST_m \times POST_c$).

[26]Wager and Athey (2018) suggest randomizing units into auxiliary and main samples at the level of treatment assignment, in our case the municipality. This parallels contemporary understandings of how standard errors should be clustered when estimating treatment effects (Bertrand et al., 2004).

$$\forall i \in N, A \begin{cases} \hat{Y}_i^1 | Z, D = 1 \\ \hat{Y}_i^0 | Z, D = 0 \end{cases}$$

Then, we take these predictions to the main sample (M), where we create a measure of predicted treatment effect heterogeneity ($\hat{S}_i$) for each individual, given the machine learning estimates from the auxiliary sample of that split.

$$\hat{S}_i = \hat{Y}_i^1 - \hat{Y}_i^0, \forall i \in N, M$$

Various approaches to using machine learning for the identification of heterogeneous treatment effects have been proposed in the literature (see, for example, (Imai et al., 2013; Zhao et al., 2018; Wager and Athey, 2018; Chernozhukov et al., 2021)). The key goal of these methods is to provide a way of formally selecting amongst a large number of covariates to estimate both baseline values and treatment effects using some form of regularization (see, for example, Tibshirani, 1996 or Athey and Imbens, 2019). Given our difference-and-difference setup, we first ensuring balance using uniformly valid post-double selection method (Belloni et al., 2014, 2017), and then use an elastic net approach following Zhao et al. (2018).

To simplify our empirical approach for the context of the elastic net, we residualize outcomes using untreated units to remove municipality and year effects from adult income rank (see, for example, Gardner, 2021). This allows us to operationalize our machine learning estimation of treatment effect heterogeneity as if we were working with random assignment, using just the treatment indicator and the residualized outcomes (Y_i^*).

We use an elastic net to generate our predictions (see, for example Chernozhukov et al. (2021), who find the elastic net to perform well relative to other machine learning procedures), using a uniformly valid post-double selection method (Belloni et al., 2014, 2017). In the first step of this approach, we select a set of control variables that are useful for predicting treatment (D_i) – the set of potentially important confounding factors. In the second step of this approach, we select the set of characteristics useful for predicting the outcome (Y_i). Third, we estimate heterogeneity in treatment effects by forcing the inclusion of both these sets of variables, and allowing the elastic net to select additional additional variables from the set of variables interacted with treatment status (see Zhao et al., 2018). This process helps to induce balance across all observable background characteristics that might be relevant in explaining either treatment assignment or outcomes. We use this information to generate predictions of $\hat{Y}_i^1$ and $\hat{Y}_i^0$ that we carry over to the main sample.

Following Chernozhukov et al. (2021), we then post-process these predictions of treatment effect heterogeneity to reduce sampling noise using a linear regression:

$$Y_i = \alpha'\mathbf{X}_{1i} + \beta_1(D_i - p(Z_i)) + \beta_2(D_i - p(Z_i))(\hat{S}_i - \bar{S}_{N,M}) + \epsilon_i, \forall i \in N, M,$$

$$E_{N,M}[w(Z_i)\hat{\epsilon}_i\mathbf{X}_i] = 0$$

The coefficient β_2 measures the extent to which $S(Z)$ predicts treatment effect heterogeneity. The vector $\mathbf{X}_i$ includes municipality and cohort fixed effects, a phase-in dummy, and the predicted outcome of each main sample individual in the absence of treatment ($\hat{Y}_i^0$). The term $p(Z_i)$ is the propensity score, bounded between zero and one, that an individual with characteristics Z is treated and $w(Z_i)$ are weights defined by $(p(Z)(1-pZ))^{-1}$.

The coefficient β_2 measures the extent to which $S(Z)$ predicts treatment effect heterogeneity. When $\hat{S}_i$ provides a strong signal of TE heterogeneity in the main sample, this has little effect. Under the unlikely scenario that $S(Z)$ provides a perfect proxy for treatment effect heterogeneity, $\beta_2 = 1$. In the case that β_2 contains no information on heterogeneity, $\beta_2 = 0$. Rejecting the null hypothesis that across all splits the median $\beta_2 = 0$ suggests both that there is heterogeneity in treatment effects, and that $S(Z)$ provides relevant information by which to predict TE heterogeneity. However, as noted by Chernozhukov et al. (2021), failing to reject the null hypothesis ($\beta_2 = 0$) does not necessarily mean that there is no TE heterogeneity; given the demands of this approach on statistical power, this is often the case. In our case, the median value of β_2 is 0.26 – lower than in the author's example application, suggesting that our machine learning exercise produced a noisy signal of heterogeneity. Since each split is estimated from just half of the overall sample, we lack statistical power to reject the null hypothesis. We take one of the target parameters the authors outline and show that it can be useful even when there is insufficient statistical power for variational analysis that takes into account split-level uncertainty that arises from randomness in splits.

The target parameter for us will be the "personalized prediction" of $S(Z)$ for each individual:

$$\hat{\theta} = \hat{\beta}_1 + \hat{\beta}_2(\hat{S}_i - \bar{S}_{N,M}), i \in N, M$$

This parameter is estimated in the main sample about 250 times for each individual (the other half of the time the individuals' municipality is randomized into the auxiliary sample). As per Chernozhukov et al. (2021), we take the median across all main splits to provide the best guess of the individual personalized prediction for each individual. Given that we can generate these out of sample predictions not just for half our sample, but the full sample we can estimate heterogeneous treatment effects without losing power.

To facilitate interpretation, we convert these main sample medians of the personal predictions to a continuous measure of rank on a scale of zero to a hundred (Davis and Heller, 2020). Since these ranks are generated from aggregating predictions of features relating to TE heterogeneity across hundreds of splits, and include sampling noise, there is no one index that corresponds to this ranking. In fact, given randomness induced by sample splitting, it is even possible that two individuals born to families with identical observables have slightly different ranks. Thus, this approach is not useful for trying to predict how a new sample might respond to a treatment in heterogeneous ways, but rather to describe the heterogeneity in the responses of the estimation sample.

We can use this index of predictions to estimate heterogeneous treatment effects just as we did with family income rank. First, we test for the identifying assumption using event-study estimates, focusing on the top fifth of and bottom fifth of the predicted heterogeneity index. Appendix Figures 5 and 7 show that prior to treatment there is no evidence of trends in the pre-period. The machine learning exercise we use for our (out-of-sample) predicted treatment effect index is based on adult income rank as an outcome – so perhaps it is not surprising to see evidence of parallel trends in the pre-period for this outcome. Reassuringly, both the other long-term outcomes and skill outcomes also show no evidence of trends in the pre-period.

Parallel to the parametric estimates we report across family income percentile, we report the parametric estimates across predicted treatment effect percentile in Appendix Tables 21 and 7. Column 1 shows the predicted change in outcome as families move from the very bottom of the predicted heterogeneity ranking to the very top. Across all outcomes—ranging from administrative measures of labor market performance, years of education, and measures of adult skills—these estimates suggest there is statistically significant heterogeneity in the effects of public childcare. For most individuals in our sample, however, the magnitude of these effects is relatively small. In Columns 2-4 of Table 21, we estimate the magnitude of treatment effects for families at the bottom 10th percentile, 50th percentile, and 90th percentile of the predicted heterogeneity rank. These estimates suggest that childcare access had both positive and negative effects on children, depending on what kind of family they came from.

As in the analysis of average treatment effects, the validity of these results rests on there being parallel trends in the potential outcomes of individuals in the absence of treatment, but with the stronger requirement that there are parallel trends at each level of predicted heterogeneity. Since the predicted heterogeneity rank is based on observed family background characteristics, balance along these measures no longer provides a test of the validity of the research design. Still, we can plot the pre-policy trends in outcomes for families at different points in the distribution of predicted heterogeneity. Appendix Figures 5 and 7 suggest that the outcomes of families in both the top and bottom twenty percent of predicted heterogeneity evolve in a parallel way prior to treatment.

To test for whether or not such a parametric approach is justified, we can test for the linearity of treatment effects across the distribution. While we lose the statistical power that we gain from pooling heterogeneous effects into one index, we can estimate these seperately for each percentile of the predicted heterogeneity index to provide a non-parametric version of these estimates. These are presented in Appendix Figures 8 and 10 for all outcomes, and, despite considerable noise in individual point estimates, suggest that there is indeed a linear pattern in the effects. We further split these subgroups in two for the granular subgroup estimates we use to estimate the correlations in the treatment effects between medium and longer-term outcomes.

Appendix Figure 20 and Table 23 describe how family background characteristics vary between individuals with different predicted treatment effects.

5.3 Alternative approaches to assessing mediating hypotheses

A comparison of the treatment effect correlations between long-term outcomes and various potential mediating outcomes provided us a basis by which to gauge the plausibility of competing mediating hypotheses. However, while resting on weaker assumptions, this approach does not provide a direct way to gauge either what portion of the effect on long-term outcome is explained by each of the potential mediating channels or how much a unit change in a mediator might affect the long-term outcome in question.

In a common approach to a causal mediation (see, for example, Heckman et al. (2013)) authors ask: How much of the effect of treatment (D) on an outcome (Y) can be explained by the mediator (M)? While this approach to mediation has the potential to provide an estimate of the degree to which particular mediators drive the effects of D on Y, it relies on stronger assumptions.

Adapted to our framework, we operationalize this approach as follows. We recover the $\hat{\beta}$ from our original estimates of effects across the family income distribution (Equation 3). Next, we compare how this estimate changes when we add measure of our mediating factors (visual-spatial skills, academic skills, and social competence) to the equation, both interacted with treatment and not interacted with treatment.

$$Y_{imc} = \beta_1(FIRST_m \times POST_c) + \beta_2(FIRST_m \times POST_c \times HET_i) + \tag{7}$$

$$\sum^{M} \beta_M(FIRST_m \times POST_c \times HET_i) + \sum^{M} \lambda_M HET_i + \sum^{M} \phi_M$$

$$\delta(FIRST_m \times PHASEIN_c) + \pi_m + \gamma_c + e_i$$

Imposing still more structure on this relationship, this exercise can be used to gauge the extent that we might expect Y to shift as a result of shifts in M (Imai et al., 2010).

In addition to standard causal assumptions, Imai et al. (2010) term the additional assumption required for causal mediation as *sequential ignorability*, consisting of a first part that requires that the potential outcomes of mediators to be independent of treatment, and a second part that requires the relationship between the causal mediator (M) and the outcome (Y) to be uncorrelated with any unobserved covariates. In most applications, however, as in our own, this assumption is unlikely to hold, since – as described in Section 4 – effects on, say social skills, are likely to be correlated with effects on, say cognitive skills.