JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12400

ARE IMPACTS OF EARLY INTERVENTIONS IN THE SCANDINAVIAN WELFARE STATE CONSISTENT WITH A HECKMAN CURVE? A META-ANALYSIS

Michael Rosholm* and Alexander Paul*

Department of Economics and Business Economics
Aarhus University
Aarhus, 8210 Denmark, TrygFonden's Centre for Child Research
Aarhus University
Aarhus, 8210 Denmark and Centre for Integrated Register-based Research, CIRRAU
Aarhus University
Aarhus Denmark

Dorthe Bleses and Anders Højen

School of Communication and Culture
Aarhus University
Aarhus, 8000 Denmark and TrygFonden's Centre for Child Research
Aarhus University
Aarhus, 8210 Denmark

Philip S. Dale

Department of Speech & Hearing Sciences
The University of New Mexico
Albuquerque NM, 87131 United States

Peter Jensen

TrygFonden's Centre for Child Research
Aarhus University
Aarhus, 8210 Denmark and Department of Economics and Business Economics

*Corresponding author contact email: rom@econ.au.dk and alexander.paul@econ.au.dk.

Journal of Economic Surveys (2021) Vol. 35, No. 1, pp. 106–140

Aarhus Universitet
Aarhus, 8210 Denmark

Laura M. Justice

Ohio State University College of Education and Human Ecology
The Ohio State University
Columbus OH, 43201 United States

Michael Svarer

Department of Economics and Business Economics
Aarhus University
Aarhus, 8210 Denmark

Simon Calmar Andersen

TrygFonden's Centre for Child Research
Aarhus University
Aarhus, 8210 Denmark and Department of Political Science
Aarhus University
Aarhus, 8000 Denmark

Abstract. "Early intervention" has been a mantra in recent debates about human capital investment. Strong theoretical models motivate this focus by predicting that investment in children is most cost-effective when they are young. The "Heckman curve" summarizes this idea visually (Heckman, 2006). However, hardly any reviews scrutinize this hypothesis empirically in modern welfare states such as those in Scandinavia that already invest heavily during early childhood. Any such review is ideally based on interventions conducted as randomized controlled trials (RCTs), set in the same welfare state, and comparable across ages through cost-standardized effects. This meta-analysis assembles cost-standardized effect estimates from 10 RCTs, including a total of 18 intervention arms and 30,578 participants (aged 1.5–24 years), conducted by the same research center in the Scandinavian welfare state of Denmark. These interventions show significant effects relative to their costs, despite the large baseline investment level. Interventions targeted at younger children tend to produce larger effects, consistent with the Heckman curve. However, variation in the effect size within age groups is as large as it is across age groups. This indicates that both the quality and timing of investments matter and that "early interventions" are not necessarily superior to later interventions.

Keywords. Early interventions; Education; Meta-analysis; Randomized controlled trial; Welfare state

1. Introduction

A large body of research shows that interventions targeted at children and youth with the aim of promoting equal opportunities can be a very good investment from society's point of view (see, e.g., Karoly, 2012; Heckman and Mosso, 2014; Elango *et al.*, 2016). Early investments are often perceived to benefit both growth and equality, two aims that are often seen as being in conflict.

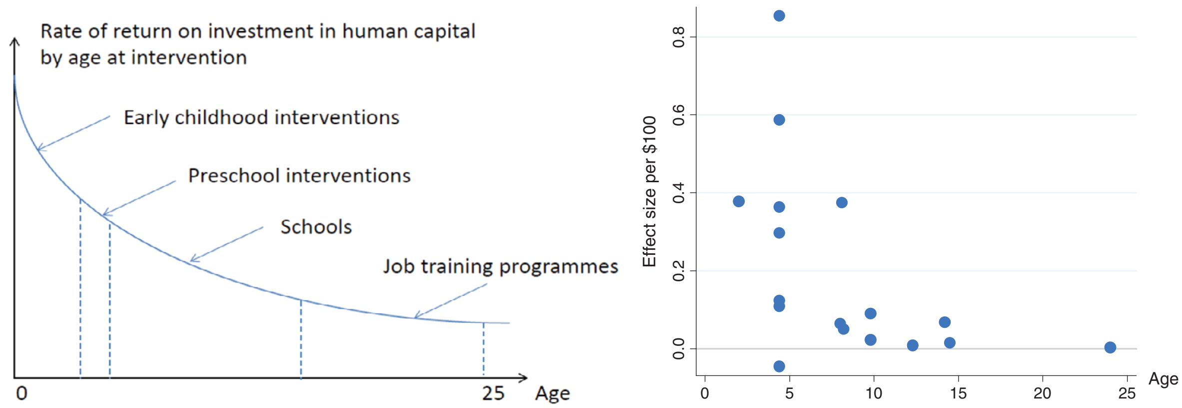

Figure 1. *Left*: Stylized Heckman Curve Depicting the Returns to Extra Marginal Investments in Individuals at Different Ages, Starting from a Situation of Low But Equal Levels of Initial Investment at All Ages. *Right*: Standardized Effect Sizes per $100 from 18 Treatment Arms of 10 Danish Interventions. [Colour figure can be viewed at wileyonlinelibrary.com]

This view is based on pioneering work by Heckman and coauthors, who propose the theoretically derived "Heckman curve" (Figure 1, left). The curve depicts the simulated returns on marginal investments in individuals at different ages (Heckman, 2006,2008). It hypothesizes that early childhood investments yield higher economic returns than later remediation investments and thus presents a strong theoretical case for investing in early childhood skill formation. Investing in disadvantaged children early in life promotes both efficiency and equality.

In recent years, some interest has centered on the Scandinavian countries, which have managed to become some of the world's richest countries (in terms of GDP per capita, for example), and yet have maintained a comprehensive welfare system and have the highest degree of income equality in the world. This interest was boosted in the 2016 U.S. presidential campaign, when the Democratic candidate Bernie Sanders pointed specifically to Denmark as a country to learn from.

One of the proposed causal factors behind Denmark's success has been its generous prototypical Scandinavian welfare state that provides universal access to a wide selection of services for all of its citizens. Beginning during pregnancy, comprehensive birth preparation programs are in place to prepare future parents for parenthood. Right after birth, parental leave systems pay parents to spend time one-on-one in the first year of their children's lives, and there is access to heavily subsidized universal day care from age 6 months on, at no cost to low-income families and highly subsidized for all other families. Enrollment rates in day care are high, reaching 91% for 1- to 2-year-olds and 97% for 3- to 5-year-olds (NOSOSKO, 2014). Moreover, the average ratio of children to caregiving staff (teacher's aides and teachers) in Danish day care is 5:1 (3.7:1 for children aged 0–2 and 6.4:1 for children aged 3–5; Dalsgaard and Tenney Jordan, 2016). This is the lowest among all OECD countries and much lower than the OECD average of 11:1 (OECD, 2016). Besides high public spending on day care, there is free access to schooling and education up to and including university (no tuition fees) combined with favorable study grants (students not living with their parents receive approximately $900 per month), free and universal public health care, and there is a large and diverse toolbox of public policies and interventions aimed at helping and supporting disadvantaged families and, in particular, children growing up in such families.

However, a recent study by Landersø and Heckman (2017) demonstrates that, despite Denmark's far more supportive welfare state with its heavily subsidized day care and free educational system, the relationship and transmission mechanisms between parents' and children's education is very similar in Denmark and the United States. Most of this association can be explained in both countries by factors set in place by age 15. Moreover, Denmark is below the United States when it comes to the fraction of 25- to 34-year-olds that have completed an upper secondary education as defined by the International Standard Classification of Education 2011 (slightly above 90% in the United States and around 84% in Denmark; see OECD, 2016). Landersø and Heckman (2017) suggest that higher levels of income mobility and equality in Denmark are mainly attributable to *ex post* redistribution through taxes, transfers, public services, and wage compression.

The significance of these surprising findings is their implication that the expensive public institutions of the Danish welfare state are not only unable to provide equality of opportunities for children growing up in disadvantaged families; they are also unable to do this to a larger extent than the disadvantaged parents manage to do without comprehensive public support in the United States. This suggests a massive policy failure and an associated considerable waste of public resources, which warrants additional research into the reasons underlying it.

This policy failure has led to renewed interest among Danish ministries and municipal authorities in the evaluation of interventions explicitly targeting children and youth, and in how to improve the performance of the institutions of the welfare state in this respect. This interest in turn has led to the emerging use of randomized controlled trials (RCTs) in public policy evaluations, a methodology that is still relatively rare in Europe. In fact, over the period 2013–2019, TrygFonden's Centre for Child Research has been involved in more than 40 RCTs testing interventions aimed at children and youth in Denmark, which is quite unique in the European context. Coupled with the availability of some of the most comprehensive data sets for children and their families, based on administrative registers, this research provides us with unique opportunities for examining the effectiveness of interventions aimed at children and youth.

In this study, we conduct a meta-analysis of standardized short-term impacts from the first 10 completed RCTs—with a total of 18 different treatment arms (i.e., excluding control arms)—of interventions aimed at children and youth, which have been evaluated by researchers at TrygFonden's Centre for Child Research in Denmark. The RCTs were conducted at a large scale and mostly administered to unselected, representative samples of the population. In total 30,578 participants aged 1.5 to 24 years were assigned to the treatment or control groups. We have access to cost measures for all interventions and can thus compute cost-standardized effect sizes to compare studies. We plot these cost-standardized effect sizes against age in the right panel of Figure 1. The graph strongly points to a negative relationship between impact and age and thus motivates our test of consistency with a Heckman curve. As the Heckman curve is a hypothetical construct, such an empirical test naturally comes with a few caveats and assumptions that we clarify carefully.

We find that, even in a Scandinavian welfare state characterized by large investments throughout childhood, additional interventions can have strong effects relative to their costs. Moreover, meta-regressions reveal a tendency for earlier interventions to produce larger cost-standardized effects, consistent with the presence of a Heckman curve. However, variation in the size of the effect within age groups is as large as it is across age groups. Early interventions can be highly effective, but they can also be virtually ineffective.

Our finding that effect sizes vary considerably within age groups has two implications. First, it offers an explanation for the paradoxical situation of large investments and discouraging results on equality of opportunity in the Danish welfare state. It could be that large public investments per se in early childhood interventions are not enough to offer a sufficiently rich learning environment for children; the quality of investment, as reflected by its type and the curriculum of day care and preschool—what caregivers do with the children—may be equally or even more important. Second, it reinforces the idea that the Heckman curve should be interpreted as a technological frontier, as was recently emphasized by its inventor (Heckman, 2020). It captures the returns on investment that may be obtained when institutions and interventions are optimally designed. The marginal effect of new interventions thus depends as much on the quality as on the timing of the intervention. Hence, the focus on ever more "early interventions" present in the recent literature (Allen, 2011; Black *et al*., 2017) may not be entirely warranted.

We contribute to the literature in the following ways; first, to our knowledge, the existence of a Heckman-type curve has not been documented empirically, at least not in a non-U.S. context, and in particular not in the Scandinavian welfare state context. Second, we offer results from a single interdisciplinary research center, using robust evaluation techniques (RCTs) in a very homogeneous cultural context—Denmark. Hence, any systematic variation in intervention effects by, for example, age is likely to arise from the mechanisms of skill formation discussed in the next section and documented empirically in the U.S context by Heckman and coauthors in a number of studies (Cunha and Heckman, 2008; Cunha *et al*., 2010).

This paper proceeds as follows. Section 2 clarifies the theoretical foundation of the Heckman curve, summarizes the existing literature on early childhood education (ECE) programs and describes the Danish welfare state. Section 3 discusses the data, the interventions included in this study, and elaborates on outcomes and cost. Section 4 presents the results of our meta-analysis. Section 5 discusses and interprets the results. Section 6 points out limitations of our study. Section 7 concludes.

2. Background

2.1 *From Theory to Application*

In a series of papers, Heckman and Cunha have formulated a model of the technology of skill formation (e.g., Cunha and Heckman, 2007, 2008; Elango *et al*., 2016). According to this model, children are born with a set of multidimensional abilities, and the subsequent development of these abilities during childhood is determined by parental skills and investments in their children, and by skills and investments acting as substitutes and/or complements during different stages of childhood. The model is summarized in Figure 2 (adapted from Elango *et al*., 2016). In the context of the welfare state, where the public sector invests heavily in children as well, we define "investments," that is, the time and monetary resources spent on children, as including parental investments as well as public interventions and institutions.

The Heckman curve (Heckman, 2006, 2008) visually summarizes the model of skill formation (Figure 1). The curve "shows the return to a marginal increase in investment at different stages of the life cycle starting from a position of low but equal initial investment at all ages" (Heckman, 2008, p. 309). It has a negative slope and thus postulates that marginal returns to investment are higher at younger ages.

Under specific assumptions, the model of skill formation gives rise to the Heckman curve. Support for these assumptions comes from Cunha *et al*. (2010), who estimate the technology parameters. Three drivers of the Heckman curve can be identified: (1) The productivity of investment is higher in early childhood than in late childhood, especially for cognitive outcomes. In other words, a dollar of investment has a larger immediate impact on skills the younger the child. One explanation for this could be critical and sensitive periods for the acquisition of certain skills (e.g., language skills) at young ages. (2) There is complementarity between investment and cognitive skills that becomes stronger as children get older. In other words, investments increase skills more if they come on top of a large stock of existing skills, especially at later stages of childhood. This implies that if levels of investment and thus skills are initially low, as assumed in the Heckman curve, then children receive more benefit from early investment than late investment. (3) There is substantive self-productivity. Skills produced in early childhood persist and become reinforced over time. Self-productivity coupled with complementarity generates so-called dynamic complementarity between early and late investments. That is, productive investments in early childhood generate a higher level of skills that boost the returns of investments in later childhood.

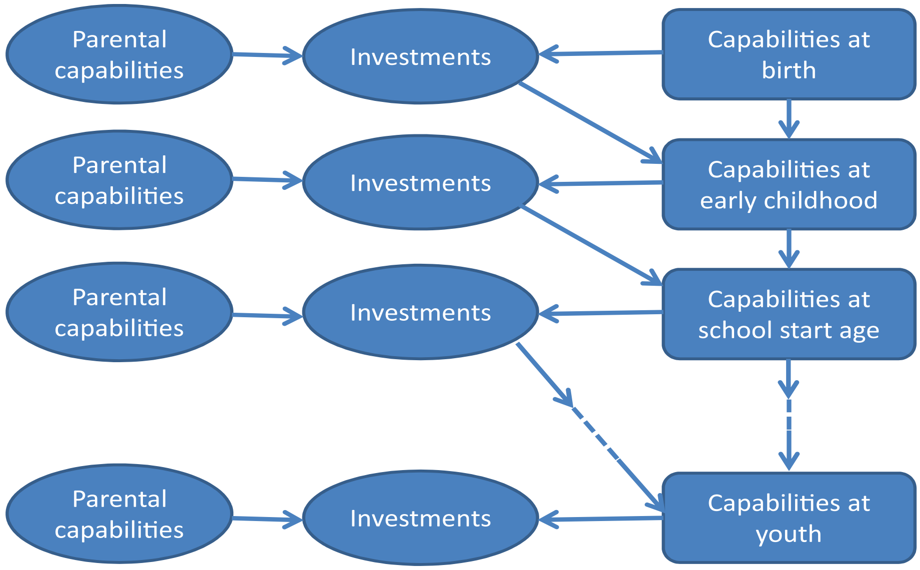

Figure 2. Stylized Model Showing How Parental Capabilities, Parental Investments and Children's Capabilities Interact in the Process of Human Skill Formation. [Colour figure can be viewed at wileyonlinelibrary.com]

Cunha *et al*. (2010) provide indirect evidence of the Heckman curve based on parameter estimates of a structural model. Our study attempts to test the validity of the Heckman curve more directly using the outcomes of randomized experiments that were conducted at various ages during childhood in a common setting. Because the Heckman curve as defined above is a hypothetical construct, such an exercise necessitates a range of practical assumptions and simplifications that we make explicit in the following.

First, the Heckman curve is defined for *marginal* returns to investment. In practice, this would require a researcher to examine 1-dollar interventions, much in contrast to the real-world interventions included in this study, which cost between 40 and 50 and several hundred dollars per participant. To address this dilemma, we calculate benefit–cost ratios: we divide the effect size by the cost of the intervention. For readability, we also scale up the result to obtain effect sizes per 100 dollar. A limitation of benefit–cost ratios is that they measure the average rather than marginal product per dollar. If the marginal product of investment is diminishing, we will underestimate the return of the first dollar, especially for highly expensive interventions. A Heckman curve could emerge simply because interventions for older children—as in our setting—are implemented at higher cost, generating small average products. We argue in the discussion why we do not think that this mechanism drives our findings.

Second, the Heckman curve is to be interpreted as a technological frontier, showing the highest feasible return of an optimally designed intervention at every age (Heckman, 2020). In our setting, effect sizes vary considerably within age group, suggesting that some studies are inefficient and below the frontier. As a result, a simple meta-regression of effect size on age will not estimate the frontier but rather the average return of all interventions including the inefficient ones. One could approach this problem using stochastic frontier analysis, but this would require a large number of studies close to the frontier. Instead, we assume that inefficiencies are unrelated to age. If true, this will allow us to test for a negative slope of the Heckman curve even though we will underestimate its location.

Third, the Heckman curve makes assumptions about counterfactual investment. Specifically, it assumes "low but equal initial investment at all ages." In contemporaneous Denmark, current investments in

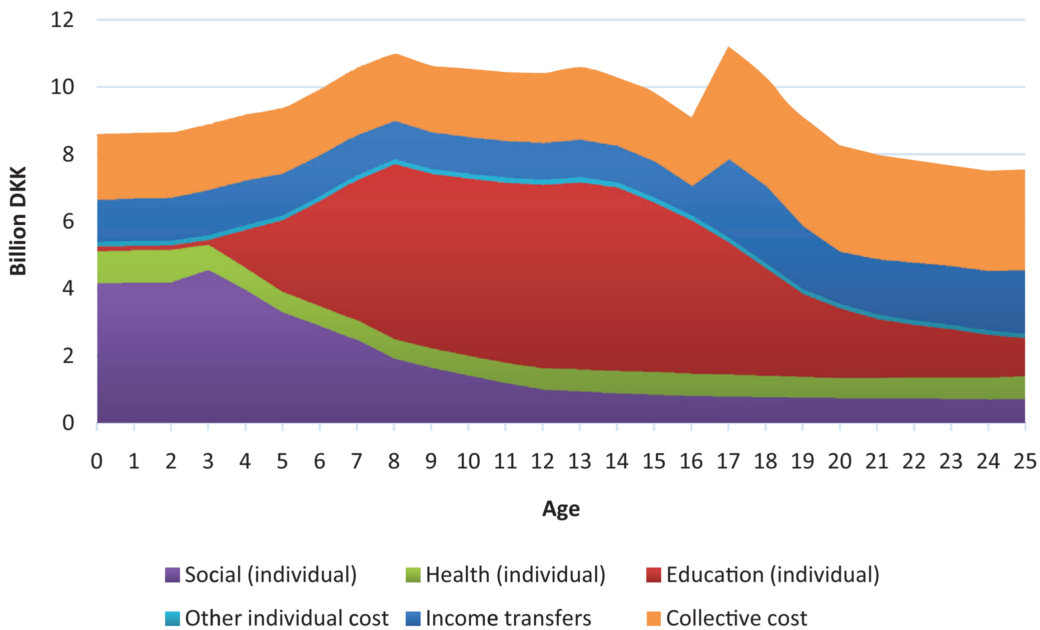

Figure 3. Types of Public Expenditures by Age and Type in 2008, Calculated Based on Data from the Danish Institute for Economic Modelling and Forecasting (DREAM). 2008 Prices. [Colour figure can be viewed at wileyonlinelibrary.com]

children are neither low nor exactly equal. As shown in Figure 3, annual total expenditures on children range between 8 and 11 billion DKK per year, which is high by international standards. With high levels of investment and diminishing returns, additional interventions might yield small or even zero returns. Figure 3 also shows that expenditure levels are not uniform across age. If older children receive higher baseline investment than younger children, then diminishing returns to investment will mechanically generate a Heckman-type curve. Figure 3 shows that current expenditures do indeed increase up until age 8, but then fall again up until 16. Such a pattern of expenditure is not sufficient to explain the emergence of a Heckman curve. In addition to total expenditures, we also have information about the specific business-as-usual cost in the control group. This cost is highest during day care and thus again not conducive to a Heckman curve.

Note also that Figure 3 potentially masks heterogeneity in spending levels across different socioeconomic groups, especially if one would additionally account for parental investment. Variation in spending could affect the shape of the curve. For example, if returns to investment are diminishing, then the lower initial spending of disadvantaged children will shift their Heckman curve up. In addition, lower initial spending could make investment in older children less productive because of complementarity, thus steepening the curve for disadvantaged children. How exactly spending heterogeneity affects the shape of the curve cannot be said with certainty, but we find it plausible to assume that the Heckman curve is sloping downward for children from all backgrounds. Heckman himself introduced the curve with reference to disadvantaged children in Heckman (2006), but did not add this qualification when giving a more explicit definition in Heckman (2008). From an empirical perspective, when our RCTs explored effect size heterogeneity, effect sizes for disadvantaged children were sometimes larger, smaller, or similar to other children (see Supplemental Information (SI)). As most treatment arms of our RCTs target the general population of children, rather than disadvantaged subgroups, our empirical curve is thus best interpreted as an "average" Heckman curve. That said, some of our interventions do target subgroups of children that lack specific skills, such as bilingual children for whom Danish is a second language. Membership in these subgroups often correlates with socioeconomic disadvantage but not perfectly so. Because it is theoretically unclear how the restriction to subgroups will influence effect sizes, we maintain the assumption that our curve captures an average Heckman curve.

Lastly, the Heckman curve captures long-term returns, whereas this paper is limited to short-term effect sizes. This implies that we cannot provide evidence for driver (3) above because self-productivity and dynamic complementarity will only take effect in the long term. However, our finding that effect sizes are larger in early childhood than in late childhood is consistent with drivers (1) and (2) above. We thus provide evidence for the presence of a Heckman curve under the assumption that the impacts of early investments do not fade out too much. In other words, there must be a sufficiently large degree self-productivity so that the large effects observed in early childhood actually persist into adulthood. We consider this assumption plausible, but ultimately we would need to observe long-run returns to verify this. Note that a Heckman curve could also arise in the absence of (1) and (2), that is, if early investments are not very productive but self-productivity over time is sufficiently large so that future investments become more productive (dynamic complementarity). In this sense, our findings are neither necessary nor sufficient for a Heckman curve to exist, but they do imply that key conditions in favor of its presence are in place.

In sum, testing the Heckman curve requires making a range of assumptions in practice. Moreover, we cannot provide a test of the Heckman curve that has the potential to falsify it. That said, we do provide evidence that some important factors giving rise to the Heckman curve are at work in the Scandinavian welfare state.

2.2 *Existing Evidence*

The interventions summarized in this study add to a large prior literature on ECE programs, school factors, and parent and family programs, most of which is United States based. We briefly review each of these three branches of the literature.

Evaluations of *ECE programs* have grown exponentially since the early demonstration programs in the 1960s and 1970s such as Abecedarian and Perry Preschool, which demonstrated substantial short- and long-term effects (Elango *et al*., 2016). Broadly speaking, this research can be divided into two lines of research. The first line of research focuses on the effects of access to and enrollment in broadly focused ECE programs on children's learning opportunities, such as Head Start, prekindergarten and other center-based programs. The impact of such programs has been summarized in several meta-analyses (Burchinal *et al*., 2010; Duncan and Magnuson, 2013; Fryer, 2017). A second line of research has taken a narrower approach and examined how the quality of children's experiences in preschools can be enriched by opening the "black box" of preschools and evaluating the effects of specific instructional practices carried out in ECE settings, such as shared book reading (Justice *et al*., 2010), phonological awareness training program like Sound Foundation (Byrne and Fielding-Barnsley, 1991), or training of mathematical skills like Building blocks (Clements and Sarama, 2007). Aggregated effects have been published in a massive number of meta-analyses of, for instance, shared book reading (e.g., Swanson *et al*., 2011), language and literacy instructional approaches (Chambers *et al*., 2016), and math interventions (Wang *et al*., 2016).

Both lines of research into ECE have demonstrated wide-ranging evidence of overall positive impacts. Across this work, among the key characteristics that have been associated consistently with positive outcomes are provision of higher quality interactions between teachers and children, appropriately targeted curricula, enriched learning environments, and professional development of teachers. In the light of the considerable increase in early childhood services offered over the past 50 years, research in the field is engaged in debates about how well early childhood services (preschools, prekindergartens, and day care programs) are actually able to boost different groups of children's development and what type of interventions can support the implementation of effective instruction (Center on the Developing Child at Harvard University, 2016; Phillips *et al*., 2017). Moreover, as argued in a review (White *et al*., 2015), the literature is not very precise on the types of interventions that are necessary; in particular, the literature is undecided on a number of important dimensions: quantity–quality trade-offs, the length of day care (hours per day and number of years), the contents of day care, etc. It has also been noted (e.g., White *et al*., 2015) that there is a difference between the aims of early childhood interventions in European countries and the United States, with U.S. interventions being driven by a desire to promote school readiness, while the European view is more holistic, focusing on the whole child.

The amount of research on *school factors* influencing student achievement is overwhelming—as witnessed in a meta-meta-analysis of more than 800 meta-analyses covering more than 50,000 individual studies (Hattie, 2008). However, studies with strong causal identification strategies are rare (which may explain why there appear to be no Campbell reviews within education with school-based interventions and academic achievement as outcome). Again, this research can be divided into two broad lines. One line of research focuses on structural factors such as class size, school size, and budget per student. The interpretation of the results of these kinds of studies has been debated (Hanushek *et al*., 2003; Krueger, 2003). However, more recent studies tend to find positive, albeit small, effects of structural factors such as smaller class size (Jepsen and Rivkin, 2009; Fredriksson *et al*., 2013) and smaller school size (Humlum and Smith, 2015). The other line of research studies what happens within the classroom in terms of teachers' instruction, including feedback to students and tailoring to individual needs (e.g., Kane and Staiger, 2012). Especially within research on early reading instruction, there is considerable evidence of the positive effects of systematic and explicit phonics instruction relative to programs, which are primarily whole-word or whole-language approaches (National Reading Panel, 2000; Ehri *et al*., 2001). However, while structural factors such as class size can be governed relatively effectively through legislation and budgeting, changing the way teachers teach in the classroom—including their use of evidence-based practices—is more difficult to control. One obvious strategy to change teachers' behavior is through professional development (courses and further training). Yet, the evidence for this approach is weak: A review of this literature found that only nine out of more than one thousand studies were capable of supporting causal interpretation (Yoon *et al*., 2007). Surveying only field experiments, Fryer (2017) finds that "managed" professional development that provides teachers with concrete training materials and instructions is more effective than unmanaged professional development.

Strong causal evaluations of *parent and family programs* have not shown quite the same growth as ECE evaluations. Much of the research on parenting knowledge, attitudes, and practices is correlational, and causal evaluations of how to bring effective parenting programs to scale are especially scarce National Academy of Sciences, Engineering, and Medicine (2016). Many parent interventions turn out to be ineffective when implemented at large scale—particularly for socioeconomically disadvantaged families (Pomerantz *et al*., 2007; Castro *et al*., 2015; Kalil, 2015). A recent, comprehensive review of primarily experimental research on how parenting matters for positive child outcomes in the areas of physical health and safety, emotional and behavioral competence, social competence, and cognitive competence identified six practices (National Academies of Sciences, Engineering, 2016)). These include contingent responsiveness ("serve and return"), showing warmth and sensitivity, routines and reduced household chaos, practices that promote children's health and safety, shared book reading and talking to children, and use of appropriate discipline. However, despite a slower start, an emerging body of experimental studies has begun to provide compelling empirical evidence that well-designed parent-directed interventions can increase parental involvement in their children's learning and, more importantly, that these activities translate into positive learning gains for children (Avvisati *et al*., 2014; Guryan *et al*., 2014; Suskind *et al*., 2016; Doyle *et al*., 2017; Mayer *et al*., 2018; York *et al*., 2019; Doyle, 2020).

The RCTs summarized in this study include investments in early childhood and schooling as well as parental interventions and thereby contribute to all three of these large research areas. They are unique, in that, they provide estimates that are not only causally interpretable, but also directly comparable with each other sharing the Danish welfare state as their common setting.

We are only aware of few studies that attempt to find empirical support for the Heckman curve. A study by Rea and Burton is based on 314 cost–benefit analyses conducted by the Washington State Institute of Public Policy, that is, U.S. data (Rea and Burton, 2020). They plot the benefit–cost ratios against the average age of the intervention group and also perform regression analyses. They were not able to confirm the theoretical prediction of declining returns on investments by age of intervention. If anything, they tend to find the opposite. The main explanations they offer are, first, that interventions within age groups differ and not all use the same active ingredients. Second, they propose that interventions placed in adulthood, that is, closer in time to the exposure to many risk factors such as drugs, teenage parenthood, and crime may be better aimed at addressing such issues and may for that reason yield high returns. Both suggested explanations are thus related to the qualitative content of the intervention (quality vs. quantity). They appear to imply that the Heckman curve is better thought of as a curve of the maximum attainable return if interventions are designed appropriately.

Gardner *et al*. (2019) conduct careful meta-analyses with a focus on parenting interventions. They combine an individual participant data meta-analysis of the Incredible Years intervention with a conventional meta-analysis of interventions aimed at reducing disruptive behavior. In both cases, they find no evidence that effect size would depend on age. Although the focus on the same (specific) parenting interventions promotes comparability of effect sizes, it could be misleading if children require different types of investments at different stages of the life cycle. The interventions included in our study are tailored to the presumed needs of the targeted age group.

Other studies compare effect sizes across ages without explicitly referring to the Heckman curve. Fryer (2017) surveys 196 field experiments from developed countries with standardized math and reading scores. He finds a negative correlation between age and effects on reading scores, but no relationship with math scores. A disadvantage of this survey is the inclusion of experiments from different countries and several decades starting in the 1950s, which raises doubts about the comparability of effect sizes across studies. Hollands *et al*. (2013) compare the cost-effectiveness of several early literacy programs aimed at different age groups ranging from kindergarten to third grade. They compute the cost to generate a unit increase in effect size and obtain values between $38 and $38,135. On average, much higher investments are needed in third grade than in kindergarten to obtain the same effect size. Most of the programs included in the study, however, target only low-performing students, which limits the comparability of these programs with programs aimed at the general population of children.

2.3 *The Danish Welfare State*

The Danish welfare state is a prototypical example of the Scandinavian (or universal) welfare state. It aims to foster equality of opportunity as well as to reduce inequality of income. The aim of reducing income inequality is pursued in two ways: directly through income transfers and progressive taxation and indirectly via the free provision of certain goods and services such as free public health insurance. The aim of equal opportunities is pursued in part by making parents' income distributions more equal, but also by large public investment in day care (family-based and center-based care serving children 0–2 years or 3–5 years), free public schools, free education and favorable study grants, free public health care for all, and a lot of specific and targeted policies and institutions aimed at helping those at risk of falling behind.

2.3.1 *Day Care*

As mentioned in the introduction, Denmark provides universal day care with enrollment rates equal to 90% and higher. Day care facilities are organized in public, semipublic, and private centers or in publicly organized nonparental home-based care. Danish day cares are typically age-divided, that is, age 0–2, 3–5, or 0–5 years, while home-based care is delivered in the caregiver's own home and includes one to five children aged 0–2 years with an average of 3.7 children per caregiver (Dalsgaard and Tenney Jordan, 2016). All types of day care facilities are audited by the municipality and are regulated by national legislation on educational quality and safety. Costs are highly subsidized with a minimum of 75% of the cost covered by the municipality, with even larger subsidies for households with low-income and more children. For further description of the universal child care system, see Datta Gupta and Simonsen (2010).

Danish day cares have comparatively high quality in terms of structural characteristics (Esping-Andersen *et al*., 2012). As pointed out in the introduction, the ratio of children to staff (teacher's aides and teachers) is low by international standards. However, there is considerable variation across municipalities, as the child-to-staff ratio is not regulated by the state (Nøhr *et al*., 2012). Approximately 60% of teachers have a 3.5-year pedagogical bachelor degree (Dalsgaard *et al*., 2014) and the other 40% are mainly teacher's aides without postsecondary education.

In Denmark, like many European countries, the day care context typically emphasizes children's social–emotional development, as opposed to a focus on preacademic skills, and is characterized by a holistic and strong free-play–oriented curriculum with little time spent on formal instruction (Bauchmüller *et al*., 2014; Sylva *et al*., 2015). Consequently, there are no "preschools" in Denmark in the American sense. Denmark does not have a national curriculum, but curricula are locally defined, usually at the center level. With a legislation implemented in 2004, all preschools are obliged to formulate so-called "learning plans" focusing on six broad themes: children's all-round individual development, social competence, language development, body and movement, nature and nature phenomena, and cultural expressions and values (Sylva *et al*., 2015). This legislation reflects a broad concept of learning through free play, creativity, and outdoor activities within a social and inclusive context (Jensen, 2009; Bauchmüller *et al*., 2014). The importance of learning through social interaction and in play situations rather than in structured instructional situations, such as circle time or academic activities, was shown in a large-scale survey among 1340 Danish teachers (Broström *et al*., 2014). The stronger focus on social–emotional learning using a free and play-based approach likely affects children's experiences and interactions with both the teacher and peers, hence exposing them to a different educational context than, for instance, that in the United States. A recent review, which examined the relation between structural and interaction quality in preschools in OECD countries using the *Classroom Assessment Scoring System PreK* (CLASS; Pianta *et al*., 2008), placed Denmark in the high range for emotional support, in the medium range for classroom organization, whereas scores for instructional support were in the low range, which is in line with international evidence (Slot, 2018; see also Slot *et al*., 2018).

2.3.2 *School*

The Danish public school is a comprehensive, integrated school covering primary and lower secondary education without tracking. Approximately 80% of all students in primary and lower secondary schools attend the public school, which covers Grade 0, as it is called, to Grade 9. With few exceptions, the rest attend private schools that receive a per student subsidy from the state equaling 75% of the costs of an average public school student. The average number of students per class in the public school is approximately 21 (Panduro, 2017). Students normally stay together in the same class from Grades 0 to 9. Teachers often teach the same class for several years. According to the Public School Act, schools must provide students with subject-specific qualifications and prepare them for further education, but they are also charged to prepare students for their role as citizens in a democratic society. National standardized, computer-based tests have been introduced in a number of subjects in Grades 2–8. School-leaving examinations are taken in Grade 9. Upper secondary education is split into a vocational and an academic (high school type) track, with the former providing qualifying skills of direct use in the labor market, while the latter prepares for further studies. Tertiary education is publicly provided and financed, and no tuition fees are charged. From the age of 18, students are eligible for a study grant (around $900 per month while in the tertiary education system, less if they live with their parents).

2.3.3 *Expenditures*

Figure 3 shows how public expenditures aimed at children and youth are distributed over age and type of expenditure.

The total costs allocated to children and youths up to age 25 years (including collective costs, that is, costs that cannot be directly attributed to any single individual and is therefore allocated evenly to all individuals; police, defense, administration, etc.) correspond to 27% of the total public expenditures and about 14% of GDP in Denmark (which has one of the highest tax pressures in the world in terms of total taxes as a percentage of GDP). In the early ages, the primary expenditures are on health (including a universal nurse home visiting program offered at birth and regular well-child care visits at the GP) and social (including day care) costs, while from ages 6 to 7, costs of education (public schools) start to dominate. Total expenditure levels are largest during primary school years and again during adolescence, around the age of 18, due to eligibility for study grants and other income transfers.

Comparing these numbers to the United States is not straightforward due to differences in definitions etc. Isaacs *et al*. report that public expenditures on children comprise 10.4% of the federal budget and 2.5% of GDP, but they exclude collective costs and only include expenditures up to and including the age of 18 (Isaacs *et al*., 2012). Computing comparable numbers for Denmark, we find expenditures of 16.3% of the public budget and 8.4% of GDP, still considerably more than in the United States. Edelstein *et al*. show that the expenditures in the United States are largest for the 0- to 2-year-old children and then decline uniformly until age 18, which is quite different from the pattern shown in Figure 3 above (Edelstein *et al*., 2012).

3. Methodology

3.1 *Data*

The data used for this study come from several sources. First, we have data from administrative registers compiled by Statistics Denmark for research purposes. This database covers the entire population of Denmark. Most registers cover a period beginning in the 1980s and onward until 2015 (2016 and 2017 for some registers), while a few are only available since 2000 or even later. The administrative data enable linking children to information about their parents and siblings. Important for our purposes, the registers comprise test results from national tests in compulsory school, school absence, school leaving exams, upper secondary and further educational attainments, out-of-home placement and other social interventions, health information, crime (charges, convictions, and incarceration), employment and earnings trajectories. The availability of administrative register data allows for using similar variable definitions, family construction algorithms, etc. across studies. The downside of the Danish administrative registers is that they do not yet include any information on outcomes before school entry. Therefore, such information has been collected as part of each intervention. As a consequence, outcome measures for early ages are more susceptible to measurement errors, which are not present in administrative registers.

Second, for each intervention included, we have specific information about treatment and control groups, the intervention itself, its design and implementation, its costs, and surveys on other relevant individual-specific outcomes not available in the administrative registers (e.g., personality traits in some cases).

3.2 *Interventions*

In this subsection, we briefly present an overview of the interventions included in this study. Inclusion criteria are that the intervention

- was carried out in Denmark by researchers from TrygFonden's Centre of Child Research;
- was evaluated using an RCT;
- is completed, and short-term impacts on cognitive/educational outcomes are available and published, either in a scientific journal or in a publicly available report;
- allows calculation of costs of the intervention (additional costs when compared to the treatment as usual administered to the control group).

All inclusion criteria are motivated by the desire to maximize comparability across studies. The 10 included interventions (comprising 18 treatment arms) are summarized in Table 1 below and described in more detail in the SI.

There is one intervention in center-based day cares serving children 0–2 years (*Play and Learn*) and two interventions in center-based day care serving children 3–5 years (*SPELL* and *LEAP*). All of these interventions involve teaching educators techniques for higher quality interactions with children with the aim of boosting language and emergent literacy skills (and also math language and numeracy skills in day cares serving children 0–2 years). The two interventions in day cares for children 3–5 years have three treatment arms each, which vary with respect to the degree of support of educators, degree of discretion in the planning and implementation of the educational practice/instruction (that is, the extent to which the intervention is based on a curriculum), with the group sizes when treated (small or large groups of children treated simultaneously), and with the degree of parental involvement. These interventions are all aimed at a broad group of children representative of the population of children in Denmark. The final day care intervention, *Suitcase*, is aimed at bilingual children, and it consists of reading material in the mother tongue of the parents and a tutorial DVD explaining the importance of talking to the child.

The *READ* intervention was inspired by the *Suitcase* project but aimed at children in primary school. The intervention was quite similar to the *Suitcase* intervention but focused also on affecting the mindset of the parents. The other two schooling interventions, *Instruction time* and *Teacher's aide*, were aimed at primary school students and consisted of either increasing the number of lessons given or the number of teachers in the classroom. Both interventions had two treatment arms; the *Instruction time* intervention varied the curriculum (fixed or flexible) of the extra lectures, while the *Teacher's aide* intervention varied the quality of the coteacher (teacher with or without teaching degree). The *TMTM* intervention (*Tidlig Matematik indsats Til Marginal elever*) attempted to improve mathematical skills and motivation among both low- and high-performing students through a dialogue-based approach to mathematical teaching. The *Turbo* intervention offered a 2-week intensive learning camp to students who were close to finishing compulsory schooling but lacked the skills to continue education. The *Mentoring* intervention was aimed at youth on social assistance and consisted of weekly sessions with a mentor with the aim of increasing educational entry rates.

Interventions usually target the general population of children. Exceptions are *Suitcase* (bilingual children learning Danish as their second language), *Instruction time* (schools with at least 10% bilingual students), *TMTM* (high- or low-achieving students), *Turbo* (not ready for postcompulsory education), and *Mentoring* (social assistance receipt).

Table 1. Summary of Interventions.

Intervention	Period of implementation	Target group: Age range in years (average)	Target group: Other	Participants and setting	Description	Primary input	Primary outcome	Standardized effect on primary outcome (standard error)	Cost of treatment (per individual in USD)	Cost of treatment as usual (per individual in USD)	Effect size per USD 100	Cost in USD per unit increase in effect size
Play and Learn (Bleses *et al.*, 2016)	2015–2016	1.5–2.9 (2.0)	In public day cares	1508 children across 88 day cares	The 20-week intervention implemented a research-based instructional curriculum called "Play and Learn," targeting language, math and executive function. Teachers received guidance on planning weekly activities and engaging in "serve and return" interactive practices with children.	Quality (improved instruction)	Composite language and math score	0.45 (0.09)	120	18,036	0.38	265
SPELL (Bleses *et al.*, 2015, 2014; Bleses, Højen, Justice, *et al.*, 2018)	2012–2015	3-06 (4.4)	In public day cares	7076 children across 142 day cares	The 20-week intervention was adapted from the U.S. *Read It Again-PreK!* program. Educators read in total 10 books with the children in a way that stimulated language skills more than regular reading does. There were 3 treatment arms: SPELL (baseline), SPELL + HOME (baseline + reading at home with parents), SPELL + PD (baseline + educator receives extra training)	Quality (improved instruction)	Composite language and pre-literacy score	SPELL: 0.27 (0.08) SPELL + HOME: 0.18 (0.05) SPELL + PD: 0.28 (0.14)	SPELL: 46 SPELL + HOME: 61 SPELL + PD: 77	10,034	SPELL: 0.59 SPELL + HOME: 0.30 SPELL + PD: 0.36	SPELL: 170 SPELL + HOME: 337 SPELL + PD: 275

(*Continued*)

Table 1. *Continued.*

Intervention	Period of implementation	Target group: Age range in years (average)	Target group: Other	Participants and setting	Description	Primary input	Primary outcome	Standardized effect on primary outcome (standard error)	Cost of treatment (per individual in USD)	Cost of treatment as usual (per individual in USD)	Effect size per USD 100	Cost in USD per unit increase in effect size
LEAP (Bleses *et al.*, 2015; Bleses, Højen, Dale, *et al.*, 2018)	2012—2015	3–6 (4.4)	In public day cares	6706 children across 154 day cares	The 20-week intervention used play-based activities to target a range of language and literacy learning objectives. Activities included, e.g., memory games, children's songs and storytelling with pictures. There were 3 treatment arms: LEAP-SMALL (baseline), LEAP-LARGE (like baseline, but with entire class rather than small groups), LEAP-OPEN (like baseline, but with more educator autonomy in activities)	Quality (improved instruction)	Composite language and pre-literacy score	LEAP-SMALL: 0.06 (0.08) LEAP-LARGE: 0.04 (0.06) LEAP-OPEN: 0.43 (0.07)	LEAP-SMALL: 49 LEAP-LARGE: 37 LEAP-OPEN: 50	10,034	LEAP-SMALL: 0.12 LEAP-LARGE: 0.11 LEAP-OPEN: 0.85	LEAP-SMALL: 809 LEAP-LARGE: 922x LEAP-OPEN:117
Suitcase (Jakobsen and Andersen, 2013)	2009–2011	4–5 (4.4)	In public day care; learning Danish as 2nd language	284 children across 61 day cares	The intervention was aimed at improving language proficiency, in particular language comprehension. Each family was offered a suitcase containing various children's books, games, and a tutorial DVD about language development techniques. The parents decided whether to use the material. Children were tested ca. 10 months after the suitcase was provided.	Quality (improved parental quality of interaction)	Composite language score	-0.08 (0.21)	179	10,034	−0.05	−-2,185

(*Continued*)

Table 1. *Continued.*

Intervention	Period of implementation	Target group: Age range in years (average)	Target group: Other	Participants and setting	Description	Primary input	Primary outcome	Standardized effect on primary outcome (standard error)	Cost of treatment (per individual in USD)	Cost of treatment as usual (per individual in USD)	Effect size per USD 100	Cost in USD per unit increase in effect size
READ (Andersen and Nielsen *et al.*, 2016)	2013–2016	8–9 (8.1)	In second grade	1587 children across 72 classrooms	The intervention used a growth mindset approach to target reading skills. Parents received not only books, but also information that (a) emphasized a growth theory of abilities by explaining that reading ability can be improved, (b) encouraged parents to support the child's autonomous engagement with books, (c) encouraged parents to praise their child's effort rather than performance. Children were tested 2 and 7 months later.	Quality (improved parental quality of interaction)	Composite reading score (2 months)	0.26 (0.07)	69	7438	0.38	267
Instruction time (Andersen, Humlum, and Nandrup, 2016)	2013–2014	9–10 (9.8)	In fourth grade	1931 students across 90 schools with at least 10% bilingual students	The 16-week intervention involved an increase in instruction time in the subject Danish by 3 hours per week (ca. 15%) at the expense of students' spare time. There were 2 treatment arms: 1. TP: Teachers were required to follow a teaching program developed by national experts, 2. NOTP: Teachers were not required to follow a teaching program, allowing for more need-based instruction.	Quantity (increased instruction time)	Composite reading score	With a teaching program (TP): 0.04 (0.06) Without a teaching program (NOTP): 0.15 (0.06)	165	7952	With a teaching program (TP): 0.02 Without a teaching program (NOTP): 0.09	With a teaching program (TP): 4,458 Without a teaching program (NOTP): 1114

(*Continued*)

Table 1. *Continued.*

Intervention	Period of implementation	Target group: Age range in years (average)	Target group: Other	Participants and setting	Description	Primary input	Primary outcome	Standardized effect on primary outcome (standard error)	Cost of treatment (per individual in USD)	Cost of treatment as usual (per individual in USD)	Effect size per USD 100	Cost in USD per unit increase in effect size
Teacher's aide (Andersen *et al.*, 2014,2018)	2012-2013	12–13 (12.3)	In sixth grade	5213 children across 105 schools	The intervention ran for almost a whole school year and assigned a teacher's aide to each class in the treatment group. There were 2 types of teacher's aides: 1. Co-teacher with a teaching degree (min. 10.5 lessons per week), 2. Teaching assistant without a teaching degree (min. 14.5 lessons per week). The types reflect the trade-off between the teacher aide's qualification and the time spent in class for a fixed wage cost.	Quantity (additional teacher)	Composite reading and math score	Co-teacher with degree: 0.07 (0.04) Teaching assistant without degree: 0.09 (0.04)	Co-teacher with degree: 963 Teaching assistant without degree: 1021	8343	Coteacher with degree: 0.01 Teaching assistant without degree: 0.01	Coteacher with degree: 13,377 Teaching assistant without degree: 11,731
TMTM (Harder et al. 2020)	2017–2018	8–9 (8.1) and 14–15 (14.2)	In second and eighth grade	1501 children across 81 schools	The intervention targeted low and high performers in mathematics. Taught 4 times a week during students' regular lectures, the 12-week intervention sought to improve the mathematical skills/competences and motivation of students through a dialogue-based approach to mathematical teaching. Treatment varied with respect to whether students received one-on-one or group teaching and whether the teacher was supervised.	Quality (change in curriculum)	Math score	Grade 2, low-achieving: 0.38 (0.122) Grade 2, high-achieving: 0.24 (0.086) Grade 8, low-achieving: 0.38 (0.144)	Grade 2, low-achieving: 593 Grade 2, high-achieving: 479 Grade 8, low-achieving: 562	Grade 2: 7438 Grade 8: 9599	Grade 2, low-achieving: 0.06 Grade 2, high-achieving: 0.05 Grade 8, low-achieving: 0.07	Grade 2, low-achieving: 1,561 Grade 2, high-ac.hieving: 1,994 Grade 8, low-achieving: 1,479

(*Continued*)

Table 1. *Continued.*

Intervention	Period of implementation	Target group: Age range in years (average)	Target group: Other	Participants and setting	Description	Primary input	Primary outcome	Standardized effect on primary outcome (standard error)	Cost of treatment (per individual in USD)	Cost of treatment as usual (per individual in USD)	Effect size per USD 100	Cost in USD per unit increase in effect size
Turbo (Rosholm *et al.*, 2020)	2016–2019	14–15	In eighth grade; at-risk students	2184 children across 148 schools	The intervention targeted Grade 8 students who were evaluated to be at risk of lacking the skills to embark on post-compulsory schooling. These students attended a 2-week intensive learning camp that took place during regular school hours at the students own school, but outside the students' normal classroom. It focused on strengthening selected competencies in Danish and math. After the learning camp, the students had weekly 1.5-hour-meetings with a mentor for 8 weeks.	Quality (change in curriculum)	Composite reading and math score	0.100 (0.133)	667	9559	0.02	6,655
Mentoring (Svarer *et al.*, 2014)	2012–2013	18-30 (24.0)	Social assistance receipt	2588 youths across 13 Danish job centers	The intervention targeted youths without a qualifying education, on social assistance, and not perceived to be immediately 'ready for education'. Individuals were matched with a personal mentor, whose primary aim was to help start an education. The secondary aim was to help find a job. Mentor and mentee held weekly meetings. The mentorship ended after 12 months, or as soon as the youth started an education or a job. Individuals could be followed for about 3 years.	Quantity (additional personal advice)	Likelihood of obtaining a vocational degree	0.09 (0.04)	3182	9433	0.003	34,586

All day care interventions feature changes to the curriculum, that is, changes in quality. Some interventions targeted at older age groups also make quantitative changes, such as an increase in instruction time.

Interventions employed a cluster-randomized design that assigned whole childcares or schools to either treatment or control group. An exception is the *Mentoring* intervention, where randomization occurred at the individual level, and the *TMTM* intervention, which randomized schools to treatment arms and assigned students to treatment and control within schools. Most interventions stratified the randomization on variables related to socioeconomic status or children's skills.

Large-scale interventions such as those included here almost inevitably suffer from some degree of noncompliance with treatment assignment. Interventions generally report intention-to-treat effects, which are a good proxy for the true effect if noncompliance is small (e.g., *Suitcase*, *Teacher's aide*, and *Mentoring*) or absent. *TMTM* and *Turbo* additionally address noncompliance by computing local average treatment effects (LATE). Most interventions also suffered from attrition, for example, due to missing responses or posttest data. Differential attrition between treatment and control group was addressed by checking if differences were actually significant and by controlling for observable characteristics.

3.3 *Outcomes*

The outcomes that we study are the primary outcomes selected by the included intervention research teams for measurement of their effectiveness, in all cases some sort of test score or educational attainment. This is specified in Table 1, and more detailed information is provided in the SI. We report short-term measures only, partly for comparative purposes, and partly because most of the interventions were conducted recently, so no medium-term outcome data were yet available, except for very few of them.

For school-based interventions, outcomes are constructed using register based data on national test scores (Nandrup and Beuchert-Pedersen, 2018), ensuring high comparability across interventions. For younger children in preschool-based interventions, no register-based outcome measures of cognitive skills exist in Denmark. Therefore, the language assessments of children were administered by their day care teachers, either by administering a published standardized assessment test (*SPELL* and *LEAP*; Bleses *et al*., 2010), or by employing a standard language test conducted by the municipality at preschool exit (*Suitcase*; Jakobsen and Andersen, 2013), or by completing a standardized checklist (*Play and Learn*, Bleses, Jensen, *et al*., 2018). Prior research shows strong congruence between educators' and researchers' assessments (Cabell *et al*., 2009), so we do not believe that this has introduced any biases.

To ensure comparability of preschool interventions both with each other and with school-based interventions, we report standardized effect sizes, calculated as the estimated impact on the primary outcome divided by the standard deviation of the outcome in the control group. In other words, we express effect sizes in terms of how much outcomes typically vary. The use of standardized effect sizes, although not guaranteeing perfect comparability, is standard practice in meta-analyses and the best we can do in this setting.

3.4 *Costs*

Steuerle and Jackson discuss best practices in the use of economic evidence to informing investments in children and youth (Steuerle and Jackson, 2016). They recommend that the time horizon for the economic evaluation be comparable across studies. In the present study, as mentioned above, all evaluations measure impacts on short-term outcomes (less than a year after the intervention).

For all included interventions, we have access to the total costs of the intervention. For some of them, but not all, we also have access to detailed information on cost items, but we do not use this in the analysis. Costs are always calculated in such way that they measure the extra costs of the intervention (net of research and evaluation costs, which include costs of data collection) over and above the costs of treatment as usual.

We divide the total costs by the number of participants in the treatment group as a measure of the cost of the intervention per participant.[1] Table 1 shows the costs of each intervention, the costs of the treatment as usual, as well as the effect size per USD 100. In the last column, we also report the cost in USD per 1 standard deviation unit increase in effect size. This ratio, which is inversely proportional to effect size per USD 100, is sometimes used in the education literature to compare cost-effectiveness of different programs (e.g., Hollands *et al*., 2013,2016; Levin and Belfield, 2015). We prefer to use the former measure of cost-effectiveness rather than the latter, because effect size per dollar corresponds more closely to the rate of return on investment in the Heckman curve (Figure 1).

Finally, we report the cost of treatment as usual. This is the amount that would have been spent on the members of the treatment group in the absence of the intervention. It can also be viewed as the amount expended on members of the control group. We obtain this information from several sources. For children in day care interventions, we draw on Lemvigh *et al*. (2015), which provides per-child expenses of the average municipality separately for 0- to 2-years-olds (Play&Learn) and 3- to 5-year-olds (SPELL, LEAP, Suitcase). The estimates focus on expenses incurred by municipalities and do not include fees or government grants. Similar numbers for schoolchildren differentiated by grade are provided in Kollin *et al*. (2015), which we use for the READ, instruction time and teacher's aide interventions. For the mentoring intervention, we estimate the cost of treatment as usual as the costs of the active policies to which the youth in the control group are exposed. This includes the cost of being in activation, meetings with case workers and the limited mentoring that also occurs in the control group. Cost of treatment as usual are highest in day care interventions; they are lowest in early compulsory school and subsequently increase again slightly (Table 1). This pattern is not entirely consistent with Figure 3 above. A potential reason for this discrepancy is that Figure 3 reports total public spending across all children, including those not in day care or attending private school, while cost of treatment as usual only applies to those enrolled in the specific public institution. In any case, we consider the numbers in Table 1 as most accurate and will use them in our analysis.

4. Results

In this section, we present comparable results from the 18 intervention arms described above. Figure 4a shows effect sizes for all the interventions against the average age of participants in each intervention.

Effect sizes range from just below zero to almost 0.5. Nearly all intervention arms produce positive effects, although they are not all statistically significant. The most effective interventions are the Play and Learn intervention and LEAP-OPEN. Effect sizes tend to be smaller during compulsory school than in day care. An exception is the TMTM intervention for low-achieving eighth graders, but the standard error is quite large. The mentoring program, which is aimed at young adults on welfare, produces significantly positive, albeit small, effect sizes. Note also that, while we find that the most effective interventions are aimed at younger children, we also find that so are three out of the four least effective interventions. In fact, the standard deviation of effect sizes among interventions aimed at preschool children is about the same as the total standard deviation of effect sizes (0.190 vs. 0.154).

Interventions differ considerably with respect to the cost of treatment. As documented in Table 1, values range between the equivalent of \$37 (LEAP-LARGE) and \$3,182 (Mentoring) per participant. More cost-intense interventions could be expected to yield larger overall effects. Therefore, to make effect sizes comparable, we standardize them by the cost of treatment. Specifically, we compute benefit–cost ratios as effect sizes per \$100 per participant (Figure 4b). Recall from our discussion in Section 2

Figure 4. The Relationship between Average Age at Time of Treatment and Effect Size for the 10 RCTS and Their 18 Treatment Arms. Panel (a): Standard Deviation Adjusted (= Standardized) Effect Size. Panel (b): Standardized Effect Size per $100 Cost. Panel (c). Standardized Effect Size per 1% Cost Increase. "Significant" Indicates If Effects Are Statistically Significant from Zero at the 5% Level. [Colour figure can be viewed at wileyonlinelibrary.com]

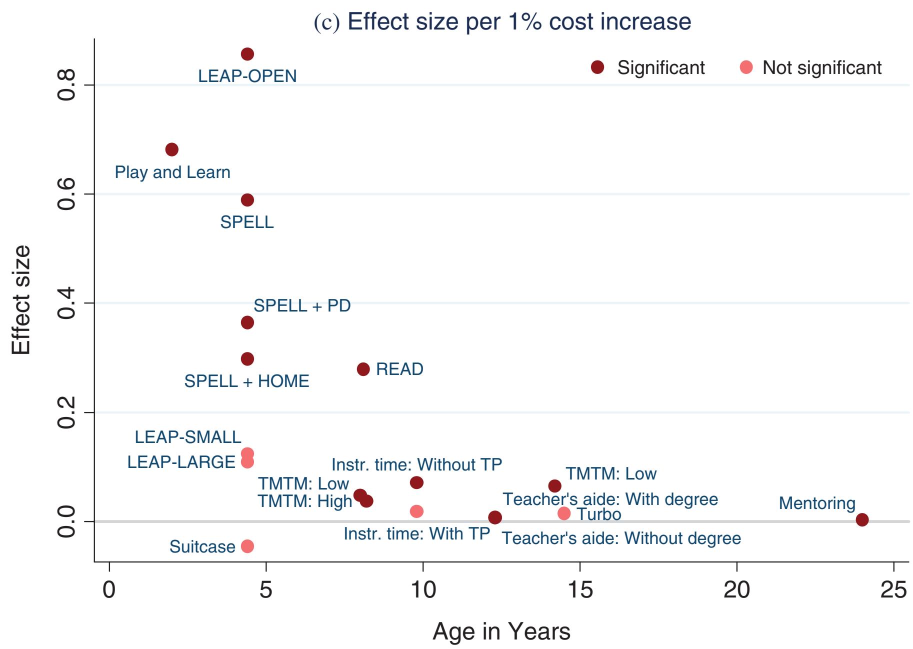

Figure 4. *Continued*

that benefit–cost ratios only reflect average returns; they might underestimate the returns to the first $100 if returns to investment are diminishing, especially for high-cost interventions.

The most cost-effective intervention is LEAP-OPEN, which has a standardized effect size of more than 0.8 per $100 invested. Note that these are not "real" effect sizes but should rather be interpreted as a scale measuring the relative efficiency of the different interventions. The day care interventions we have tested all have relatively low costs, and therefore the most effective among them are highly effective when compared to interventions targeting children in compulsory school or young adults. On the other hand, we once again find that three of the least cost-effective interventions are also aimed at children in day care.

Finally, we standardize treatment effects not only by the cost of treatment, but also by the baseline level of spending at the given age ("cost of treatment as usual"). This yields effect sizes per 1% cost increase. The results are presented in Figure 4c. Since current spending levels are highest during day care (Table 1), the relatively large treatment effects per $100 in the day care interventions (Figure 4b) become reinforced in Figure 4c.

To summarize the outcomes of our interventions, we also conducted a random-effects meta-analysis, using restricted maximum likelihood to estimate the between-studies variance (Thompson and Sharp, 1999), see Figure 5. The overall effect size is equal to 0.18. In other words, the RCTs on average generate gains equal to almost 18% of a standard deviation of the respective outcome. The overall effects per $100 and per 1% cost increase are 0.16 and 0.17, respectively. All overall estimates are statistically significant (Figures 5a–c).

The overall treatment effects mask substantial heterogeneity across interventions ($I^2 = 75\%–100\%$). Specifically, although there is considerable effect size variation within age groups, there also seems to be a strong association between the treatment effect and the average age of the participants in each intervention. Impacts appear larger on average at younger ages, as postulated by the Heckman curve. To investigate this relationship more formally, we have run meta-regressions with the study participants' average age as an explanatory variable (mean = 8.6, st. dev. = 5.4). As Table 2 shows, the estimated

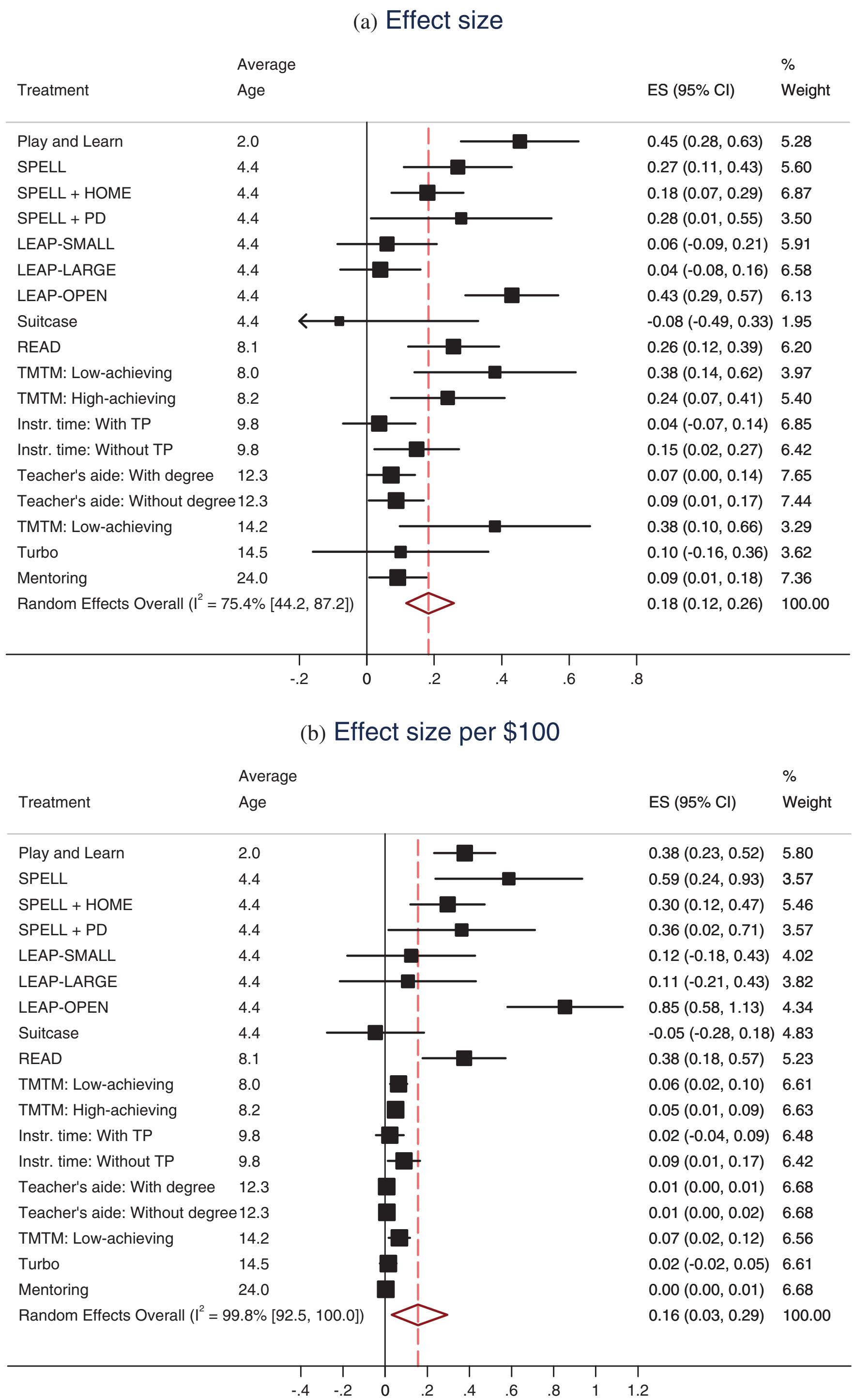

Figure 5. Forest Plots of a Random Effects Meta-Analysis for Different Measures of Effect Size for the 10 RCTS and Their 18 Treatment Arms. Panel (a): Standard Deviation Adjusted (= Standardized) Effect Size. Panel (b): Standardized Effect Size per \$100 Cost. Panel (c). Standardized Effect Size per 1% Cost Increase. 95% Bias-Corrected Confidence Intervals from 2000 Bootstrap Replications Clustered at the Intervention-Level Are Reported. [Colour figure can be viewed at wileyonlinelibrary.com]

(c) Effect size per 1% cost increase

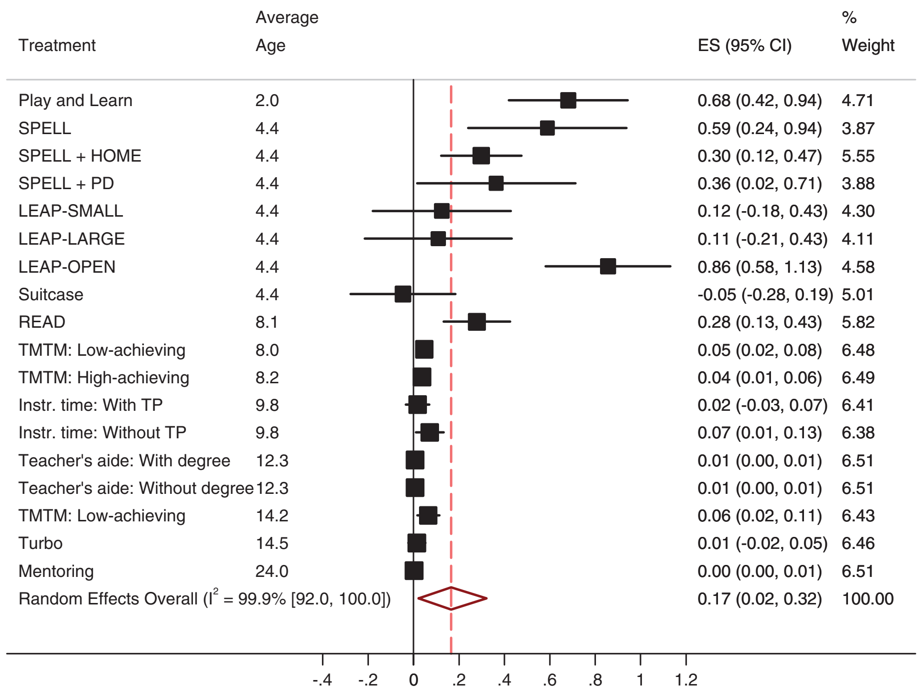

Figure 5. *Continued*

Table 2. Estimates from Meta-Regressions of Various Measures of Effect Size on Age. Brackets (parentheses) Include 95% (90%) Bias-Corrected Confidence Intervals from 2000 Bootstrap Replications Clustered at the Intervention Level.

	(a) Effect size	(b) Effect size per \$100	(c) Effect size per 1% cost increase
Age	−0.009	−0.020	−0.024
	[−0.025, 0.001]	[−0.042, −0.004]	[−0.052, −0.002]
	(−0.022, −0.002)	(−0.040, −0.005)	(−0.048, −0.003)
Constant	0.259	0.339	0.379
	[0.147, 0.399]	[0.075, 0.526]	[0.047, 0.614]
	(0.169, 0.370)	(0.089, 0.501)	(0.061, 0.568)
Adjusted R^2	0.14	0.39	0.36
N	18	18	18

coefficient on age has a negative sign and, additionally, it is significantly different from zero at least at the 10% level. Age explains between 14% and 39% of the total variation in raw and cost-adjusted effect sizes across studies. In Figure 6, which plots the regression results, the lower bound of the 95% prediction interval crosses the zero-effect line somewhere between 13 and 16 years of age, suggesting that additional investments tend to become ineffective in early adolescence.

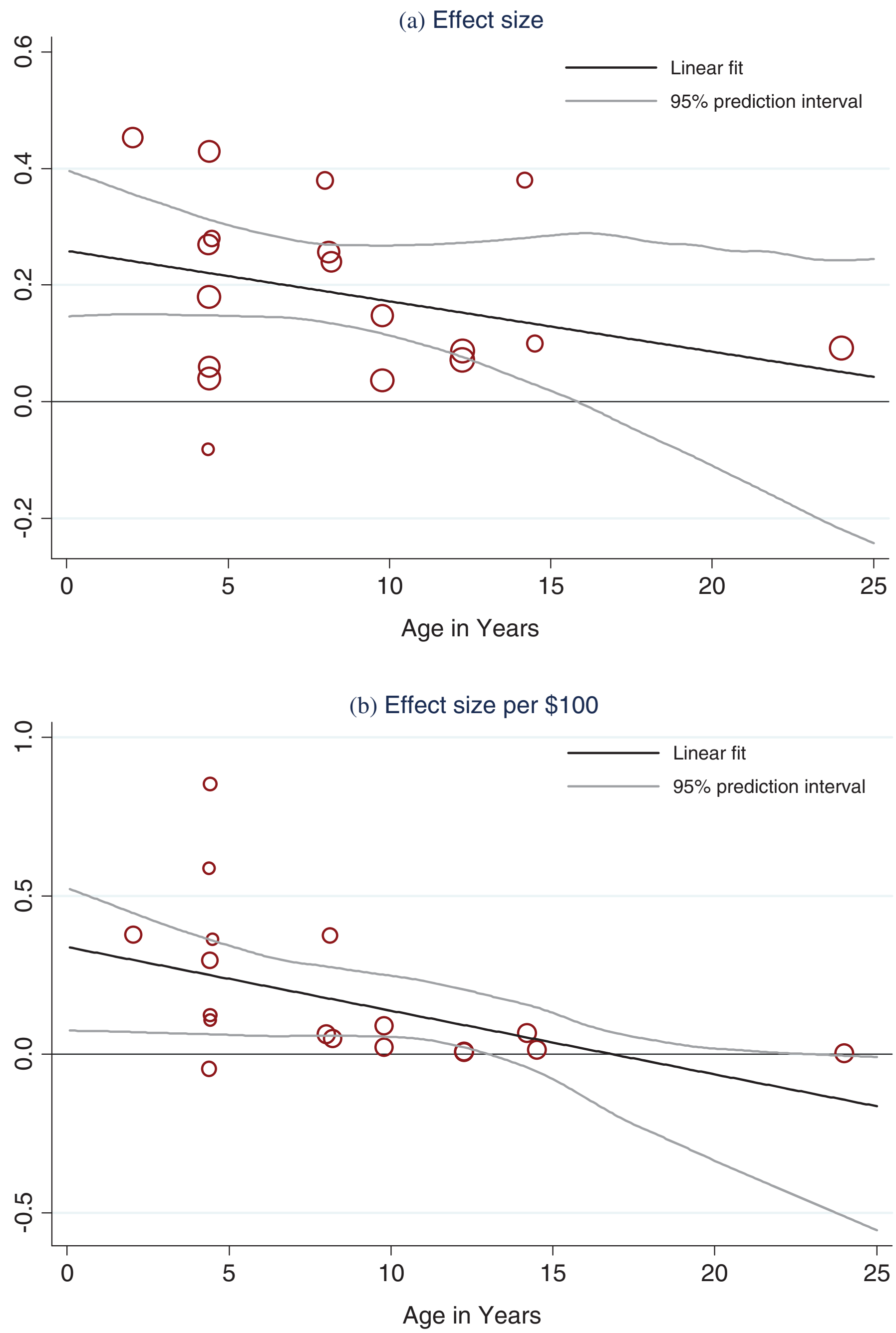

Figure 6. Meta-Regressions of Various Measures of Effect Size on Average Age in Years at Time of Treatment for the 10 RCTS and Their 18 Treatment Arms (= *N*). Panel (a): Standard Deviation Adjusted (= Standardized) Effect Size. Panel (b): Standardized Effect Size per \$100 Cost. Panel (c). Standardized Effect Size Per 1% Cost Increase. The Size of the Circle Indicates the Weight of Each Study, Which Is Equal to the Inverse of the Overall Study Error Variance (Between-Studies Variance Plus Within-Study Error Variance). 95% Bias-Corrected Prediction Intervals Are Based on 2000 Bootstrap Replications Clustered at the Intervention Level. [Colour figure can be viewed at wileyonlinelibrary.com]

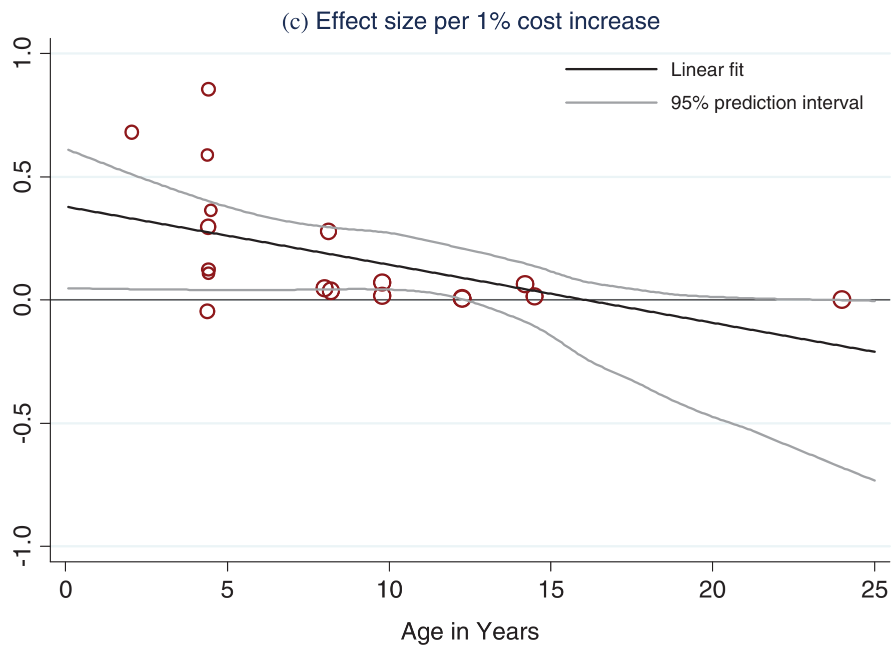

Figure 6. *Continued*

Even though publication status was not an inclusion criterion in this meta-analysis, there might still be selective reporting of findings based on the sign and significance of effects. We apply a battery of modern techniques that address potential publication bias (see SI). In summary, there is some indication that selective reporting pushes up our estimate of the overall effect size, which after bias-correction falls from 0.18 to around 0.10. However, estimates are imprecise, and our baseline estimate almost always lies within the confidence band of the corrected estimate. Importantly, selective reporting appears to be unrelated to age and does not alter our finding from the meta-regression that effect size decreases with age.

5. Discussion

Our main finding is that almost all interventions have positive impacts. More specifically, several interventions aimed at improving the learning environment in day care have remarkably large impacts.

This is quite impressive for a collection of large-scale effectiveness ("real world") studies, especially in a Scandinavian welfare state, where spending levels are already high (in the form of high-quality universal day care, free public schools, etc.).

All the interventions we study added relatively few components on top of the "treatment as usual" during a fairly short span of time. With diminishing returns to investment, this marginal investment might not have been expected to yield large returns. Indeed, the Heckman curve is explicitly defined for settings with low initial investment, unlike the Danish welfare state. Yet, we often find fairly large effect sizes across all ages, suggesting that it is not (very) difficult to obtain better results than what is already provided by the comprehensive welfare state policies. Perhaps this reflects a low level of quality in existing investment, such as poorly designed curricula and ineffective day care and classroom practices. Irrespective of what explains these large returns, they should be interpreted as the highest feasible returns under the current Danish welfare state.

Furthermore, we find a tendency for cost-standardized treatment effects (= benefit–cost ratios) to decline with age and become small in adolescence. Interventions in day cares generate large effects in some cases, while school-based interventions have positive but smaller impacts. Many interventions aimed at older individuals, including the mentoring program, are effective in terms of effect size, but have a very low return per dollar invested due to their high cost. Overall, the observation that investments are most cost-effective at young ages provides evidence that the Heckman curve (Figure 1) can be justified empirically.

The Heckman curve is defined for marginal returns, but benefit–cost ratios likely underestimate marginal returns in the presence of diminishing returns to investment, in particular for costly interventions (see Section 2). A Heckman curve could emerge simply because interventions for older children are more expensive. In our setting, interventions for older children do indeed tend to be more expensive. However, we show that also the raw effect sizes *before* dividing by cost are smaller for older children. In other words, interventions for young children are overall more productive despite being cheaper. This strongly suggests that also the marginal dollar invested is more productive at younger ages.

Our findings are consistent with the model of skill formation. A rationalization for the large effects of early investments is that they boost skills more (larger productivity of skills) and depend less on the existing stock of skills (lower complementarity) than late investments. Cunha *et al*. (2010) previously provided evidence for these two features of the skill formation process. A caveat is that we can only measure short-term effects. This prevents us from verifying that these effects actually persist into adulthood (self-productivity) and enhance the productivity of future investments (dynamic complementarity). That said, we provide evidence that a key driver of the Heckman curve, that is, large short-term returns to early investments, is at work. While this is strictly speaking not necessary for the (long-term) Heckman curve to arise, it is substantial evidence in favor of it.

Our findings contrast with, for example, Rea and Burton (2020), who do not find evidence of a Heckman curve. There are several potential reasons for this difference. First, our meta-analysis assembles studies from a homogeneous setting, while Rea and Burton draw on studies from various places (United States and elsewhere) over a longer time span. Second, Rea and Burton focus on benefit–cost ratios, which—as discussed above—are only imperfect proxies for marginal returns. Unlike Rea and Burton, we are able to show that also raw returns are larger for young children, suggesting a Heckman curve in not only benefit–cost ratios, but also marginal returns. Finally, we focus on short-term effects measured immediately after the end of the interventions. In contrast, Rea and Burton consider long-term effects. Although long-term effects are consistent with the definition of the Heckman curve, in practice they require various assumptions that might entail considerable bias.

Early investment can be highly effective, but we also find that the variation in effect sizes during early childhood is about as large as the total variation in effect size across all ages. Age only explains at most between 14 and 39 percent of the variation in cost-standardized effect sizes and some childhood interventions are practically ineffective. This finding emphasizes the important point that the Heckman curve is to be interpreted as a technological possibility frontier (see Section 2). The curve shows the largest potential return on investment during various stages of development. This potential return is largest during early childhood and will only be realized if the intervention is optimally designed. The return of a suboptimally designed intervention might fall significantly short of this best-practice return.

The interpretation of the Heckman curve as a frontier may explain why Scandinavian welfare states fail to realize high levels of equality of opportunity despite large expenditures. The investments of the welfare states do not attain the possibility frontier, and therefore children do not reach their full potential because resources are wasted on less than optimal interventions. The question then is, how can we remedy this failure of the welfare state?

To address this question, we present three broad lessons that emerge from our analysis and the studies underlying it:

- Establish a culture of evaluation and systematically collect robust evidence on what type of early life interventions are effective, depending on the setting. As a starting point, existing programs can be assessed with respect to their gaps and limitations. The experience from the studies included here is that interventions work well if carefully designed in close collaboration with practitioners.
- Ensure a more equal focus on all aspects of school readiness, including both content-specific cognitive (language, math) and social–emotional skills and more general domain skills like executive functions.
- Reallocate resources toward high-quality day care at the expense of school investments. Assuming that public budgets are constrained, spending on children should be more concentrated in effective day care, where it is most productive.

6. Limitations

This study has several limitations that influence the generalizability of the conclusions. First, we only consider short-term effects, which are not necessarily persistent. Short-term effects could fade out over time, but they could also be reinforced if they make future investments more productive. There is a vivid debate about fading-out of early childhood interventions (e.g., Duncan and Magnuson, 2013; Elango *et al*., 2016; Bailey *et al*., 2017). We restrict the focus to short-term effects because all interventions took place recently, and we do not have long follow-up periods for many of them. However, we plan to investigate medium- to long-run outcomes in a future paper. It is reassuring that in the READ intervention, effect sizes measured after 2 months are still significant, though slightly diminished, when measured again after 7 months. Follow-up data (not yet published) show that these effects remain at this level after 31 months.

Second, we investigate only impacts on cognitive outcomes in this study. There is a large body of research pointing to the importance of socio-emotional skills for later-life outcomes (e.g., Heckman *et al*., 2013). Effect sizes in the studies included in this analysis may be less than optimal in terms of academic performance and cognitive skills, but it may be that there are large and persistent effects on other aspects of the children's development—which partly is the intention of the Danish system. In the present study, however, the available register data and surveys constrain our focus to cognitive ability, which also plays a particularly important role when it comes to providing equality of opportunity in education.

Third, when evaluating interventions in day cares, we cannot use register data but must rely on assessments made by the educators that also implemented the intervention. These assessments are more prone to be subject to measurement error, especially if educators evaluate their own performance in implementing the treatment rather than children's progress. However, as argued above, previous research shows strong congruence between educators' and researchers' assessments (Cabell *et al*., 2009), so that the resulting bias is presumably negligible.

Fourth, costs are treated as fixed in this study. In practice, however, costs cannot be measured without errors and uncertainty; this should be borne in mind when interpreting our findings on cost-standardized effects.

Fifth, it should also be noted that we have only 18 treatment arms and, hence, effect sizes. However, they do represent a large number of children in total ($N = 30{,}578$). Still, conclusions on the age distribution of effect sizes suffer from the lack of interventions aimed at young adults.

Sixth, we normalize effect sizes only by their standard deviation to make them comparable across studies. We do not anchor outcomes in a common metric, such as monetary returns or years of schooling, as previous studies have done. This is because we do not observe long-run outcomes and projecting short-term effects into the future would require too strong assumptions.

Seventh, we consider the internal validity of our study as high given that all interventions took place in the same country and around the same time. However, most of our interventions are implemented at the classroom level (all except for the mentoring program). Therefore, the estimated treatment effect not only captures the intervention's direct impact on the individual child, but also potential spillovers stemming from other children in the treated classroom. Indeed, spillover effects within classrooms could be an important channel through which our interventions operate. This should be borne in mind when interpreting the absolute size of the treatment effects, but is not relevant for comparisons across studies, given that almost all studies are implemented at the classroom level. Regarding the external validity, our results are most relevant to other Scandinavian welfare states. Most likely, the interventions would give even larger effect sizes in most other countries, which have lower spending on children, particularly during day care.

Eighth, each of our interventions was carried out in a small subset of Danish day cares or schools. Participation was voluntary. Scaling up the interventions to the whole country could affect implementation fidelity and thereby treatment benefits. It should be noted, however, that most interventions were already conducted at a relatively large scale (see Table 1), with no or limited reinforcement or compliance support. Relatedly, all interventions modify a baseline scenario to which the control group continues to be exposed. We generally view this baseline scenario as representative of the ordinary instruction going on in Danish childcares and schools. That said, voluntary participation could have led to selection bias, even if some interventions tried to ensure a diverse distribution (*SPELL*, *LEAP*, *Teacher's aide*).

Ninth, there might be alternative explanations for the negative relationship of treatment effect size with age. For example, if children are more similar to each other at young ages with respect to their skill level, a more unequal skill distribution at later ages could affect treatment impacts, for example, through Matthew effects.

Finally, the Heckman curve is defined for marginal returns to investment, while the interventions in our study cost between 40 and 50 and several hundred dollars per participant. We hence use benefit–cost ratios, but these underestimate marginal returns in the presence of diminishing productivity. In addition, interventions tend to be more expensive for older children, which might mechanically generate a Heckman curve. Although we argued in the discussion that the use of benefit–cost ratios is unlikely to drive our findings, they would be a better proxy for marginal returns if cost were small and uniform across interventions. This should be borne in mind for future attempts to test the Heckman curve.

7. Conclusion

This paper surveys the results from 10 RCTs including 18 treatments arms conducted recently in the Scandinavian welfare state of Denmark. We find that, despite already large levels of early investment, additional marginal investments can have sizable effects, also after standardizing effect sizes with the cost of treatment. Graphical evidence together with meta-regressions confirm that effect sizes decline as age increases, consistent with the presence of a Heckman curve.

Variation in the size of the effect within age groups is as large as it is across age groups, suggesting that the impact of an intervention is not only determined by its timing, but also its quality. This calls attention to the interpretation of the Heckman curve as a production possibility frontier, showing the attainable effect when the intervention combines the most effective ingredients. Lack of quality might explain why the large investments of the Scandinavian welfare state do not translate into more equality of opportunity. The call for ever more early investment is unjustified if interventions are designed poorly, while carefully designed late interventions can have respectable impacts.

Note

1. To conduct full cost-benefit analyses on all interventions, we would need access to longer-term outcome data anchored to adult outcomes such as educational attainments.

References

Allen, G. (2011) *Early Intervention: The Next Steps*. London, UK: HM Government. https://www.gov.uk/government/publications/early-intervention-the-next-steps-2 (last accessed: November 28, 2020).

Andersen, S.C., Beuchert, L., Nielsen, H.S. and Thomsen, M.K. (2014) 2L Rapport – Undersøgelse Af Effekten Af Tolærerordninger [Evaluating the effect of co-teaching]. https://childresearch.au.dk/fileadmin/childresearch/dokumenter/Publikationer/Rapport_om_forsoeg_med_tolaererordninger_3_.pdf (last accessed: November 28, 2020).

Andersen, S.C., Beuchert, L., Nielsen, H.S. and Thomsen, M.K. (2018) The effect of teacher's aides in the classroom: evidence from a randomized trial. *Journal of the European Economic Association*, December. https://doi.org/10.1093/jeea/jvy048.

Andersen, S.C., Humlum, M.K. and Nandrup, A.B. (2016) Increasing instruction time in school does increase learning. *Proceedings of the National Academy of Sciences* 113(27): 7481–7484. https://doi.org/10.1073/PNAS.1516686113.

Andersen, S.C. and Nielsen, H.S. (2016) reading intervention with a growth mindset approach improves children's skills. *Proceedings of the National Academy of Sciences* 113(43): 12111–12113. https://doi.org/10.1073/pnas.1607946113.

Avvisati, F., Gurgand, M., Guyon, N. and Maurin, E. (2014) Getting parents involved: a field experiment in deprived schools. *The Review of Economic Studies* 81(1): 57–83. https://doi.org/10.1093/restud/rdt027.

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. https://doi.org/10.1080/19345747.2016.1232459.

Bauchmüller, R., Gørtz, M. and Rasmussen, A.W. (2014) Long-run benefits from universal high-quality preschooling. *Early Childhood Research Quarterly* 29(4): 457–70. https://doi.org/10.1016/J.ECRESQ.2014.05.009.

Black, M.M., Walker, S.P., Fernald, L.C.H., Andersen, C.T., DiGirolamo, A.M., Lu, C., McCoy, D.C., *et al*. (2017) Early childhood development coming of age: science through the life course. *The Lancet* 389(10064): 77–90. https://doi.org/10.1016/S0140-6736(16)31389-7.

Bleses, D., Højen, A., Andersen, M.K., Dybdal, L. and Sehested, K. (2015) Sproget Kan Styrkes! Rapport Om SPELL Og Fart På Sproget. Undersøgelser Af Effekten Af to Sprogindsatser. http://www.e-pages.dk/aarhusuniversitet/1526/ (last accessed: November 28, 2020).

Bleses, D., Højen, A., Dale, P.S., Justice, L.M., Dybdal, L., Piasta, S., Markussen-Brown, J., Kjærbæk, L. and Haghish, E.F. (2018) Effective language and literacy instruction: evaluating the importance of scripting and group size components. *Early Childhood Research Quarterly* 42(January): 256–269. https://doi.org/10.1016/J.ECRESQ.2017.10.002.

Bleses, D., Højen, A., Dybdal, L., Dale, P., Justice, L., Piasta, S., Markussen-Brown, J., *et al*. (2014) A randomized controlled trial of a large scale systematic and explicit language and literacy intervention in Danish daycares: the SPELL study. *British Journal of Education, Society & Behavioural Science* 4(11): 1504–1524. https://doi.org/10.9734/BJESBS/2014/11691.

Bleses, D., Højen, A., Justice, L.M., Dale, P.S., Dybdal, L., Piasta, S.B., Markussen-Brown, J., Clausen, M., and Haghish, E. F. (2018) The effectiveness of a large-scale language and preliteracy intervention: the SPELL randomized controlled trial in Denmark. *Child Development* 89(4): e342–e363. https://doi.org/10.1111/cdev.12859.

Bleses, D., Jensen, P., Højen, A. and Dale, P.S. (2018) An educator-administered measure of language development in young children. *Infant Behavior and Development* 52(August): 104–113. https://doi.org/10.1016/J.INFBEH.2018.06.002.

Bleses, D., Jensen, P., Nielsen, H. and Rathe, A. (2016) Indsats i Vuggestuen Kan Styrke Børns Tidlige Kompetencer. Rapport Om Effekten Af Leg Og Læring i Vuggestuen – En Tidlig Målrettet Indsats i Danske Vuggestuer. http://www.e-pages.dk/aarhusuniversitet/1600/ (last accessed: November 28, 2020).

Bleses, D., Vach, W., Jørgensen, R.N. and Worm, T. (2010) The internal validity and acceptability of the Danish SI-3: a language-screening instrument for 3-year-olds. *Journal of Speech, Language, and Hearing Research* 53(2): 490–507. https://doi.org/10.1044/1092-4388(2009/08-0132).

Broström, S., Johansson, I., Sandberg, A. and Frøkjær, T. (2014) Preschool teachers' view on learning in preschool in Sweden and Denmark. *European Early Childhood Education Research Journal* 22(5): 590–603. https://doi.org/10.1080/1350293X.2012.746199.

Burchinal, M., Vandergrift, N., Pianta, R. and Mashburn, A. (2010) Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly* 25(2): 166–176. https://doi.org/10.1016/J.ECRESQ.2009.10.004.

Byrne, B. and Fielding-Barnsley, R. (1991) evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology* 83(4): 451–455.

Cabell, S.Q., Justice, L.M., Zucker, T.A. and Kilday, C.R. (2009) Validity of teacher report for assessing the emergent literacy skills of at-risk preschoolers. *Language, Speech, and Hearing Services in Schools* 40(2): 161–173. https://doi.org/10.1044/0161-1461(2009/07-0099).

Castro, M., Expósito-Casas, E., López-Martín, E., Lizasoain, L., Navarro-Asencio, E. and Gaviria, J.L. (2015) Parental involvement on student academic achievement: a meta-analysis. *Educational Research Review* 14(February): 33–46. https://doi.org/10.1016/J.EDUREV.2015.01.002.

Center on the Developing Child at Harvard University (2016) From best practices to breakthrough impacts: a science-based approach to building a more promising future for young children and families. http://www.developingchild.harvard.edu/ (last accessed: November 28, 2020).

Chambers, B., Cheung, A.C.K. and Slavin, R.E. (2016) Literacy and language outcomes of comprehensive and developmental-constructivist approaches to early childhood education: a systematic review. *Educational Research Review* 18(May): 88–111. https://doi.org/10.1016/J.EDUREV.2016.03.003.

Clements, D.H. and Sarama, J. (2007) Effects of a preschool mathematics curriculum: summative research on the building blocks project. *Journal for Research in Mathematics Education* 38(2): 136–163.

Cunha, F. and Heckman, J.J. (2007) The technology of skill formation. *American Economic Review* 97(2): 31–47. https://doi.org/10.1257/aer.97.2.31.

Cunha, F. and Heckman, J.J. (2008) Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43(4): 738–782. https://doi.org/10.3368/jhr.43.4.738.

Cunha, F., Heckman, J.J. and Schennach, S.M. (2010) Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78(3): 883–931. https://doi.org/10.3982/ECTA6551.

Dalsgaard, C., Nøhr, K. and Jordan, A.L.T. (2014) Personale Og Børn i Kommunernes Dagtilbud - En Undersøgelse Af Perioden 2007-2012. https://www.vive.dk/da/udgivelser/personale-og-boern-i-kommunernes-dagtilbud-9053/ (last accessed: November 28, 2020).

Dalsgaard, C. and Jordan, A.L.T. (2016) Udvikling i Personaleforbrug i Daginstitutionerne: Normeringer Og Pædagogandele i Kommunerne 2010-2014. https://www.vive.dk/da/udgivelser/udvikling-i-personaleforbrug-i-daginstitutionerne-8817/ (last accessed: November 28, 2020).

Datta Gupta, N. and Simonsen, M. (2010) Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics* 94(1–2): 30–43. https://doi.org/10.1016/j.jpubeco.2009.10.001.

Doyle, O. (2020) The first 2,000 days and child skills. *Journal of Political Economy* 128(6): 2067–2122. https://doi.org/10.1086/705707.

Doyle, O., Harmon, C., Heckman, J.J., Logue, C. and Moon, S.H. (2017) Early skill formation and the efficiency of parental investment: a randomized controlled trial of home visiting. *Labour Economics* 45(April): 40–58. https://doi.org/10.1016/j.labeco.2016.11.002.

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

Edelstein, S., Isaacs, J.B., Hahn, H. and Toran, K. (2012) How do public investments in children vary with age? A kids' share analysis of expenditures in 2008 and 2011 by age group. https://www.urban.org/research/publication/how-do-public-investments-children-vary-age-kids-share-analysis-expenditures-2008-and-2011-age-group (last accessed: November 28, 2020).

Ehri, L.C., Nunes, S.R., Stahl, S.A. and Willows, D.M. (2001) Systematic phonics instruction helps students learn to read: evidence from the national reading panel's meta-analysis. *Review of Educational Research* 71(3): 393–447. https://doi.org/10.3102/00346543071003393.

Elango, S., García, J.L., Heckman, J.J. and Hojman, A. (2016) Early childhood education. In Robert A. Moffitt (ed.), *Economics of Means-Tested Transfer Programs in the United States*, Vol. 2 (pp. 235–297). Chicago, IL: University of Chicago Press. https://www.nber.org/chapters/c13489.

Esping-Andersen, G., Han, W.-J., Magnuson, K. and Wagner, S. (2012) Child care and school performance in Denmark and the United States. *Children and Youth Services Review* 34(3): 576–589. https://doi.org/10.1016/J.CHILDYOUTH.2011.10.010.

Fredriksson, P., Öckert, B. and Oosterbeek, H. (2013) Long-term effects of class size. *The Quarterly Journal of Economics* 128(1): 249–285. https://doi.org/10.1093/qje/qjs048.

Fryer, R.G (2017) The production of human capital in developed countries: evidence from 196 randomized field experiments. In Banerjee, Abhijit Vinayak & Duflo, Esther (eds.), *Handbook of Economic Field Experiments*, Vol. 2 (pp. 95–322). North-Holland:Elsevier. https://doi.org/10.1016/BS.HEFE.2016.08.006.

Gardner, F., Leijten, P., Melendez-Torres, G.J., Landau, S., Harris, V., Mann, J., Beecham, J., Hutchings, J. and Scott, S. (2019) The earlier the better? Individual participant data and traditional meta-analysis of age effects of parenting interventions. *Child Development* 90(1): 7–19. https://doi.org/10.1111/cdev.13138.

Guryan, J., Kim, J. and Quinn, D. (2014) *Does reading during the summer build reading skills? Evidence from a randomized experiment in 463 classrooms*. 20689. NBER Working Paper. Cambridge, MA. https://www.nber.org/system/files/working_papers/w20689/w20689.pdf.

Hanushek, E.A., Kain, J.F., Markman, J.M. and Rivkin, S.G. (2003) Does peer ability affect student achievement? *Journal of Applied Econometrics* 18(5): 527–544. https://doi.org/10.1002/jae.741.

Harder, J., Færch, J. V., Maim, S. G., Overgaard, S., Rasmusse, K., Tonnesen, P. and Rosholm, M. (2020) *Sammenfatning af følgeforskningen på Matematikindsats 2017–TMTM, Tidlig Matematikindsats Til Marginalgruppeelever*. Trygfondens Centre for Child Research and University College Copenhagen, https://childresearch.au.dk/nyheder/nyhed/artikel/sammenfatning-af-foelgeforskningenpaa-matematikindsats-2017-tmtm-tidlig-matematikindsats-til-margi/ (last accessed: November 28, 2020).

Hattie, J. (2008) *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London, UK: Routledge. https://www.taylorfrancis.com/books/9780203887332.

Heckman, J.J. (2006) Skill formation and the economics of investing in disadvantaged children. *Science* 312(5782): 1900–1902. https://doi.org/10.1126/SCIENCE.1128898.

Heckman, J.J. (2008) Schools, skills, and synapses. *Economic Inquiry* 46(3): 289–324. https://doi.org/10.1111/j.1465-7295.2008.00163.x.

Heckman, J.J. (2020) Comment on "New Evidence on the Heckman Curve" by David Rea and Tony Burton. *Journal of Economic Surveys*.

Heckman, J.J. and S. Mosso (2014) The economics of human development and social mobility. *Annual Review of Economics* 6(1): 689–733. https://doi.org/10.1146/annurev-economics-080213-040753.

Heckman, J.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): 2052–2086. https://doi.org/10.1257/aer.103.6.2052.

Hollands, F.M., Kieffer, M.J., Shand, R., Pan, Y., Cheng, H. and Levin, H.M. (2016) Cost-effectiveness analysis of early reading programs: a demonstration with recommendations for future research. *Journal of Research on Educational Effectiveness* 9(1): 30–53. https://doi.org/10.1080/19345747.2015.1055639.

Hollands, F.M., Pan, Y., Shand, R., Cheng, H., Levin, H.M., Belfield, C.R., Kieffer, M., Brooks Bowden, A. and Hanisch-Cerda, B. (2013) Improving early literacy: cost-effectiveness analysis of effective reading programs. https://www.cbcse.org/publications/improving-early-literacy-cost-effectiveness-analysis-of-effective-reading-programs (last accessed: November 28, 2020).

Humlum, M.K. and Smith, N. (2015) Long-term effects of school size on students' outcomes. *Economics of Education Review* 45(April): 28–43. https://doi.org/10.1016/J.ECONEDUREV.2015.01.003.

Isaacs, J., Toran, K. and Hahn, H. (2012) Kids' share 2012 – report on federal expenditures on children through 2011. https://www.urban.org/research/publication/kids-share-2012 (last accessed: November 28, 2020).

Jakobsen, M. and Andersen, S.C. (2013) Coproduction and equity in public service delivery. *Public Administration Review* 73(5): 704–713. https://doi.org/10.1111/puar.12094.

Jensen, B. (2009) A Nordic approach to early childhood education (ECE) and socially endangered children. *European Early Childhood Education Research Journal* 17(1): 7–21. https://doi.org/10.1080/13502930802688980.
Jepsen, C. and Rivkin, S. (2009) class size reduction and student achievement. *Journal of Human Resources* 44(1): 223–250. https://doi.org/10.3368/jhr.44.1.223.
Justice, L.M., McGinty, A.S., Piasta, S.B., Kaderavek, J.N. and Fan, X. (2010) Print-focused read-alouds in preschool classrooms: intervention effectiveness and moderators of child outcomes. *Language, Speech, and Hearing Services in Schools* 41(4): 504–520. https://doi.org/10.1044/0161-1461(2010/09-0056).
Kalil, A. (2015) Inequality begins at home: the role of parenting in the diverging destinies of rich and poor children. In Paul R. Amato, Alan Booth, Susan M. McHale, and Jennifer Van Hook (eds.), *Families in an Era of Increasing Inequality: Diverging Destinies*, pp. 63–82. New York: Springer. https://doi.org/10.1007/978-3-319-08308-7_5.
Kane, T.J. and Staiger, D.O. (2012) Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. MET Project. *Bill & Melinda Gates Foundation*. https://eric.ed.gov/?id=ED540960.
Karoly, L.A. (2012) Toward standardization of benefit-cost analysis of early childhood interventions. *Journal of Benefit-Cost Analysis* 3(1): 1–45. https://doi.org/10.1515/2152-2812.1085.
Kollin, M.S., Jordan, A.L.T., Dalsgaard, C. and Madsen, A. (2015) Så Meget Koster En Skoleelev – Kommunernes Enhedsudgifter På Skoleområdet 2009/2010-2012/2013. https://www.vive.dk/da/udgivelser/saa-meget-koster-en-skoleelev-9008/ (last accessed: November 28, 2020).
Krueger, A.B. (2003) Economic considerations and class size. *The Economic Journal* 113(485): F34–63. https://doi.org/10.1111/1468-0297.00098.
Landersø, R. and James, J.H. (2017) The scandinavian fantasy: the sources of intergenerational mobility in Denmark and the US. *The Scandinavian Journal of Economics* 119(1): 178–230. https://doi.org/10.1111/sjoe.12219.
Lemvigh, K., Petersen, J.S., Dalsgaard, C. and Haslund-Thomsen, J. (2015) Så Meget Koster et Barn i Dagtilbud – Kommunernes Enhedsudgifter Til Dagtilbud 2009-2013. https://www.vive.dk/da/udgivelser/saa-meget-koster-et-barn-i-dagtilbud-9005/.
Levin, H.M. and Belfield, C. (2015) Guiding the development and use of cost-effectiveness analysis in education. *Journal of Research on Educational Effectiveness* 8(3): 400–418. https://doi.org/10.1080/19345747.2014.915604.
Mayer, S.E., Kalil, A., Oreopoulos, P. and Gallegos, S. (2018) Using behavioral insights to increase parental engagement: the parents and children together intervention. *Journal of Human Resources*, April, 0617-8835R. https://doi.org/10.3368/jhr.54.4.0617.8835R.
Nandrup, A.B. and Beuchert-Pedersen, L.V. (2018) The Danish national tests at a glance. *Nationaloekonomisk Tidsskrift* 2018(2):1–37.
National Academies of Sciences, Engineering, and Medicine (2016) *Parenting Matters: Supporting Parents of Children Ages 0-8*. Vivian L. Gadsden, Morgan Ford, and Heather Breiner (eds.). Washington, DC: National Academies Press. https://www.nap.edu/catalog/21868/parenting-matters-supporting-parents-of-children-ages-0-8.
National Reading Panel (2000) Report of the national reading panel: teaching children to read. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf.
Nøhr, K., Dalsgaard, C., Meldgaard, J., Bækgaard, M. and Kloppenborg, H.S. (2012) Sådan Passer Kommunerne Børn. Kommunernes Dagtilbudsstruktur 2004-2011. https://www.vive.dk/da/udgivelser/saadan-passer-kommunerne-boern-9151/ (last accessed: November 28, 2020).
NOSOSKO (2014) *Social tryghed i de nordiske lande 2012/13. Nordisk Socialstatistik Komité. Version 57:2014*, http://norden.diva-portal.org/smash/get/diva2:774222/FULLTEXT02.pdf (last accessed: November 28, 2020).
OECD (2016) Education at a Glance 2016.
Panduro, B. (2017) Klassekvotienter Og Andel Elever i Privatskoler i Kommunerne. Notat. https://www.vive.dk/da/udgivelser/klassekvotienter-og-andel-elever-i-privatskoler-i-kommunerne-8370/.
Phillips, D.A., Lipsey, M.W., Dodge, K.A., Haskins, R., Bassok, D., Burchinal, M.R., Duncan, G.J., Dynarski, M., Magnuson, K.A. and Weiland, C. (2017) Puzzling it out: the current state of scientific knowledge

on pre-kindergarten effects—a consensus statement. *Issues in Pre-Kindergarten Programs and Policy*, 19–30.

Pianta, R.C., Paro, K.M.L. and Hamre, B.K. (2008) *Classroom Assessment Scoring System™: Manual K-3*. Baltimore, MD: Paul H Brookes Publishing https://psycnet.apa.org/record/2007-18799-000.

Pomerantz, E.M., Moorman, E.A. and Litwack, S.D. (2007) The how, whom, and why of parents' involvement in children's academic lives: more is not always better. *Review of Educational Research* 77(3): 373–410. https://doi.org/10.3102/003465430305567.

Rea, D., and Burton, T. (2020) New evidence on the Heckman curve. *Journal of Economic Surveys* 31(2):241–262 https://doi.org/10.1111/joes.12353.

Rosholm, M., Hvidman, C., Nielsen, S.A., Gumede, K., Nafziger, J., and Koch, A. (2020) Turboforløb til ikke-uddannelsesparate elever i 8. klasse: Afsluttende rapport. Trygfondens Centre for Child Research. https://childresearch.au.dk/publikationer/nyhed/artikel/turboforloeb-til-ikke-uddannelsesparate-elever-i-8-klasse/ (last accessed: November 28 2020).

Slot, P.L. (2018) *Structural characteristics and process quality in early childhood education and care: a literature review*. OECD Education Working Papers 176. Paris: OECD Publishing. https://www.oecd-ilibrary.org/education/structural-characteristics-and-process-quality-in-early-childhood-education-and-care_edaf3793-en.

Slot, P.L., Bleses, D., Justice, L.M., Markussen-Brown, J. and Højen, A. (2018) Structural and process quality of Danish preschools: direct and indirect associations with children's growth in language and preliteracy skills. *Early Education and Development* 29(4): 581–602. https://doi.org/10.1080/10409289.2018.1452494.

Steuerle, E. and Jackson, L.M. (2016) *Advancing the Power of Economic Evidence to Inform Investments in Children, Youth, and Families*. Washington, DC: National Academies Press. https://www.nap.edu/catalog/23481/advancing-the-power-of-economic-evidence-to-inform-investments-in-children-youth-and-families.

Suskind, D.L., Leffel, K.R., Graf, E., Hernandez, M.W., Gunderson, E.A., Sapolich, S.G., Suskind, E., Leininger, L., Goldin-Meadow, S. and Levine, S.C. (2016) A parent-directed language intervention for children of low socioeconomic status: a randomized controlled pilot study. *Journal of Child Language* 43(2): 366–406. https://doi.org/10.1017/S0305000915000033.

Svarer, M., Rosholm, M., Havn, L., and Høeberg, L. (2014) Evaluering Af Mentorindsats Til Unge Uden Uddannelse Og Job (Evaluation of a mentor intervention for youth not in employment or education). https://star.dk/om-styrelsen/publikationer/2014/10/evaluering-af-mentorindsats-til-unge-uden-uddannelse-og-job/ (last accessed: November 28, 2020).

Swanson, E., Vaughn, S., Wanzek, J., Petscher, Y., Heckert, J., Cavanaugh, C., Kraft, G. and Tackett, K. (2011) A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. *Journal of Learning Disabilities* 44(3): 258–275. https://doi.org/10.1177/0022219410378444.

Sylva, K., Ereky-Stevens, K. and Aricescu, A.-M. (2015) Overview of European ECE curricula and curriculum template. http://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_1_European_ECEC_Curricula_and_Curriculum_Template.pdf (last accessed: November 28, 2020).

Thompson, S.G. and Sharp, S.J. (1999) Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 18(20): 2693–2708. https://doi.org/10.1002/(SICI)1097-0258(19991030)18:20<2693::AID-SIM235>3.0.CO;2-V.

Wang, A.H., Firmender, J.M., Power, J.R. and Byrnes, J.P. (2016) Understanding the program effectiveness of early mathematics interventions for prekindergarten and kindergarten environments: a meta-analytic review. *Early Education and Development* 27(5): 692–713. https://doi.org/10.1080/10409289.2016.1116343.

White, A., Prentice, S. and Perlman, M. (2015) The evidence base for early childhood education and care programme investment: what we know, what we don't know. *Evidence & Policy: A Journal of Research, Debate and Practice* 11(4): 529–46. https://doi.org/10.1332/174426415X14210818992588.

Yoon, K.S., Duncan, T., S.W.-Y. Lee, Scarloss, B. and Shapley, K.L. (2007) Reviewing the evidence on how teacher professional development affects student achievement. https://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=70.

York, B.N., Loeb, S. and Doss, C. (2019) One step at a time: the effects of an early literacy text messaging program for parents of preschoolers. *Journal of Human Resources* 54(3): 537–566. https://doi.org/10.3368/jhr.54.3.0517-8756R(last accessed: November 28, 2020).

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure A.1 Estimates from meta-regressions of various measures of effect size on age.
Table A1. Play and Learn intervention
Table A2. SPELL intervention
Table A3: LEAP intervention
Table A4. Suitcase intervention
Table A5. READ intervention
Table A6. Instruction time intervention
Table A7. Teacher's aide intervention
Table A8. TMTM intervention
Table A9. Turbo intervention
Table A10. Mentoring intervention