MEMORANDUM

No 16/2015

Kindergarten for All: Long-run Effects of a Universal Intervention

Nina Drange, Tarjei Havnes and Astrid M. J. Sandsør

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 15/15	Paolo G. Piacquadio *The Ethics of Intergenerational Risk*
No 14/15	Finn R. Førsund *Productivity Interpretations of the Farrell Efficiency Measures and the Malmquist Index and its Decomposition*
No 13/15	Christian N. Brinch, Erik Hernæs and Zhiyang Jia *Salience and Social Security Benefits*
No 12/15	Florian K. Diekert *Threatening Thresholds? The Effect of Disastrous Regime Shifts on the Cooperative and Non-cooperative Use of Environmental Goods and Services*
No 11/15	André K. Anundsen and Ragnar Nymoen *Did US Consumers 'Save for a Rainy Day' before the Great Recession?*
No 10/15	Finn Førsund *Economic Perspectives on DEA*
No 09/15	Andreas Kotsadam, Eivind Hammersmark Olsen, Carl Henrik Knutsen and Tore Wig *Mining and Local Corruption in Africa*
No 08/15	Eric Nævdal *Catastrophes and Expected Marginal Utility – How the Value of The Last Fish in a Lak is Infinity and Why We Shouldn't Care (Much)*
No 07/15	Niklas Jakobsson and Andreas Kotsadam *The Economics of Trafficking for Sexual Exploitation*
No 06/15	Geir B. Asheim and Stéphane Zuber *Evaluating Intergenerational Risks: Probability Adjusted Rank-Discounted Utilitarianism*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/

Kindergarten for all: Long run effects of a universal intervention*

Nina Drange† Tarjei Havnes‡ Astrid M. J. Sandsør§

Memo 16/2015-v1
(This version August 2015)

Abstract

Theory and evidence point towards particularly positive effects of high-quality child care for disadvantaged children. At the same time, disadvantaged families often sort out of existing programs. To counter differences in learning outcomes between children from different socioeconomic backgrounds, governments are pushing for universal child care. However, it is unclear how effective programs with universal participation may be at addressing the needs of disadvantaged children. We provide evidence on the long-run effect on schooling of mandating kindergarten at age 5–6. Our identifying variation comes from a reform that lowered school starting-age from 7 to 6 in Norway in 1997. The new program was designed as a low intensity kindergarten program, similar to voluntary child care programs available before mandating. Our precise DD estimates reveal hardly any effect, both overall, across subsamples, and over the grading distribution. A battery of specification checks support our empirical strategy.

Keywords: kindergarten, early childhood intervention, distributional effects, difference-in-differences, child care, child development
JEL codes: J13, H40, I28

*Thanks to Michael Baker, Torbjørn Hægeland, Magne Mogstad, Mari Rege, Marianne Simonsen and Kjetil Telle, as well as participants at a number of seminars and conferences. The project is part of the research activities at the ESOP center at the Department of Economics, University of Oslo. ESOP is supported by The Research Council of Norway. Havnes also gratefully acknowledges support from Statistics Norway and funding from The Research Council of Norway (S/194339 and S/212305). Drange also gratefully acknowledges support from The Research Council of Norway (160965/V10).

†Statistics Norway, nina.drange@ssb.no
‡University of Oslo, tarjei.havnes@econ.uio.no.
§University of Oslo, a.m.j.sandsor@econ.uio.no.

1 Introduction

Universally available child care of high quality can benefit child development, also in the long run (Almond and Currie, 2010). Returns are often found to be particularly high for children from disadvantaged families.[1] At the same time, children from disadvantaged families are underrepresented in existing programs. This sorting into the programs coupled with particularly large estimated benefits among disadvantaged children, suggests a potentially strong social gradient in expanding or mandating early childhood interventions (Barnett and Belfield, 2006). Indeed, in an effort to counter differences at school entry depending on social background, many countries are currently moving towards subsidized child care available for the general population.[2]

Policies and proposals promoting universal interventions in early childhood pose a challenge to the existing literature, which has reserved most of its attention for programs targeted at disadvantaged children. Existing studies on universally available programs typically reveal the impact on children from families with a strong preference for out-of-home care. For instance, Baker, Gruber, and Milligan (2008) and Havnes and Mogstad (2011) study the introduction of a universally available program but with actual enrollment being far from universal, while Gupta and Simonsen (2010) explicitly exploit rationing of child care for identification. Since both theory and evidence point towards important heterogeneity in the effects of early childhood interventions, it remains an open question how well the current evidence can inform about the impact of truly universal interventions. In particular, it is unclear how effective programs with universal participation may be at addressing the needs of disadvantaged children.

In the current paper, we provide evidence on the long-run effect on schooling of a reform that mandated kindergarten at age 5–6. We first consider the impact on children's school performance at the end of compulsory schooling at age 15–16. We also consider the impact on high school dropout and on enrollment in the academic track in upper secondary school measured at age 18 and 16, respectively. These are interesting in their own right, and help confront the concern of fading out of cognitive effects from early intervention programs, even when long-term effects on substantive outcomes may persist.[3] Our identifying variation comes from a 1997-reform in Norway that lowered school

[1] Havnes and Mogstad (2012) document large heterogeneity in the effects on adult outcomes from child care for 3–6 year old children in the late 1970s in Norway. Ludwig and Miller (2007) interpret the effects of the US Head Start as an upper bound because children are among the most disadvantaged. Further, effects found in the targeted Perry Preschool project (e.g. Karoly, Kilburn, and Cannon, 2005) are larger than what could plausibly be expected in the general population.

[2] For instance, US President Obama stated in his 2013 State of the Union Address that he wants to "make high-quality preschool available to every child in America". In Europe, the European Union Commission proclaims that early childhood education and care "is the essential foundation for successful lifelong learning, social integration, personal development and later employability" (European Union, 2011, p. 1).

[3] See for example Heckman, Moon, Pinto, Savelyev, and Yavitz (2010) or Heckman, Pinto, and Savelyev (2013).

starting age from seven to six. The new program for six year olds was designed as a low intensity kindergarten program, aimed at preparing children for school by learning through play, similar to early US kindergarten programs (Cascio, 2009). The goal of the new program was to counter differences in learning outcomes between children from different socioeconomic backgrounds. While disadvantaged children were thought to benefit most from kindergarten programs, they were strongly underrepresented in the existing voluntary programs prior to the reform.

Because the implementation of the reform was nationwide, the most direct assessment compares cohorts just young enough to be affected with cohorts just old enough not to be affected. An immediate objection to this strategy is that we may be confounding effects of the policy with unrelated cohort effects. To get around this issue, we take advantage of voluntary enrollment in child care prior to the reform. Since the new program for six year olds bears strong resemblance to kindergarten programs that were widely available prior to the reform, it should have little impact on children that would voluntarily enroll in such programs. This motivates a difference-in-differences (DD) approach where we compare outcomes before and after the implementation of the mandatory kindergarten reform, of children who enroll in voluntary kindergarten at age six (i.e. the control group) and children who do not enroll in voluntary kindergarten at age six (i.e. the treatment group). Because voluntary enrollment at age six is unobserved by definition after the reform, we use enrollment in child care at age five to determine treatment. This should be a good proxy since children who are enrolled at age five are almost universally enrolled at age six.

Results reveal that the program had little impact on affected children. In our baseline estimation, the precisely estimated effect on the child's school performance is negative but below 2 % of a standard deviation. Meanwhile, we find a modest increase in high school dropout rates, and no impact on academic tracking in upper secondary school. These results are robust to including or excluding a large set of observable characteristics, as well as a battery of specification checks confronting the key identifying assumption of common trends in treatment and comparison groups. Importantly, we find no evidence of a separate effect of the reform on our comparison group that may attenuate effects on the treatment group, nor of a delayed effect of the reform on later cohorts. We also find no evidence of important heterogeneity, when we look across subsamples reflecting the child's background and home environment, across different segments of the grading distribution, or across school subjects where we may expect children to benefit from different types of skills.

To help interpret our estimates, we take a close look at the contents of the program, which was specifically intended to be play-oriented, with little focus on specific learning activites. As a comparison, the program appears to be quite similar in content to the early US kindergarten programs, as its focus was more on children's social development

than on academic training, though the compulsory nature of the Norwegian program is an important difference.[4] The program also seems comparable to the US Head Start program, with its low intensity educational content, as well as similar costs and contents.[5] While the program we study served the entire population of 5–6 year olds, however, Head Start serves children 3–5 years old and is targeted at poor families.

Our paper contributes to the rapidly increasing literature on how early childhood interventions in general, and kindergarten programs in particular, can promote the formation of skills in children.[6] This literature is divided in two distinct branches, one focussed on targeted programs, and another focussed on universal programs available to the general population. While studies of targeted programs often find positive effects,[7] the literature on universal programs is smaller and findings are mixed.[8] Perhaps as a consequence, the discussion on child care policies is based largely on the targeted literature and descriptive evidence, even when the policies considered are universal.

There are several reasons why effects from programs targeted at disadvantaged children could differ importantly from more universal programs, as discussed by Baker (2011). First, the effect of such programs is related to the alternative mode of care had the programs not been in place. Since disadvantaged children would be expected to have poorer alternatives, they likely have more to gain from interventions (Knudsen, Heckman, Cameron, and Shonkoff, 2006). Second, targeted interventions are often quite intensive, sometimes including home visits, nutritional advice and several years of daily activities. In comparison, a program serving a large part of the population will necessarily have to provide a less intensive intervention. This might produce effects from large-scale and

[4]See Cascio (2009) for details and discussion on development of the US kindergarten program, and Norwegian Ministry of Education (2010) on the Norwegian program.

[5]See Deming (2009) for details and discussion on Head Start.

[6]For recent reviews, see Almond and Currie (2010), Ruhm and Waldfogel (2012), or Baker (2011). Our paper also relates to the literature on early enrollment into formal schooling (see e.g. Leuven, Lindahl, Oosterbeek, and Webbink (2010) or Black, Devereux, and Salvanes (2011) for an overview). An important issue in this literature has been to resolve the collinearity of age at test and age at school start. This is not an issue in our case, since age at test is both common across treatment groups and unaffected by the reform. However, the literature on child care and early childhood interventions may in general be said to face a similar collinearity between age at program start and years of enrollment. As in the rest of the literature, we estimate the combined effect of an additional year in kindergarten and lower age of entry.

[7]The Perry Preschool and Abecedarian programs are examples of targeted randomized programs (see Barnett (1995) and Karoly, Kilburn, and Cannon (2005) for surveys of the literature), while the US Head Start program provides an example of a targeted non-randomized program (see e.g. Currie (2001) or McKey, Condelli, Ganson, Barrett, McConkey, and Plantz (1985) for a review of the findings).

[8]Several studies from Canada show a negative impact on a variety of child outcomes (Baker, Gruber, and Milligan, 2008; Lefebvre and Merrigan, 2008b; DeCicca and Smith, 2013), while Cascio (2009) and Gupta and Simonsen (2010) find essentially no impact from child care programs in the US and Denmark, respectively. In contrast, positive impacts on long run outcomes are found from child care programs in several countries, including the US (Fitzpatrick, 2008), Uruguay (Berlinski, Galiani, and Manacorda, 2008), Norway (Havnes and Mogstad, 2011), Germany (Dustmann, Raute, and Schonberg, 2013; Felfe and Lalive, 2013), and Spain (Felfe, Nollenberger, and Rodríguez-Planas, 2012). Also, while the picture is somewhat mixed, the most robust evidence on the US Head Start program tends to show positive effects on long-run outcomes such as high school dropout, college attendance and crime (Currie and Thomas, 1995; Garces, Thomas, and Currie, 2000; Ludwig and Miller, 2007; Deming, 2009).

universal programs that differ substantially from the effects of intensive small-scale interventions.

The current paper contributes to the literature on universal child care programs in two distinct ways. First, since kindergarten is not rationed prior to mandating, the estimated effect should derive from the particular group of children that do not voluntarily enroll. Our study therefore provides a rare opportunity to learn about the group of never-takers, to use the terminology of Imbens and Angrist (1994). This is of particular interest since these families may have quite different characteristics compared to families that select into child care voluntarily, many of which may be unobserved. If so, then existing estimates may tell us little about the potential effect of child care among these children. Second, while the program we study is universal, the reform may be viewed as targeted since affected children come disproportionately from disadvantaged families. Our results may therefore shed light on how a universal low intensity program can improve outcomes among the disadvantaged. That is, can the positive effects for the disadvantaged, often seen from targeted interventions, be reproduced in a universal program? Our evidence suggests that this is not the case. This is true, even though the estimates likely reflect shifts mostly from parental care, rather than informal care.

Our results differ from some previous studies that have found positive effects of child care, particularly for low income children. One reason may be that the larger operations and broader scope involved in a universal compared to a targeted program may come with particular challenges, for instance by making it harder to see children's needs and to tailor activities to these needs. This might suggest that the unstructured child-centered approach to instruction (Stipek, Feiler, Daniels, and Milburn, 1995), which has been a hallmark of low intensity child care programs, may be less suitable in a universal program. An alternative interpretation is that parents who chose not to send their children to early education when such a program was available and affordable, may have done so partly because they expected little benefits to their children. In any case, while early childhood investments through subsidized child care can be an important tool in facilitating equal opportunities, our evidence emphasizes that this is hardly automatic, and that the structuring of the program and its content could be key to generating the intended benefits.

The paper proceeds as follows. We first discuss the institutional background for the 1997-reform in Section 2. Section 3 describes our data and gives descriptive statistics while Section 4 discusses our empirical approach. Section 5 then presents our main results, before Section 6 presents a battery of specification checks and investigates potential mechanisms. Section 7 concludes.

2 Background

Until 1997, Norwegian children started school in August the year they turned seven. This was late compared to children in most western countries.[9] At the same time, slots in child care institutions were widely available following a child care reform in 1975. In 1996, 89 % of non-immigrant families enrolled their six year olds in a kindergarten program.[10] However, from the mid-1980s, there was widespread worry that children entered school on different footings, depending on their socioeconomic background.

Figure 1 shows the strong social gradient in school performance and kindergarten enrollment.[11] In Panel (a), we draw the average grade of students at exams administered at the end of compulsory school, in the deciles of family income at age five. The figure shows a strong positive relationship between the two: On average, children in the lowest decile, with family income of about USD 16,000, receive a grade of less than 3.5, while children in the upper decile, with family income of about USD 170,000, receive a grade of almost 4.5.[12] This difference in exam performance is equivalent to a difference of just under one standard deviation, comparable to the 90-10 income achievement gaps in the US reported by Reardon (2011).

At the same time, children in lower deciles have a much lower probability of being enrolled in kindergarten. Panel (b) of Figure 1 shows that enrollment in kindergarten at age five among children in the lowest decile of family income is just over 50 %, compared to over 90 % for deciles 6–10. This serves on the one hand to illustrate the political background for the reform discussed above, and on the other hand to show that the children that were affected by the reform should come disproportionately from low-income familes.

While children enrolled in formal child care were offered school preparation in kindergarten groups within their child care center, this was not available for children not enrolled in formal child care.[13] On this background, a proposal to lower the mandatory school starting age from seven to six was widely discussed. Compulsory programs for six year olds, as opposed to voluntary kindergarten programs, would expose all children to the same educational program, and was argued to counter differences in learning outcomes between children from different socioeconomic backgrounds. A reform was finally pro-

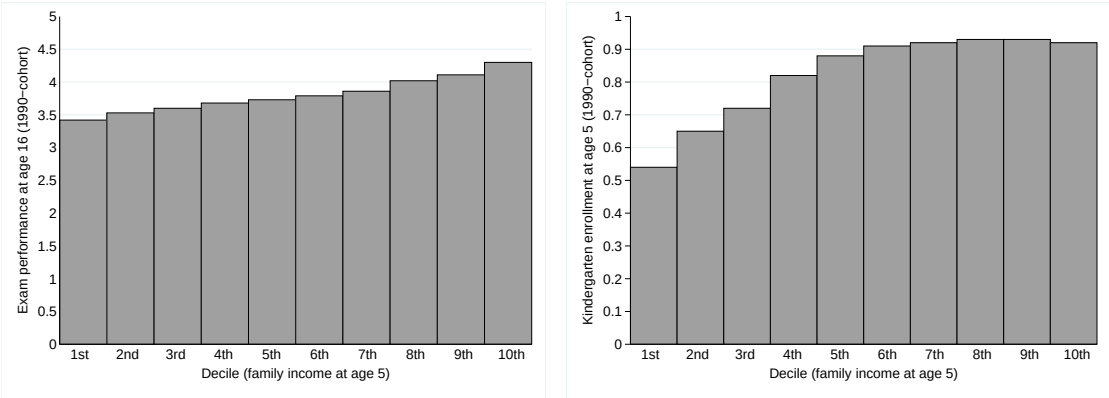
[9] For instance, school starting age in Germany, France and the US was six, while England had a starting age of five.

[10] For simplicity, we use age a to refer to the year the child turns a years old in the following.

[11] For details, see Table A1 in Appendix A.

[12] Throughout, we refer to 2011-USD adjusted using the consumer price index, USD/NOK = 6.

[13] Voluntary programs for six year olds were allowed on school grounds from 1991, managed by kindergarten teachers (Norwegian Ministry of Education, 1990-91, Ot.prp. nr. 57). Government support was the same as for six year olds enrolled in kindergarten, and the parental copayment and educational content of the program were essentially the same as that offered in regular child care institutions. In the remainder, we do not distinguish between the two, as we cannot identify in which program a particular child was enrolled.



(a) School performance (b) Enrollment in kindergarten

Figure 1: Social gradient in school performance and enrollment in kindergarten among children born in 1990.

Note: Family income is measured in 1996, when the child is five years old, adjusted for CPI-growth, and converted to USD using $USD/NOK = 6$. Average exam grade and enrollment in kindergarten refers to the mean among children from families with income in each decile of the distribution of family income. Data descriptions and variable definitions are found in Section 3.

posed in a government White Paper published in the spring of 1993 (Norwegian Ministry of Education, 1992-93), and passed the Norwegian Parliament in May 1994 (Norwegian Ministry of Education, 1993-94). The reform was implemented in August 1997, at the start of the 1997–1998 school year. The first children affected were those born in 1991, who started school in August 1997, the year they turned six years old.

Note that the cutoff for school starting age in Norway is January 1st. In Norway schools employ strict enrollment rules, and nearly all children start school the year they turn the school starting age. Any exemption from this rule requires a formal application from the parents which then has to be approved by specialists and decided upon by the local government.

Structural content. The group size was capped at 20 children, under supervision of two kindergarten teachers, identical to the previous child care programs. The government cost of the program was about 8,800 USD per child per year, similar to the cost of the previous child care programs which cost about 9,700 USD per child per year. There was no parental copayment for the core four hour program, but the voluntary after-school program required a copayment of about 170 USD per month on average. This was a reduction from the copayment for center-based kindergarten, which ranged from about 290 to 630 USD per month, depending on income. The reduced cost could imply that the reform caused an economic windfall for children who would attend kindergarten in the absence of the reform, with possibly positive effects on their school performance. We investigate the potential for income effects extensively below, finding little cause for concern.

Educational content. The new program was aimed at combining the best of school and

kindergarten traditions. These were grounded in the tradition of social pedagogy that dominates child care practices in Norway since the 1970s.[14] Learning through play was stated as essential, and formal learning was given little credence (Norwegian Ministry of Education, 1992-93). The curriculum specifically stated that *"The first year is to have a distinct kindergarten character, and one has to emphasize learning through play and age-mixed activities throughout elementary school [years 1–4]"* (Norwegian Ministry of Education, 1996a).

Note that the curriculum for all grades in primary school was revised, and implemented for grades 1, 2, 5 and 8 in 1997, grades 3, 6 and 9 in 1998 and the remaining grades in 1999 (Norwegian Ministry of Education, 1996b). Children in our main sample, born 1990–1991, were therefore subject to the same, new curriculum throughout primary school.

In the new mandatory kindergarten program, the minimum requirement was one teacher or kindergarten teacher for every 18 children. By comparison, the minimum requirement for pedagogical staff in child care centers for six year olds was one per 14–18 children. Beyond this, the municipalities should themselves judge the need for further staff, but they had to secure sufficient care for the children (Norwegian Ministry of Child and Family Affairs, 1995).[15] Similarly, in the new kindergarten program integrated into schools, in addition to the kindergarten teacher, assistants were hired depending on the size of the group. Starting in 1991, both kindergarten teachers and school teachers were allowed to work with six year olds. This was part of the gradual implementation of voluntary programs for six year olds on school grounds. After the 1997-reform was passed in parliament, both kindergarten teachers and teachers could work in first grade, while some continued education was needed for kindergarten teachers to work in grades 2–4. The reform explicitly aimed to bring together the best of school and kindergarten traditions, and bringing kindergarten teachers into the first grades of school was part of this goal. The transfer of kindergarten teachers from child care centers into schools, meant that six year olds in 1997 were likely to experience a similar pedagogical environment to the one they would have experienced in absence of the reform.

The teaching requirement for the new first graders was set to 20 hours per week (Norwegian Ministry of Education, 1992-93). To ensure the care of six year olds during normal work hours but outside school hours, the government also expanded the access to the pre-existing after-school program, available for children in grades 1–4. After-

[14]The social pedagogy tradition for early education has been especially influential in the Nordic countries and Central-Europe. In contrast, a so-called pre-primary pedagogic approach to early education has dominated many English and French-speaking countries, favoring formal learning processes to meet explicit standards for what children should know and be able to do before they start school.

[15]The head kindergarten teacher was responsible for planning, observing, collaborating and evaluating the work being done, under the requirements specified in the regulations for subsidized child care. Teachers typically worked closely with one or two assistants, and were responsible for the educational programs in separate groups of 6–18 children and for day-to-day interaction with parents. The kindergarten teacher education is a college degree, while there are no educational requirements for assistants.

school programs were available throughout our period of study, were subject to similar requirements as regular child care providers, and were usually situated on school premises. The programs offered free play under the supervision of non-qualified adults, with no educational content.

Other reforms. We may worry that there were other reforms that could have affected our cohorts differently. However, the closest reform in primary education prior to the 1997-reform was implemented in 1986, while there were no additional reforms until the start of the school year 2007–2008. This ensures that the 1990 and 1991 cohorts completed their entire compulsory schooling with the national curriculum introduced in 1997. A nationwide cash-for-care reform was implemented in 1998, and expanded in 1999, paying families with children below two years old (from 1998) and three years old (from 1999) that did not utilize subsidized child care a substantial monthly cash allowance.[16] While this reform did not affect the children in our sample directly, it could have had an effect on younger siblings and therefore an indirect effect on the children in our sample (Bettinger, Hægeland, and Rege, 2014). However, the impact of the cash-for-care reform does not differ between children born in 1990 and 1991, which constitute our baseline estimation sample.[17] This suggests that the cash-for-care reform does not pose a threat to our empirical strategy.

3 Data

Dataset and variables. Our data are based on administrative registers from Statistics Norway. Specifically, we use a rich longitudinal database which covers every resident from 1992 to 2007. It contains individual demographic information (e.g. sex, age, immigrant status, marital status, number of children), socioeconomic data (e.g. years of education, income, employment status), and geographic identifiers for municipality of residence. Information on school performance, educational attainment and school enrollment for every individual is based on annual reports from Norwegian educational establishments. Income and employment data are collected from tax records and other administrative registers. The household information is from the Central Population Register, which is updated annually by the local population registries and verified by the Norwegian Tax Authority. We also have access to registry data on municipal child care coverage reported by the child care institutions themselves. The reliability of Norwegian register data is considered to be very good, as is documented by the fact that they received the highest rating in a data quality assessment prepared for the OECD by Atkinson, Rainwater, and Smeeding

[16] See Schøne (2004) or Drange and Rege (2013) for a detailed description of the cash-for-care reform.

[17] To investigate this directly, we estimated our baseline DD model in equation (1) using as dependent variable a dummy equal to one if the child has a younger sibling born 1996 or later (i.e. partly or fully eligible for the subsidy) and zero otherwise. The estimate is almost exactly zero (0.004, $SE = 0.007$).

(1995).

Estimation sample. We start with the universe of children born 1990–1991, who reside in Norway the year they turn five years old and who graduate from lower secondary school in 2005–2006. We then restrict our sample to children born to native-born parents, constituting about 96 % of the population, in order both to focus our study on the effect of mandating, and to sidestep problems of comparability between native and immigrant children. Our paper does not, therefore, speak to the debate on early interventions to provide language training among non-native speakers. We also exclude a handful of children with missing values on our dependent variable. Rather than exclude children with missing values on control variables, we construct dummy variables for missing and include these in our regressions. Our main sample then consists of 111,397 individuals, of which just over 16 % are in the treatment group. In our extended sample, we consider the analogous population of children born 1988–1992.

Measuring kindergarten enrollment. There is, unfortunately, no register of individual kindergarten or child care enrollment. However, parents may claim the cost of child care as a deduction on their earned income. To identify whether a child is enrolled in kindergarten, we therefore follow Black, Devereux, Løken, and Salvanes (2014) in using a binary variable equal to one if the child's parents claimed a tax deduction for child care for the year the child turned five years old. Of course, if a child has siblings, we cannot verify which of the children the deduction is claimed for (if not all). To get around this issue, we assume that child care enrollment is monotonous in age, such that older children are in child care whenever younger children are in child care. This ensures that at least the older child in child care age is enrolled whenever the parents claim the deduction. We might worry that low income households did not take the tax deduction even if eligible. However, payment for child care for each child is often reported directly from the child care provider to the tax authorities. Indeed, since 1994 all public child care institutions have been required to do so.[18] To verify that our measure of kindergarten enrollment is sound, we have calculated the municipal enrollment implied by this measure. We then compare these numbers to the actual enrollment from administrative registers, reported by the child care institutions themselves. The correspondence is very high, with a correlation of about 0.94 (see also Figure A1 in the appendix).

Measuring school performance. Our main outcome is an average of grades on nationally administered end-of-school exams. At graduation from compulsory school, children are tested on two or three exams in randomly drawn theoretical subjects—one or two written exams and one oral exam. The written exam is uniform across the country and provided

[18] While the vast majority of deductions claimed are for child care costs, some other costs may also be claimed under the same statute, e.g. outlays for support of children with disabilities or with other special needs.

by the Central Education Authority, and is corrected by external evaluators who typically grade exams from several schools simultaneously. The oral exam is also evaluated by an external examiner, and takes place at the school at which the child is enrolled. Grades are awarded on a scale from one to six, where six indicates excellence and one indicates very little competence (in our estimations, we standardize grades to mean zero and standard deviation one). Grade retention is illegal, hence all children are allowed to graduate regardless of their grades. In addition, teachers assign each child grades in 12–13 subjects, based on performance throughout the year. There are nine theoretical subjects and four practical subjects.[19]

Measuring high school dropout. In our data we can observe whether an individual is enrolled in education and how many years they have successfully completed. We define high school dropout as either not being enrolled in education, or not being on year for age in graduating year. That is, we code high school dropout as not being registered in your 13th year of education in the fall of the year you turn 18.[20] Note that this definition is somewhat strict, since it requires that children are not delayed.

Measuring academic track. In Norway, children are first tracked when they start upper secondary school. There are two main tracks (which are divided into 13 more specialized sub-disciplines): The academic track which is required for entry into university and college studies, and the vocational track which qualifies for a practical occupation. To consider whether the reform had an impact on academic tracking, we use a dummy equal to one if the child started on the academic track in the year following graduation from compulsory schooling, i.e. at age 16. Note that enrollment in upper secondary school is almost complete in these cohorts, with about 94 % of children enrolling in one of the two tracks. We have also estimated the effect on the decision to enroll, finding no impact of the reform. If a child does not enroll in upper secondary school the year following graduation, he or she is excluded from these estimations.

Covariates. To account for possible observable changes in composition between years, we include a number of child and parent characteristics in our analysis, measured when the child is five years old. Child characteristics include municipality of residence, gender, number of siblings, and finally a dummy measuring if the child lived in a densely populated area. Background characteristics include a dummy measuring if the mother/father worked full time, a dummy for whether the mother/father completed high school and a dummy indicating if the mother/father finished a college education. In addition, we include a

[19]Theoretical subjects are written and oral Norwegian, written and oral English, mathematics, nature and science, social science, and religion. Practical subjects are home economics, physical education, music, and arts and crafts.

[20]Final graduation from high school should occur the year they turn 19 in the academic track and the year they turn 20 in the vocational track. This information is not yet available in the data.

dummy capturing missing observations on mothers/fathers education. Further, we include a dummy that captures whether the mother/father was younger than 22 when the child was born. We also include a dummy for having missing observations on either the mother or the father. If both parents are missing we exclude the observation. Finally, we include a dummy capturing if one or both parents received welfare benefits, a dummy measuring if the family was low income (defined as earnings below the 10th decile in the family income distribution in the cohort born in 1990), and a dummy capturing if the child lives with only one of its parents.

Descriptive statistics. Means of the outcome variables are presented in Figure 2 of Section 4, and are discussed there. In Table 1, we present characteristics for the entire sample in the first two columns, and differences between the two groups by cohort in the remaining columns. All covariates are measured when the child is five years old. We see no evidence of changes over time for characteristics of children or their parents between the treated and the comparison group. As discussed above, it is clear, however, that the treated children to a greater extent come from families with younger and less educated parents, and are more likely to belong to a family on welfare and/or to a single parent family. They are also overrepresented in the low income family group. This suggests that the children in our treatment group have a more disadvantaged background, in line with the expressed motivation of the policy-maker (Norwegian Ministry of Education, 1992-93).

4 Empirical strategy

Because the implementation of the reform was nationwide, the most direct assessment of how it affected children's long-term outcomes compares cohorts just young enough to be affected with cohorts just old enough not to be affected. An immediate objection to this strategy is that we may be confounding effects of the policy with unrelated cohort effects. To get around this issue, we exploit the temporal and spatial variation in pre-reform kindergarten enrollment in a difference-in-differences setup. Ideally, we want to compare the child outcomes before and after the implementation of the mandatory kindergarten reform of children who would enroll in voluntary kindergarten at age six (i.e. the control group) and children who would not enroll in voluntary kindergarten at age six (i.e. the treatment group). Our basic difference-in-differences (DD) model estimated by OLS, can then be expressed as

$$Y_{it} = \alpha_t + \gamma_1 Treated_i + \lambda Treated_i \times Post_t + X_{it}^{'}\beta + \epsilon_{it} \tag{1}$$

where i indexes child, t indexes cohort, $Post_t$ is a dummy equal to one if the child is affected by the reform (i.e. $t \geq 1991$) and zero otherwise, and $Treated_i$ is a dummy equal to one if the child is in the treatment group. Note that the cohort-specific constant

Table 1: Summary statistics

	Mean (SD)		**T − C, by cohort**				
	Treated (T)	*Comp. (C)*	*1988*	*1989*	*1990*	*1991*	*1992*
A. Child and family characteristics							
Female	0.49 (0.50)	0.49 (0.50)	0.00	0.00	0.00	0.01	0.01
1 sibling	0.40 (0.49)	0.51 (0.50)	-0.11	-0.10	-0.11	-0.12	-0.12
2 siblings	0.28 (0.45)	0.24 (0.43)	0.04	0.04	0.04	0.05	0.02
3 siblings +	0.12 (0.32)	0.06 (0.23)	0.05	0.06	0.06	0.07	0.06
Densely populated area	0.57 (0.49)	0.58 (0.49)	0.00	0.00	0.00	0.00	0.00
On welfare	0.19 (0.40)	0.08 (0.26)	0.13	0.13	0.12	0.14	0.16
Low income	0.08 (0.27)	0.00 (0.05)	0.08	0.08	0.07	0.08	0.09
Single parent	0.29 (0.46)	0.17 (0.38)	0.12	0.13	0.12	0.11	0.14
B. Mother characteristics							
Employed	0.18 (0.38)	0.70 (0.46)	-0.55	-0.54	-0.52	-0.53	-0.64
– full time	0.07 (0.26)	0.33 (0.47)	-0.26	-0.27	-0.26	-0.28	-0.32
High school	0.34 (0.47)	0.54 (0.50)	-0.21	-0.22	-0.20	-0.22	-0.26
College	0.13 (0.34)	0.28 (0.45)	-0.17	-0.16	-0.15	-0.16	-0.20
Young mother	0.21 (0.41)	0.13 (0.33)	0.10	0.10	0.08	0.07	0.11
C. Father characteristics							
Employed	0.63 (0.48)	0.73 (0.44)	-0.10	-0.10	-0.10	-0.10	-0.11
High school	0.49 (0.50)	0.60 (0.49)	-0.13	-0.12	-0.11	-0.10	-0.12
College	0.18 (0.39)	0.27 (0.44)	-0.10	-0.09	-0.09	-0.08	-0.10
Young father	0.07 (0.26)	0.05 (0.21)	0.04	0.04	0.03	0.03	0.05

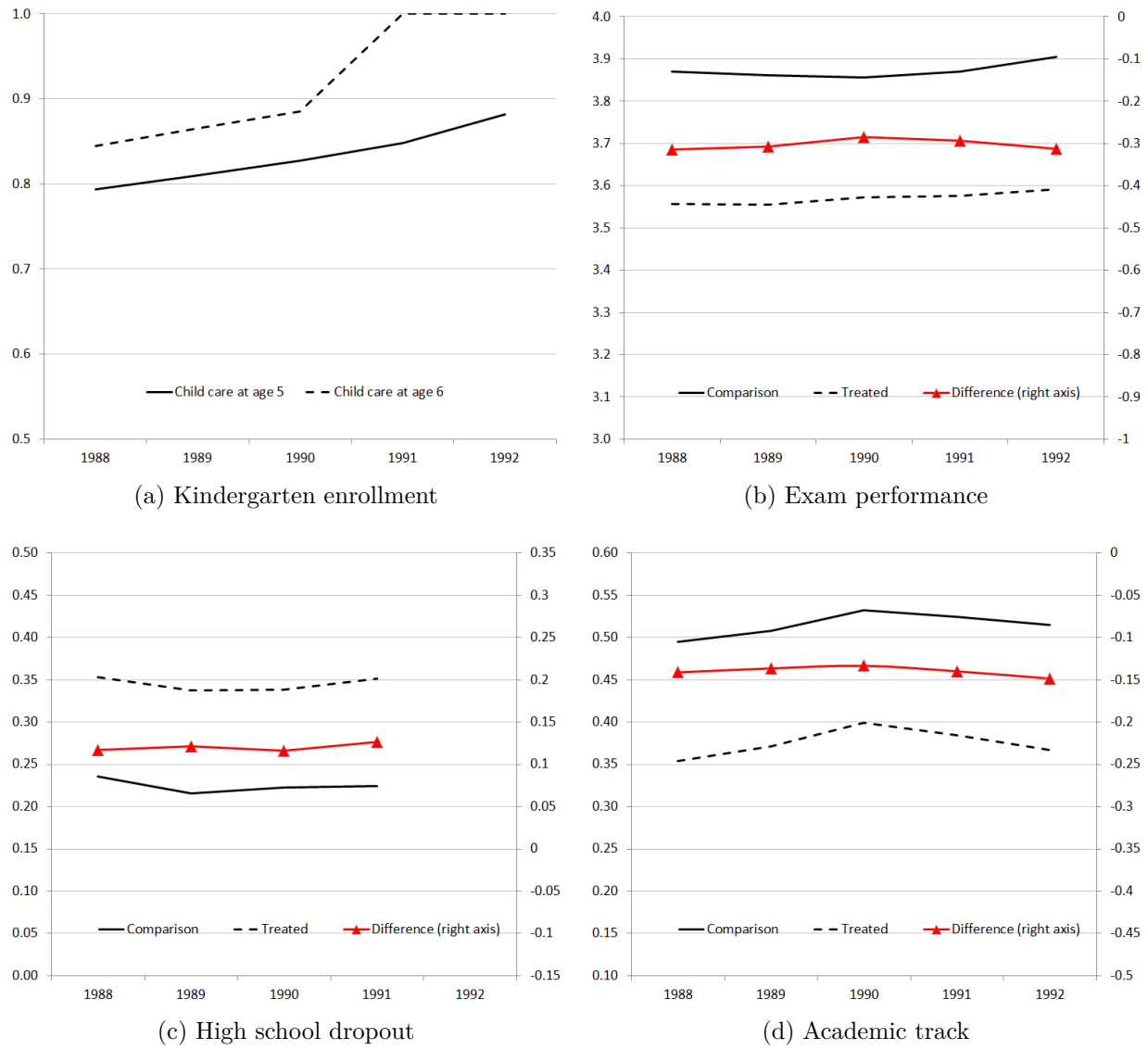
Note: The treatment group includes children whose parents did not report a tax deduction for child care expenses the year the child turned five. Outcome and control variables are defined in Section 3. Standard deviations are in parentheses.

term consumes the separate effect of the *Post*-dummy. We estimate the model with and without a large set of control variables for child and parental characteristics X_{it}, including the child's sex, the mother's and the father's age, years of education, and family size (see also Section 3). We also include municipality fixed-effects to capture time-invariant unobserved differences between children from different municipalities. All control variables are measured prior to the impact of the reform and standard errors are robust to heteroskedasticity.

In practice, whether a child would enroll in voluntary kindergarten cannot be observed for post-reform cohorts, since all children are enrolled at age six. To estimate equation (1), we therefore use enrollment in kindergarten at age five to determine treatment. This should be a good proxy since most children who are enrolled in kindergarten at age five are also enrolled at age six. That is, children who are enrolled in child care the year they turn five are placed in the control group, while children who are not enrolled in child care at age five are placed in the treatment group.

Panel (a) of Figure 2 displays the trend in kindergarten enrollment at age five and six

in our estimation sample. We note the close relationship between the two series over time in the pre-reform period, where the two lines are virtually parallel. This suggests that enrollment at age five captures the counterfactual evolution of enrollment at age six well. Furthermore, we note that there is no spike in the enrollment of five year olds following the reform, when children age six are no longer taking up places in child care centers. This suggests that there was no discernible rationing of kindergarten for these age groups in our period of study. Finally, we note that kindergarten enrollment in pre-reform years is around 84 %, giving a treatment group of about 16 % of the total sample.



(a) Kindergarten enrollment

(b) Exam performance

(c) High school dropout

(d) Academic track

Figure 2: Kindergarten enrollment and children's schooling outcomes by treatment for cohorts born 1988–1992.

Note: Vertical axes are scaled to approximately one standard deviation. High school dropout is not yet available for the 1992-cohort. Variables are defined in Section 3.

The validity of our DD strategy hinges on the assumption that the trend in school performance among children in the treatment group would have been the same as for

children in the control group, in the absence of the reform. As emphasized by Besley and Case (2000), this essentially assumes common time effects and no compositional changes between the treatment and control group. The richness of our registry data allows us to condition on a large set of observable characteristics, to investigate how changes in the composition of the groups may affect our estimates.

To investigate the time effects, Panels (b)–(d) of Figure 2 display mean outcomes of cohorts born 1988–1992 separate for the treatment and control group, and the difference between the two groups over time (on the right axis). The vertical axes are scaled to about one standard deviation in all the figures. The trends are quite flat and strikingly similar across the treatment and control group throughout the period. The similarity of the trend in the pre-reform period, supports the assumption of common time effects. That there is no jump in the treatment group from the 1991-cohort onwards, nor a divergence in the trends in the post-reform period, is first evidence that the reform had little impact on children's school performance.

One immediate objection to our empirical approach could be that the increasing trend in kindergarten enrollment generates a change in the composition of our treatment and comparison groups. We did not, however, see any evidence of a change in the composition on the observable characteristics of Table 1. Also, the increasing trend in kindergarten enrollment is constant before and after the reform. However, Figure 2 does not reveal any divergence in the school outcomes of the two groups in the pre-reform period, which would be expected if changes in the composition were associated with unobservable determinants of school performance. In Section 6, we also formally challenge our strategy both with a placebo reform in the pre-reform period and by including treatment-specific trends in the specification, neither of which give cause for concern.

Though Section 2 suggests little change in the contents, we could also worry that the new kindergarten program integrated in schools in fact was different from the former program, and thus could have had an effect also on children in the comparison group. We pay close attention to this in our robustness analysis provided in Section 6, finding no support for an effect on the comparison group. To further challenge the validity of our empirical specification, Section 6 also reports results from a series of specification checks.

The difference in enrollment between five and six year olds should not be a threat to the internal validity of our estimates. Higher enrollment at age six than at age five may, however, dilute the estimated treatment effect by misplacing some children in the treatment group who enroll in kindergarten only at age six. Our estimates may therefore be interpreted similar to intention-to-treat (ITT) estimates, and should be scaled in order to arrive at the average treatment effect on the treated. In 1990, 48 % of children who are not enrolled at age five are enrolled at age six, suggesting that only 52 % of the treatment group are in fact affected by the reform. In interpreting our results, we should bear this in mind, scaling the estimated effects by a factor of $1/.52 = 1.9$ to arrive at the average

effect on the treated (ATT).[21]

5 Empirical results

In this section, we first report estimated mean effects of mandating kindergarten on children's long-term schooling, before we investigate potential heterogeneity in the effects across subsamples and across the grading distribution. All specifications are estimated with municipality fixed effects to account for time-invariant differences between municipalities. To address concerns about compositional changes we have estimated the baseline model with and without the set of covariates capturing important child and parent characteristics. We stress that our estimates should be interpreted as ITT-effects, and should be scaled by about 1.9 to arrive at the ATT.

Mean effect. Table 2 reports our difference-in-differences estimates based on equation (1) from the sample of children born 1990–1991. In Panel A, we report the estimated effects on exam performance at the end of compulsory school, with and without the set of covariates. The estimates indicate that the reform had little effect on children's school performance, with a precisely estimated point estimate of about 1 % of a standard deviation. Excluding covariates in the second row of Panel A hardly moves the estimate. This indicates that there are no important compositional changes between the two cohorts, as expected from historical reports and descriptive statistics. Given the precision of the estimate and scaling for take-up, we can rule out effects above 3.3 % and below -7.1 % of a standard deviation at a confidence level of 5 %.

While studies of how early interventions affect child cognitive outcomes often find positive effects in the short run, these effects are often found to dissipate over time (see e.g. Knudsen, Heckman, Cameron, and Shonkoff, 2006). At the same time, persistent effects are often found on outcomes that may also reflect non-cognitive traits. In Panels B and C of Table 3, we consider effects on high school dropout rates and enrollment in the academic track in upper secondary school, where earlier studies have often found an improvement from early intervention programs. However, again we find little evidence of any substantial effect, whether or not we include covariates. Indeed, if anything, we find a small negative impact on children's schooling of mandating kindergarten, with a slight rise in high school dropout rates of 1.3 percentage points (from a pre-reform mean of about 33 % in the treatment group).

Heterogeneous effects. Though we find little support for an effect of mandating kindergarten on mean school performance, high school dropout or choice of academic track, we

[21] The opposite misclassification is almost completely absent: More than 97 % of children who are enrolled in kindergarten at age 5 are also enrolled at age 6.

Table 2: Mean effects on school performance, high school dropout rates and academic track in upper secondary school.

	Coeff	*SE*	*Mean [SD]*
A. School performance			
Baseline	-0.01	(0.014)	0 [1]
No covariates	-0.013	(0.016)	
B. High school dropout			
Baseline	0.013	(0.008)	0.33 [0.47]
No covariates	0.014	(0.008)	
C. Academic track			
Baseline	-0.008	(0.007)	0.40 [0.49]
No covariates	-0.009	(0.008)	

Note: $N = 111,397$ ($N = 107,707$ for academic track). Estimations are based on OLS on equation (1). The controls are listed in Table 1 and the dependent variables are defined in Sections 3 and 5. In Panel A, coefficients are standardized to the standard deviation of the dependent variable. Mean refers to pre-reform mean in the treatment group. Standard errors (SE) are robust for heteroskedasticity and all models include municipality fixed effects.

have already emphasized the general expectation of heterogeneous effects of early childhood interventions. A worry may therefore be that estimated mean effects mask large but offsetting effects among different groups of children. One concern might be that effects would be offsetting over the distribution of the outcome, e.g. beneficial in the lower parts of the grading distribution but negative in the upper parts. Another concern might be that effects differ depending on characteristics of the child, the family or the local school, either due to heterogeneous responses, or due to different exposure to the treatment (i.e. different take-up rates).

To address the first concern, we have estimated the impact of the reform on school performance at every point in the grading distribution (Table A2). Specifically, we estimate equation (1) over the sample of children born 1990–1991, where the outcome variable is a dummy equal to one if the child's school performance at end of compulsory school is above the given percentile, and zero otherwise. Estimates should then be interpreted as the percentage point change following the reform in the probability of performing above a given percentile for a child in the treatment group compared to a child in the comparison group.[22] Estimates for selected percentiles covering the bottom, the middle and the top of the distribution are reported in Table A2 in the appendix.[23] Results show that there is essentially no heterogeneity across the grading distribution.

To address the second concern, we have estimated separate reform effects for all outcomes in subsamples defined from a number of background characteristics. The estimates are reported in Table 3. To facilitate comparison of estimates across subsamples, we also report the mean outcome among treated children from the pre-reform cohort, the share of

[22]This procedure is essentially the first step in the RIF-procedure proposed by Firpo, Fortin, and Lemieux (2009) and applied to the DID-framework by Havnes and Mogstad (2012).

[23]We have estimated effects at all percentiles, which yields the same picture.

treated children and the take-up rate in the subsample.[24] Note that there is no systematic relationship between the take-up and the estimated coefficient (see also Figure A2 in the appendix).

The results give little indication of important heterogeneity in the impact of mandatory kindergarten, which is estimated to be very small. However, lower sample size implies less precision, and some patterns in the point estimates may warrant comment. First, girls seem to benefit more than boys, in line with what is often found in the literature on cognitive impact of early childhood interventions (Anderson, 2008). Second, though very imprecise, we also note a pattern that children that initially perform well, as measured by mean exam grade pre-reform, may tend to receive the most harm from mandatory kindergarten. In particular, children of higher educated families on average do experience a modest negative effect of the reform. Though estimates are too imprecise to provide much confidence, this could be interpreted as an indication that parents with high levels of human capital provide a good alternative to kindergarten, in line with Cunha, Heckman, and Schennach (2010) and estimates in Havnes and Mogstad (2012).

6 Specification checks

To improve our confidence in the estimates, we now challenge our empirical approach in different ways. First, we confront the key identifying assumption of our empirical strategy, namely the common trend assumption. Second, we consider whether there may be a separate effect of the reform on our comparison group that may attenuate effects on our treatment group. Third, we investigate whether there might be a delayed effect of the reform on later cohorts, before we look at how the bedding-down of the new curriculum could threaten our estimates. Finally, we consider some alternative and less aggregated school outcomes to understand whether there may be effects on some particular sets of skills that are washed out in our aggregated measure, and investigate whether the reform may have had an effect on the labor supply of mothers. For brevity, we focus on school performance. Results are similar for high school dropout and enrollment in the academic track (cf. Tables A3 and A4 in the appendix).

Common trend assumption. The primary threat in DD estimation is that the change in the observed outcome in the comparison group in the absence of the reform differs from the change in the potential outcome of the treatment group in the absence of the reform. One example could be anticipation effects, e.g. knowing that kindergarten would be free from age six, children born in 1991 might be more likely to be enrolled at age five. To investigate the common trend assumption, we start by considering a placebo reform,

[24] As discussed in Section 4, take-up is defined as the probability that a child who does not enroll in child care at age five, and is therefore in our treatment group, does not enroll in kindergarten at age six, and should therefore be affected by the reform.

Table 3: Heterogeneous responses

	School performance			High school dropout			Academic track			Treated	Take-up
	Coeff	*SE*	*Mean*	*Coeff*	*SE*	*Mean*	*Coeff*	*SE*	*Mean*		
A. Child and family characteristics											
Girls	0.012	(0.020)	3.73	0.008	(0.009)	0.22	0.000	(0.011)	0.58	0.16	0.52
Boys	-0.032	(0.021)	3.44	0.017	(0.010)	0.26	-0.015	(0.011)	0.44	0.16	0.52
Single parent	-0.035	(0.029)	3.31	0.017	(0.015)	0.39	0.002	(0.015)	0.40	0.24	0.56
Low income	-0.008	(0.043)	3.30	-0.006	(0.022)	0.45	0.013	(0.022)	0.34	0.45	0.65
Young mother	-0.035	(0.034)	3.31	0.019	(0.018)	0.37	0.002	(0.018)	0.35	0.24	0.53
On welfare	-0.006	(0.038)	3.22	-0.001	(0.020)	0.43	0.014	(0.019)	0.34	0.35	0.63
B. Mother characteristics											
No high school	0.003	(0.019)	3.35	0.020	(0.010)	0.41	-0.007	(0.010)	0.33	0.22	0.57
High school	-0.021	(0.023)	4.00	-0.003	(0.010)	0.21	-0.012	(0.012)	0.59	0.11	0.40
College	-0.047	(0.036)	4.28	0.014	(0.014)	0.14	0.003	(0.018)	0.72	0.08	0.32
C. Father characteristics											
No high school	0.016	(0.021)	3.31	0.016	(0.011)	0.43	-0.002	(0.011)	0.29	0.20	0.55
High school	-0.025	(0.020)	3.84	0.005	(0.008)	0.25	-0.013	(0.010)	0.51	0.14	0.48
College	-0.021	(0.032)	4.18	0.007	(0.012)	0.17	-0.029	(0.016)	0.70	0.12	0.47

Note: $N = 111,397$ ($N = 107,707$ for academic track). Estimations are based on OLS on equation (1). The subsamples are defined in Section 3. The controls are listed in Table 1 and the dependent variables are defined in Section 3. Standard errors (SE) are robust for heteroskedasticity and all models include municipality fixed effects.

pretending that the reform was implemented in the pre-reform period. The first row of Panel A in Table 4 reports the estimate from equation (1) estimated over the sample of children born 1989–1990, where $Post_t$ is redefined to be equal to one for children born in 1990 and zero otherwise. A significant estimate in this specification would put in doubt our identifying assumption. However, the estimate is almost precisely zero and nowhere near statistical significance.

Allowing treatment and comparison groups to follow separate trends is another way to challenge the common trend assumption. By extrapolating pre-reform trends into the post-reform period, we essentially restrict our estimates to reflect how outcomes deviate from the pre-reform trajectory. As emphasized by Besley and Case (2000), this is a simple yet potentially powerful test, which can often kill otherwise large and significant DD estimates.

To allow estimation of a trend, we extend the estimation sample to the start of our data series in 1988, and include the 1992-cohort, which is the last cohort that we can confidently use due to the cash-for-care reform in 1998. We then set $Post_t = 1$ for $t = 1991$ and $t = 1992$, and zero otherwise. For a correct comparison, results on this sample using our main regression, equation (1), are reported in the first row of Panel B. Estimates conform to those in the baseline. In row 2 of Panel B, we include a linear treatment-specific trend, while row 3 includes a second-order polynomial treatment-specific trend. Both specifications confirm the baseline estimates of essentially no effect of introducing mandatory kindergarten.

As an alternative, we can instead follow Duflo (2001) in allowing children to follow different trends depending on underlying characteristics. Specifically, we first estimate equation (1) including a linear trend interacted with baseline covariates.[25] We then relax the assumption of a linear trend, interacting instead the baseline covariates with the cohort fixed effects. Results are reported in Panel C, again confirming our baseline estimate of hardly any impact of the mandatory kindergarten reform on children's school performance at end of compulsory schooling.

Effect on comparison group. The transfer of kindergarten teachers from child care centers into schools, meant that six year olds in 1997 were likely to experience a similar pedagogical environment to the one they would have experienced in absence of the reform. Even so, the potential for changes in the content of the program coinciding with the reform raises the concern that the reform could have had an effect also on children in the comparison group. If the kindergarten program offered prior to the reform was of lower quality than the program offered after the reform, then the comparison group would experience a positive impact of the reform. In this case, the new mandatory kindergarten program

[25]The baseline covariates are measured the year the child turns five years old, and include an overall measure of school size, the education level of the mother and father, the average income in the municipality of residence, and a dummy indicating whether the child lives in an urban area.

Table 4: Robustness – School performance

	Sample	Post	Coeff	SE	N
A. Key specification check					
Placebo	1989–1990	1990	0.006	(0.016)	110,171
B. Treatment-specific trends					
Extending pre-reform	1988–1992	1991–92	-0.007	(0.010)	267,745
Linear trend	1988–1992	1991–92	-0.017	(0.017)	267,745
Quadratic trend	1988–1992	1991–92	-0.008	(0.022)	267,745
C. Flexible trends					
Trend × covar	1988–1992	1991–92	0.016	(0.009)	267,745
Year FE × covar	1988–1992	1991–92	0.009	(0.009)	267,745
D. Other					
1st diff.: Treatment	1990–1991	1991	-0.006	(0.013)	18,108
1st diff.: Comparison	1990–1991	1991	0.004	(0.008)	93,288
Delayed effect	1988–1992	1991	-0.004	(0.011)	267,745
		1992	-0.011	(0.014)	

Note: Column 2 gives the estimation sample. In all estimations, $Post_t = 1$ for t is given in Column 3. In Panel A estimation is based on OLS on equation (1). In Panel B, estimations are based on equation (1), including a linear (row 2) and a quadratic (row 3) treatment-specific trend. In Panel C, estimations are based on equation (1), including a linear trend (row 1) or cohort dummies (row 2) interacted with a set of baseline covariates (school size; mother's and father's education level; municipal income; urban area). In rows 1 and 2 of Panel D, estimations are based on equation (2), while row 3 is based on equation (3). The controls are listed in Table 1 and the dependent variables are defined in Section 3. Standard errors (SE) are robust for heteroskedasticity and all models include municipality fixed effects.

may in fact have a positive effect for children in the treatment group that is simply netted out in our DD-setup against a positive effect (of similar size) in the comparison group. Similarly, a negative impact on the comparison group could mask a negative impact in the treatment group. To investigate this, we consider the two groups of children separately, to reveal whether there are in fact substantial changes in the grades of children around the implementation of the reform. Looking back at Panel (b) of Figure 2, we see no indication of such changes in neither the treatment nor the comparison group. More formally, and including covariates, we estimate first difference regressions separately for the two groups based on

$$Y_{it} = \alpha + \lambda Post_t + X_{it}^{'}\beta + \epsilon_{it} \tag{2}$$

Results reported in rows 1–2 in Panel D of Table 4 give no reason to believe that mandating kindergarten had much impact on neither the comparison group, nor the treatment group.

Bedding down effect. Children in our main sample, born 1990–1991 and entering grades 1 and 2 in 1997, started school in the same year, and were subject to the same, new curriculum throughout primary school. Children born in 1990 were the first to experience a new curriculum for grades 2–4, while children born in 1991 were the first to experience a

new curriculum in grade 1, but the second to experience a new curriculum for grades 2–4. If the bedding down of the curriculum was a disadvantage for the children born in 1990, then our estimates could be biased upwards, and might mask an otherwise negative effect. To investigate this, we first note that if bedding down effects are important, then we would expect the performance of the 1990-cohort to dip compared to the 1989 cohort. Figure 2 shows no evidence of such a dip. Of course, our estimates would not be affected by such overall effects in any case, since our DD estimation strategy removes cohort effects. We may, however, worry that children who did not attend kindergarten before starting school (our treatment group) are more sensitive to the bedding down of the curriculum. To investigate this, we take advantage of the fact that the 1989 cohort were never exposed to the new curriculum as a first cohort. If bedding down effecs are important and affect our treated group disproportionately, then we would expect a negative DD-effect when we compare the 1989 and 1990-cohort. Again, we see no evidence of this in Figure 2. Also, this is precisely what is estimated in the placebo test in Table 4, where the estimate is essentially zero.

We have also performed an additional placebo estimation, analogous to the previous, where we compare the 1988 cohort to the 1989 cohort. These cohorts were both transferred to the new curriculum in 1999, and the difference in exposure is therefore modest. Again, we find no evidence of diverging trends (b=0.008, SE=0.015).

Delayed effect. A further worry may be that a positive effect of the mandatory kindergarten reform was offset, completely or in part, by adjustment problems in the year of implementation. Unfortunately, the cash-for-care reform implemented in late 1998 (see Section 2), creates problems for identifying effects on cohorts born in 1993 and onwards. We can, however, plausibly estimate effects on children born in 1992. To create balance between the comparison group and the treatment group, and to provide better identification of control variables, we also use the extended sample of children born 1988–1992. To allow for different treatment effects on children born in the two years, we expand on equation (1) to include a separate interaction term for the 1992-cohort, i.e.

$$\begin{aligned} Y_{it} &= \alpha_t + \gamma_1 Treated_i + \lambda_{91} Treated_i \times 1\,(t = 1991) \\ &\quad + \lambda_{92} Treated_i \times 1\,(t = 1992) + X_{it}^{'}\beta + \epsilon_{it} \end{aligned} \tag{3}$$

where $1\,(t = s)$ is an indicator equal to one if $t = s$ and zero otherwise. Results are reported in the final row of Panel D, again revealing no evidence that the reform had an important impact on children's school performance at end of compulsory schooling. If anything, point estimates indicate that the 1992 cohort was doing worse than the 1991 cohort, suggesting that there was no delayed benefit of the program.

Alternative outcomes. We may also worry that the dependent variable is not picking up the relevant margin of the effect. For instance, if kindergarten affects mostly oral skills or

mostly skills that are relevant in one or a few particular subjects, then our estimate may be small simply because it is diluted by including subjects in our outcome that test skills that are not affected. To investigate this, we have also considered alternative outcomes that should reflect different sets of skills. For brevity, estimates are reported in Table A5 in the appendix.

We start by separating the written and oral exams that make up our main dependent variable. Since there are usually two written exams for each oral exam, any effect on the oral exam may be diluted by a zero or counteracting effect on the written exam. In Panel A of Table A5, we report results from our baseline regression where the dependent variable is replaced by first the average of written exam grades and then by the grade on the oral exam. Estimated effects are virtually identical. Next, we consider teacher-assigned grades at end of compulsory school, available for 13 subjects (see Section 3). For comparison with our main estimates, we first run our baseline regression on the overall grade point average (GPA; the mean grade across all subjects). Not surprisingly, the estimated effect on the overall GPA is virtually identical to the effect on the average exam grade used in our main analysis (cf. Panel B of Table A5). We then separate out the subjects that are tested on the written and oral exams used in our main analysis, and those that are not tested on these exams.[26] Again, estimates are virtually identical to the baseline.

Finally, we group subjects according to the types of skills expected to determine the performance. Specifically, we group subjects into the following categories: "Sciences" (mathematics, natural science, and social science), "Languages" (written and oral Norwegian, and written and oral English), and "Culture" (religion, music, home economics and arts and crafts). Estimates are reported in rows 4–6 in Panel B of Table A5, and are again virtually identical. We conclude, therefore, that there is no evidence of substantial effects that were not picked up in our main analysis.

Income effects. There are two alternative channels for income effects: One from a change in the use or price of child care/after school programs, and another from a change in parental labor supply following the reform. For children that would attend child care before the reform and the after school program after, the price difference is relatively small (cf. p. 6). For children that would not attend child care before the reform and not attend the after school program after, the price is always zero. Finally, some children who would not attend child care before the reform may choose to opt into the after-school program after the reform. These children would experience an increase in the price, and hence a negative income effect that could adversely affect their children. We believe that this latter effect is small for two reasons. First, parents who opt out of voluntary kindergarten should be likely also to opt out of after-school care. Second, parents who

[26]Subjects tested are written and oral Norwegian, written and oral English, mathematics, nature and science, social science and religion. Subjects not tested are home economics, physical education, music, and arts and crafts.

choose to opt into the after-school program would likely be close to indifferent between attending kindergarten in the first place. This may suggest that the price is not very high compared to their income.

To investigate effects on parental labor supply, we have estimated the impact of the mandatory kindergarten reform on maternal labor supply. This is important in itself, but could also indirectly affect child performance by increasing family income (Dahl and Lochner, 2012; Løken, Mogstad, and Wiswall, 2010).[27] Following standard practice, we restrict our analysis to mothers with their youngest child born in 1990 or 1991, where we would expect the strongest labor supply responses. We apply an analogous empirical strategy to the one in our main analysis, estimating the DD-model in equation (1) where t refers to the cohort of the youngest child, $Treat$ is equal to one for mothers of children who enroll in child care at age five, while $Post$ is equal to one if the youngest child is born in 1991 and zero if the child is born in 1990.[28] Results reveal that the reform had no impact on the labor market attachment of mothers in our sample (cf. Panel C of Table A5).

7 Concluding remarks

Evidence on the impact of child care interventions has been dominated by estimates from targeted programs. These may be hard to apply to the general population. Recent research provides some insight into the effects of large-scale programs. While the literature is expanding rapidly, there is still a lot we do not know about the impact of the universal programs advocated in many western countries (Baker, 2011). This is particularly worrisome given the heterogeneity created by wide differences in individual alternatives to subsidized care. The high returns found for children from disadvantaged families, coupled with much lower participation rates in existing programs compared to children from more advantaged backgrounds, suggests a potentially strong social gradient in expanding or mandating early childhood interventions (Barnett and Belfield, 2006). Indeed, in an effort to counter differences at school entry depending on social background, many countries are currently moving towards subsidized kindergarten or child care available for the

[27] In a survey of the early literature, Blau and Currie (2006) report elasticities of maternal employment with respect to the price of child care ranging from 0 to -1. More recently, using more plausible identification, Baker, Gruber, and Milligan (2008) find a positive effect on maternal labor supply following the introduction of heavily subsidized universally available child care in Quebec. Meanwhile, Lundin, Mork, and Ockert (2008) find no such effect when studying a childcare reform which capped childcare prices in Sweden. See also Schlosser (2005); Cascio (2009); Havnes and Mogstad (2011); Lefebvre and Merrigan (2008a) and Berlinski and Galiani (2007). For a review of the literature, see Blau and Currie (2006).

[28] As our outcome of interest, we consider labor supply of mothers in the year when the child turns seven years old. That is, ideally we consider the labor supply of mothers whose youngest child was enrolled or was not enrolled in kindergarten in the months January through late August. We have also estimated effects when the child is six (when the child may be enrolled September through December), finding no impact of the reform.

general population.

In the current paper, we provide first evidence on the effect of mandating kindergarten at age 5–6 on children's schooling outcomes. Specifically, we consider the impact on school performance at the end of compulsory schooling at age 15–16, on high school dropout and on the likelihood of enrolling in an academic track. Our identifying variation comes from a 1997-reform in Norway that lowered school starting age from seven to six. The goal of the reform was to counter differences in learning outcomes between children from different socioeconomic backgrounds. The contents and structure of the program bear resemblance to the US Head Start program and to the early US kindergarten program, in focussing mostly on social development and less on the aquisition of academic skills.

Our results reveal that the reform did little to counter differences in schooling outcomes between children from different socioeconomic backgrounds. In our baseline estimation, the precisely estimated effect on the child's exam performance is below 2 % of a standard deviation. Estimates are similarly small when we consider effects across the grading distribution and in different subsamples defined from characteristics of the child or parents. A number of specification checks lend support to our empirical strategy. A lack of effects on medium-term school performance could still mask a long-term effect on substantive outcomes. Evidence from the Perry Preschool program suggests that although an initial increase in IQ among treated children faded out and effects on school performance for boys were at best modest, the program still generated positive effects on longer-term outcomes such as crime and labor market behavior (Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010; Heckman, Pinto, and Savelyev, 2013). However, we find negligible impacts also on high school dropout and academic tracking at ages 18 and 16, respectively, which may arguably be of direct substantive interest.

While the program we study is universal, the reform may be viewed as targeted since the affected children come disproportionately from disadvantaged families. One may therefore interpret our results as shedding light on how a universal low intensity program can improve outcomes among the disadvantaged. Previous evidence suggests that low intensity child care has the potential to improve child outcomes in this group. Our evidence suggests that this may no longer be the case when the program is truly universal. This may not necessarily come as a surprise, since the larger operations and broader scope involved in a universal compared to a targeted program may come with particular challenges, for instance by making it harder to see children's needs and to tailor activities to these needs. One interpretation may then be that the unstructured child-centered approach to instruction (Stipek, Feiler, Daniels, and Milburn, 1995), which has been a hallmark of low intensity child care programs, may be less suitable in a universal program. Our study may then lend support to policies aimed at improving the intensity of targeted programs over policies aimed at expanding the reach of these programs. Another interpretation may be that parents are able to sort children relatively efficiently into child

care programs based on their children's individual needs, suggesting that mandating participation may not be effective in reducing socioeconomic differences between children. This might be particularly true when child care programs prior to mandating are widely available and affordable, as in our case. We emphasize that our analysis is based on children from non-immigrant families, and does not, therefore, speak to the debate on early interventions to provide language training among non-native speakers.

We believe that our evidence along with e.g. Baker, Gruber, and Milligan (2008), may call for caution in the current push towards using universal child care as a tool to promote the development of children from disadvantaged families. For instance, it could look as if the European Union Commission is reading too much into descriptive, rather than causal, evidence when they proclaim that "[t]here is clear evidence that universal access to quality ECEC is more beneficial than interventions targeted exclusively at vulnerable groups" (European Union, 2011, p. 5). While we agree that early childhood investments can be an important tool in facilitating equal opportunities, our evidence emphasizes that this is hardly automatic, and suggests that the structuring of the program and its content may be key to generating the intended benefits.

Finally, the conclusion that mandating kindergarten had little impact on children's school performance may cut both ways: While the large benefits expected by proponents can be firmly rejected, our results also lend little support to claims of strong negative effects from opponents. This is true even though the reform implemented a fully mandated program affecting families that did not voluntarily enroll their children, and who would otherwise care for their children themselves. It should be noted, however, that these estimates are driven mostly by children from relatively lower socioeconomic backgrounds, and may not be representative for children from higher socioeconomic backgrounds. Also, it is clear that the evidence on how universal or large-scale child care affects child development is mixed and still quite scarce, and that one reason could be that the alternative mode of care differs across countries. It is therefore of great importance to accumulate more evidence on how child care programs can affect child development.

References

Almond, D., and J. Currie (2010): "Human Capital Development Before Age Five," NBER Working Papers 15827, National Bureau of Economic Research, Inc.

Anderson, M. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Intervention Training Projects.," *Journal of the American Statistical Association*, 103(484), 1481–1495.

Atkinson, A. B., L. Rainwater, and T. M. Smeeding (1995): *Income distribution in OECD countries : evidence from the Luxembourg Income Study*. OECD Publications and Information Center, Paris.

Baker, M. (2011): "Innis Lecture: Universal early childhood interventions: what is the evidence base?," *Canadian Journal of Economics*, 44(4), 1069–1105.

Baker, M., J. Gruber, and K. Milligan (2008): "Universal Child Care, Maternal Labor Supply, and Family Well-Being," *The Journal of Political Economy*, 116(4), pp. 709–745.

Barnett, W. S. (1995): "Long-term effects of early childhood programs on cognitive and school outcomes," *Future of Children*, pp. 22–50.

Barnett, W. S., and C. R. Belfield (2006): "Early Childhood Development and Social Mobility," *The Future of Children*, 16(2), 73–98.

Berlinski, S., and S. Galiani (2007): "The effect of a large expansion of pre-primary school facilities on preschool attendance and maternal employment," *Labour Economics*, 14(3), 665–680.

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

Besley, T., and A. Case (2000): "Unnatural Experiments? Estimating the Incidence of Endogenous Policies," *Economic Journal*, 110(467), F672–94.

Bettinger, E., T. Hægeland, and M. Rege (2014): "Home with Mom: The Effects of Stay-at-Home Parents on Children's Long-Run Educational Outcomes," *Journal of Labor Economics*, 32(3), 443–467.

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

Black, S. E., P. J. Devereux, and K. G. Salvanes (2011): "Too Young to Leave the Nest? The Effects of School Starting Age," *The Review of Economics and Statistics*, 93(2), 455–467.

Blau, D., and J. Currie (2006): *Pre-School, Day Care, and After-School Care: Who's Minding the Kids?*vol. 2 of *Handbook of the Economics of Education*, chap. 20, pp. 1163–1278. Elsevier.

Cascio, E. U. (2009): "Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing Kindergartens into Public Schools," Working Paper 14951, National Bureau of Economic Research.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931.

Currie, J. (2001): "Early Childhood Education Programs," *Journal of Economic Perspectives*, 15, 213–238.

Currie, J., and D. Thomas (1995): "Does Head Start make a Difference?," *American Economic Review*, 85(3), 341–364.

Dahl, G. B., and L. Lochner (2012): "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit," *American Economic Review*, 102(5), 1927–56.

DeCicca, P., and J. Smith (2013): "The Long-Run Impacts of Early Childhood Education: Evidence from a Failed Policy Experiment," *Economics of Education Review*, 36, 41–59.

Deming, D. (2009): "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start," *American Economic Journal: Applied Economics*, 1(3), 111–134.

Drange, N., and M. Rege (2013): "Trapped at home: The effect of mothers' temporary labor market exits on their subsequent work career," *Labour Economics*, 24, 125–136.

Duflo, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91(4), 795–813.

Dustmann, C., A. Raute, and U. Schonberg (2013): "Does Universal Child Care Matter? Evidence from a Large Expansion in Pre-School Education," Mimeo.

European Union (2011): *Early Childhood Education and Care: Providing all our children with the best start for the world of tomorrow.*

Felfe, C., and R. Lalive (2013): "Early Child Care and Child Development: For Whom it Works and Why," CEPR Discussion Papers 9274, C.E.P.R. Discussion Papers.

Felfe, C., N. Nollenberger, and N. Rodríguez-Planas (2012): "Can't Buy Mommy's Love? Universal Childcare and Children's Long-Term Cognitive Development," *Journal of population economics*, 28(2), 393–422.

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77(3), 953–973.

Fitzpatrick, M. D. (2008): "Starting School at Four: The Effect of Universal Pre-Kindergarten on Children's Academic Achievement," *The B.E. Journal of Economic Analysis & Policy*, 8(1).

Garces, E., D. Thomas, and J. Currie (2000): "Longer Term Effects of Head Start," Working Paper 8054, National Bureau of Economic Research.

Gupta, N. D., and M. Simonsen (2010): "Non-cognitive child outcomes and universal high quality child care," *Journal of Public Economics*, 94(1-2), 30 – 43.

Havnes, T., and M. Mogstad (2011): "No Child Left Behind. Subsidized Child Care and Children's Long-Run Outcomes," *American Economic Journal: Economic Policy.*

——— (2012): "Is Universal Child Care Leveling the Playing Field?," CESifo Working Paper Series 4014, CESifo Group Munich.

Heckman, J., R. Pinto, and P. Savelyev (2013): "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *The American Economic Review*, 103(6), 2052–2086.

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

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–75.

Karoly, L. A., M. R. Kilburn, and J. S. Cannon (2005): *Early Childhood Interventions: Proven Results, Future Promise*. RAND Corporation, Santa Monica, CA.

Knudsen, E. I., J. J. Heckman, J. L. Cameron, and J. P. Shonkoff (2006): "Economic, Neurobiological, and Behavioral Perspectives on Building America's Future Workforce," *Proceedings of the National Academy of Sciences of the United States of America*, 103(27), pp. 10155–10162.

Lefebvre, P., and P. Merrigan (2008a): "Child-Care Policy and the Labor Supply of Mothers with Young Children: A Natural Experiment from Canada," *Journal of Labor Economics*, 26(3), 519–548.

——— (2008b): "Family Background, Family Income, Cognitive Tests Scores, Behavioural Scales and their Relationship with Post-secondary Education Participation: Evidence from the NLSCY," Cahiers de recherche 0830, CIRPEE.

Leuven, E., M. Lindahl, H. Oosterbeek, and D. Webbink (2010): "Expanding schooling opportunities for 4-year-olds," *Economics of Education Review*, 29(3), 319–328.

Løken, K. V., M. Mogstad, and M. Wiswall (2010): "What Linear Estimators Miss: Re-Examining the Effects of Family Income on Child Outcomes," IZA Discussion Papers 4971, Institute for the Study of Labor (IZA).

Ludwig, J., and D. L. Miller (2007): "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *The Quarterly Journal of Economics*, 122(1), 159–208.

Lundin, D., E. Mork, and B. Ockert (2008): "How far can reduced childcare prices push female labour supply?," *Labour Economics*, 15(4), 647–659.

McKey, R. H., L. Condelli, H. Ganson, B. J. Barrett, C. McConkey, and M. Plantz (1985): "The Impact of Head Start on Children, Families and Communities. Final Report of the Head Start Evaluation, Synthesis and Utilization Project.," Discussion paper, U.S. Department of Health and Human Services.

Norwegian Ministry of Child and Family Affairs (1995): *Lov om barnehager. Rundskriv Q-0902B. (The Child Care Act).*

Norwegian Ministry of Education (1990-91): *Ot.prp. nr. 57: Om lov om endring av lov 6.juni 1975 nr. 30 om barnehage m.m.,.*

——— (1992-93): *St.meld. nr. 40: ...vi smaa, en Alen lange; Om 6-åringer i skolen - konsekvenser for skoleløpet og retningslinjer for dets innhold.*

——— (1993-94): *Innst. O. nr. 36: Innstilling fra kirke-, utdannings- og forskningskomiteen om lov om endringer i lov av 13. juni 1969 nr. 24 om grunnskolen.*

——— (1996a): *L97: Læreplanverket for den 10-årige grunnskolen.*

——— (1996b): *Reform 97: Dette er grunnskolereformen.*

——— (2010): *NOU 2010: 8: Med forskertrang og lekelyst. Systematisk pedagogisk tilbud til alle førskolebarn (Systematic educational offering for preschool children)*, no. 8.

Reardon, S. F. (2011): "The widening academic achievement gap between the rich and the poor: New evidence and possible explanations," *Whither opportunity*, pp. 91–116.

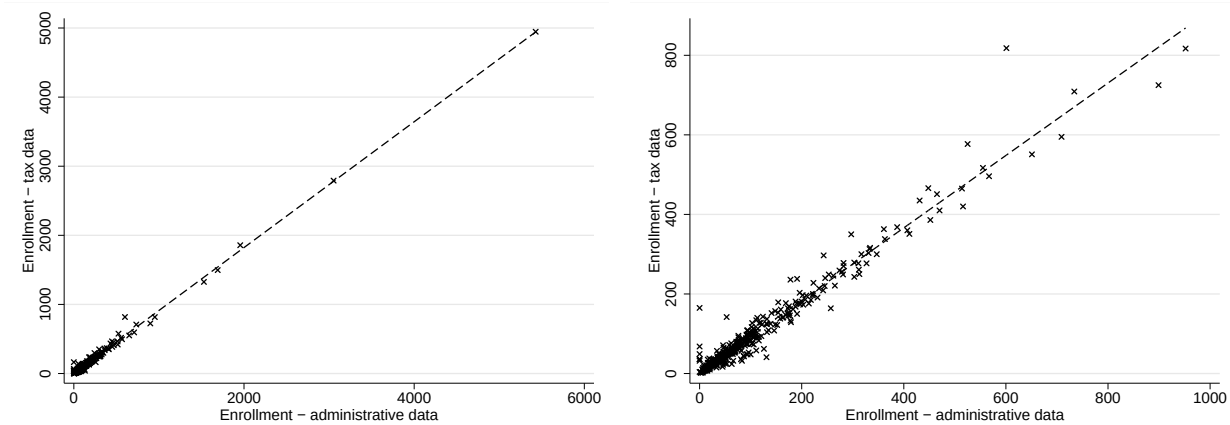
Ruhm, C., and J. Waldfogel (2012): "Long-term effects of early childhood care and education," *Nordic Economic Policy Review*, 1(1), 23–51.

Schlosser, A. (2005): "Public Preschool and the Labor Supply of Arab Mothers: Evidence from a Natural Experiment," Discussion paper, Mimeo, The Hebrew University of Jerusalem.

Schøne, P. (2004): "Labour supply effects of a cash-for-care subsidy," *Journal of Population Economics*, 17(4), 703–727.

Stipek, D., R. Feiler, D. Daniels, and S. Milburn (1995): "Effects of Different Instructional Approaches on Young Children's Achievement and Motivation," *Child Development*, 66, 209–223.

A Appendix



(a) Full sample

(b) Truncated sample

Figure A1: Enrollment – tax data and administrative data

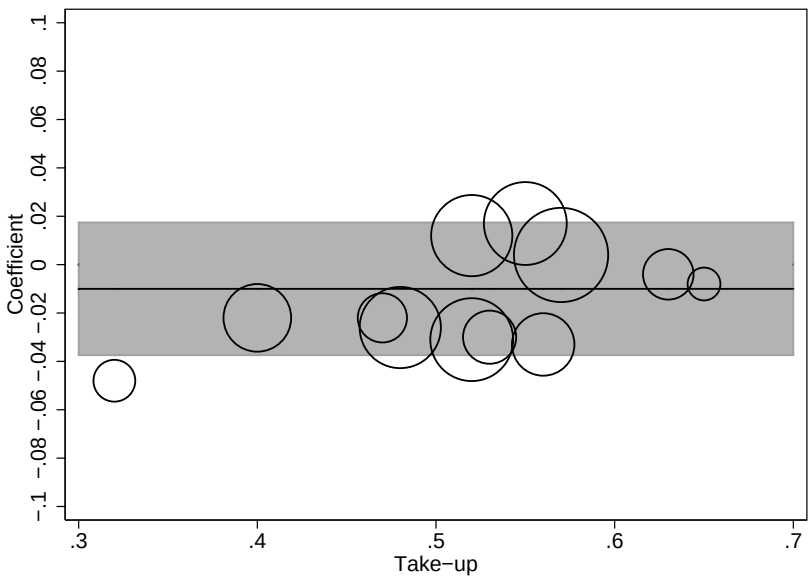


Figure A2: Estimated reform effect and take-up across subsamples

Note: The horizontal line and the shaded area correspond to the baseline estimate and its 95% confidence interval. Circles mark the estimates and take-up rates for subgroups reported in Table 3. The size of the circle indicates the size of the subgroup among treated children born in 1990. Take-up is defined as the probability that a child born in 1990 and not enrolled in child care at age five was also not enrolled in kindergarten at age six, see Section 4. Coefficients and take-up rates for subgroups are reported in Table 3.

Table A1: School performance and enrollment in kindergarten by family income decile at age five, children born in 1990.

Decile	Family income	School performance	Enrollment in kindergarten
1	15,912 (12,640)	3.42 (1.00)	0.54 (0.50)
2	45,124 (4,462)	3.53 (0.98)	0.65 (0.48)
3	57,408 (2,962)	3.60 (0.97)	0.72 (0.45)
4	66,968 (2,576)	3.68 (0.97)	0.82 (0.38)
5	75,547 (2,380)	3.73 (0.95)	0.88 (0.33)
6	83,524 (2,271)	3.79 (0.93)	0.91 (0.29)
7	91,695 (2,462)	3.86 (0.95)	0.92 (0.27)
8	101,363 (3,249)	4.02 (0.93)	0.93 (0.26)
9	116,100 (5,757)	4.11 (0.91)	0.93 (0.26)
10	169,179 (90,272)	4.30 (0.89)	0.92 (0.27)

Note: This table corresponds to Figure 1. School performance is measured as the average exam performance at the end of compulsory schooling (age 15–16). Enrollment in kindergarten is measured at age five. Standard deviations are in parentheses.

Table A2: Distributional effects on school performance

	Coeff	SE	Perc. value
5th percentile	0.002	(0.004)	2.0
10th percentile	-0.001	(0.006)	2.5
25th percentile	-0.011	(0.007)	3.0
50th percentile	-0.001	(0.006)	4.0
75th percentile	-0.005	(0.005)	4.5
90th percentile	-0.005	(0.004)	5.0
95th percentile	-0.002	(0.002)	5.5

Note: $N = 111,397$. Estimations are based on OLS on equation (1). The controls are listed in Table 1 and the dependent variable is defined in Section 3 and 5. Percentile values refer to pre-reform percentiles in the treatment group. Standard errors (SE) are robust for heteroskedasticity and all models include municipality fixed effects.

Table A3: Robustness – High school drop out

	Sample	Post	Coeff	SE	N
A. Key specification check					
Placebo	1989–1990	1990	-0.001	(0.007)	110,171
B. Treatment-specific trends					
Extended sample	1988–1991	1991	0.013	(0.005)	218,485
Linear trend	1988–1991	1991	0.011	(0.009)	218,485
Quadratic trend	1988–1991	1991	0.019	(0.023)	218,485
C. Flexible trends					
Trend × covar	1988–1991	1991	0.005	(0.006)	213,472
Year FE × covar	1988–1991	1991	0.008	(0.006)	213,472
D. Other					
1st diff.: Treatment	1990–1991	1991	0.017	(0.007)	18,108
1st diff.: Comparison	1990–1991	1991	0.004	(0.003)	93,288

Notes: Column 2 gives the estimation sample. In all estimations, $Post_t = 1$ for t is given in Column 3. In Panel A, estimation is based on OLS on equation (1). In Panel B, estimations are based on equation (1), including a linear (row 2) and a quadratic (row 3) treatment-specific trend. In Panel C, estimations are based on equation (1), including a linear trend (row 1) or cohort dummies (row 2) interacted with a set of baseline covariates (school size; mother's and father's education level; municipal income; urban area). In rows 1 and 2 of Panel D, estimations are based on equation (2), while row 3 is based on equation (3). The controls are listed in Table 1 and the dependent variables are defined in Section 3. Standard errors (SE) are robust for heteroskedasticity and all models include municipality fixed effects.

Table A4: Robustness – Academic track

	Sample	Post	Coeff	SE	N
A. Key specification check					
Placebo	1989–1990	1990	-0.003	(0.008)	105,894
B. Treatment-specific trends					
Extending pre-reform	1988–1992	1991–92	-0.003	(0.005)	258,112
Linear trend	1988–1992	1991–92	-0.008	(0.009)	258,112
Quadratic trend	1988–1992	1991–92	-0.016	(0.011)	258,112
C. Flexible trends					
Trend × covar	1988–1992	1991–92	0.005	(0.005)	252,495
Year FE × covar	1988–1992	1991–92	0.002	(0.005)	252,495
D. Other					
1st diff.: Treatment	1990–1991	1991	-0.020	(0.007)	17,203
1st diff.: Comparison	1990–1991	1991	-0.011	(0.003)	90,504
Delayed effect	1988–1992	1991	-0.009	(0.006)	258,112
		1992	0.005	(0.007)	

Note: Column 2 gives the estimation sample. In all estimations, $Post_t = 1$ for t is given in Column 3. In Panel A, estimation is based on OLS on equation (1) for years 1989 and 1990, years 1988 and 1989 and years 1988 and 1990. In Panel B, estimations are based on equation (1), including a linear (row 2) and a quadratic (row 3) treatment-specific trend. In Panel C, estimations are based on equation (1), including a linear trend (row 1) or cohort dummies (row 2) interacted with a set of baseline covariates (school size; mother's and father's education level; municipal income; urban area). In rows 1 and 2 of Panel D, estimations are based on equation (2), while row 3 is based on equation (3). The controls are listed in Table 1 and the dependent variables are defined in Section 3. Standard errors (SE) are robust for heteroskedasticity and all models include municipality fixed effects.

Table A5: Alternative outcomes

	Coeff	SE	N	Mean
A. Separating written and oral exams				
Exam, written subjects	-0.012	(0.015)	108,473	3.21
Exam, oral subjects	-0.012	(0.015)	105,224	4.05
B. Teacher-assigned grades				
Grade point average	-0.013	(0.014)	111,185	3.79
Exam subjects	-0.014	(0.014)	111,021	3.64
Non-exam subjects	-0.007	(0.015)	111,038	4.12
Sciences	-0.011	(0.014)	111,225	3.57
Languages	-0.015	(0.015)	110,951	3.66
Culture	-0.008	(0.014)	111,157	4.11
C. Mothers labor supply, child age 7				
Earnings (2011-USD)	1,540	(914)	46,742	20,110
Employment	-0.002	(0.010)	46,742	0.23
Full time	-0.000	(0.008)	46,742	0.10

Note: In Panel A and B estimations are based on OLS on equation (1) including covariates. "Sciences" includes mathematics, natural science, and social science; "Languages" includes written and oral Norwegian, and written and oral English; "Culture" includes religion, music, home economics and arts and crafts. In Panel C estimations are based on OLS on equation (1), and the sample is restricted to mothers with youngest child of relevant age. We define a mother as being employed if she is registered with more than four working hours per week, and in full time employment if she is registered with more than 30 working hours per week while earning more than two times the basic amount in the Norwegian pension system (about USD 26,000). We restrict full time employment also on the level of earnings to correct for lags in the submission of employee information by firms, which causes some individuals with low or even zero earnings to be recorded as full time workers. The basic amount of the Norwegian Social Insurance Scheme is used to define labor market status, and determine eligibility for unemployment benefits as well as disability and old age pension. Covariates included are listed in Table 1 (in Panel C, we exclude measures of mothers employment and include municipality-specific unemployment rates). Standard errors (SE) are robust for heteroskedasticity and all models include municipality fixed effects.