Childcare quality and children's educational outcomes: a discontinuity approach*

Jo Blanden
University of Surrey

Emilia Del Bono
University of Essex

Kirstine Hansen
University College London

Birgitta Rabe
University of Essex

12th April 2019

Abstract

We investigate if the effect of an additional term of pre-school education on children's educational achievement varies by the quality of the setting attended; as measured by staff qualifications and nursery inspection ratings. We exploit date-of-birth discontinuities that entitle children born a few days apart to different amounts of free pre-school and use administrative data on children's school outcomes and the nursery they attend. There is no beneficial effect of more time spent in nurseries with highly qualified staff, but children have better educational outcomes at age 5 if they spend more time in a setting with the highest inspection ratings.

JEL Classification:I22, I24, C21
Keywords: childcare, child outcomes, regression discontinuity

*We thank the Department for Education for making available, under terms and conditions, data from the National Pupil Database and the Early Years Census. The Family Resources Survey is available through the UK Data Service. We are grateful to seminar participants at the Universities of York, St Gallen, Surrey, Hamburg, Essex, DIW Berlin, UCL, Strathclyde and the Institute for Fiscal Studies for helpful comments and suggestions. This work was funded by the ESRC under grant ES/K003232/1 and by the Nuffield Foundation. Emilia Del Bono and Birgitta Rabe acknowledge the support provided by the ESRC Centre on Micro-Social Change at ISER (award no. RES/518/28/001). Corresponding author: Jo Blanden, J.Blanden@surrey.ac.uk.

1 Introduction

In his 2013 State of the Union Address President Barack Obama stated his ambition "to make high-quality preschool available to every single child in America". The focus on "high quality" has become standard in the discussion of Early Childhood Education and Care in the academic literature (Cascio, 2015, Anders and Ulferts, 2016) and among policy makers across the world (OECD, 2015), but we know relatively little about the policy levers that will improve quality and benefit children's life chances. Policy emphasis on Early Childhood Education and Care (ECEC) is supported by evidence from small scale, randomized interventions targeted at disadvantaged children which found positive effects on children's attainment during the school years that, in some cases, extended into adulthood. These have provided important evidence that early years interventions have the potential to improve children's outcomes and reduce inequality (Barnett, 1995; Heckman et al., 2010; Karoly et al., 2005). However, international evidence on the impact of universal provision of ECEC on children's outcomes is more mixed with some studies finding positive and others no or even negative effects.[1]

Given this mixed evidence, understanding the features of successful universal ECEC systems is essential. The idea that only high quality child care should matter very often rests on *ex-post* comparisons between the characteristics of successful programmes (e.g. Norway) with those of programmes found to be less beneficial (e.g. Quebec), and resonates strongly with policy-makers. Despite the intuitive appeal of this argument, however, we cannot point out precisely what features of early education provision define its quality and whether quality indicators are causally and positively associated with child outcomes (Cascio 2017, Sabol et al., 2013).

In this paper we ask three questions. The first is whether there is any benefit to pre-school education for children's educational attainment. Here we exploit the fact that in the UK the rules governing eligibility to an extra term (about 3.5 months) of free part-time pre-school at age 3 depend on date-of-birth, and that this is randomly assigned in the proximity to the relevant cut-off, to arrive at a causal estimate. The second is whether the effect of eligibility to an extra term of pre-school varies with the quality of the setting attended in a way that suggests that higher quality is related to better child outcomes. The third important question is whether we can establish which measures of quality are most strongly related to differences in educational attainment, in order to inform policy.

Specifically, we use two established indicators of quality, staff qualifications and ratings from nationwide nursery inspections, and relate variation in these measures across nurseries in England to child outcomes in the first years of school. Previous studies are often based on researcher-collected quality indicators, typically available only in small datasets. These detailed measures of quality are often only weakly associated with the levers available to policy-makers, and therefore

[1] Positive effects are found in Gormley and Gayer (2005) and Fitzpatrick (2008) for the US and in Havnes and Mogstad (2011) for Norway. However, Baker et al. (2008) and Herbst and Tekin (2010) find negative effects of the introduction of subsidised universal childcare in Quebec and of subsidies for child care provided to working mothers in the US, respectively, with particularly strong negative effects found for aspects of social and emotional development. Datta Gupta and Simonsen (2010) find no effects of ECEC enrolment on outcomes at age 7 in Denmark.

of limited practical relevance. Instead, the quality measures we use are collected because of the existence in England (as well as many other countries) of a regulatory framework governing both school inspections and staff qualifications. So, they are of immediate relevance for policy decisions.

Staff qualifications are a fundamental measure of *structural* quality (the inputs used). Having a carer with a degree level qualification or equivalent in the setting is often considered best practice[2] and many studies emphasise the importance of staff qualifications for high quality childcare (Ulferts and Anders, 2016). Quality ratings can instead be considered closer to a measure of *process* quality, i.e. they take into account the interaction between children and carers in the settings, as well the interactions among the children themselves. Regulation and inspection are used to promote both process and structural quality and these have become more widespread in recent years (OECD, 2015, Gambaro et al., 2014), and can be either enforced by law or through voluntary schemes, such as the QRIS (Quality Rating and Improvement System) in the US.

Our analysis takes place in the English context where universal ECEC comes in the form of the free entitlement to part-time childcare (hereafter the "free entitlement"), a subsidy which costs the government around £2bn per year (Department for Education, 2013).[3] The policy was rolled out across England in the early 2000s, and 94 percent of children now benefit from part-time ECEC at age 3, delivered in public and private sector settings (National Audit Office, 2016). We focus here on the universe of children receiving their entitlement in the private sector where provision has been shown to be most heterogeneous in terms of quality (Gambaro et al., 2015, Mathers et al., 2007) and which caters for roughly half of children in each birth cohort.

We use administrative data from the National Pupil Database to measure the effects of eligibility for childcare and its quality on teacher-assessed measures of academic and social skills recorded at the end of the first year of school, at age 5, and on test results in English, Maths and Science at age 7. We have information on the precise date of birth for four cohorts of children (270,000 who attended private settings) born close to the relevant eligibility cut-off dates who started school from academic years 2008 to 2011. Children are linked to information on the characteristics of childcare settings attended at age 4, including staff qualifications and inspection ratings published by the English education regulator Ofsted.

To assess the causal effect of eligibility to an extra term of part-time childcare we use a regression discontinuity design, exploiting variation in eligibility to a free place due to strictly enforced date of birth discontinuities which mean that children born just a few days apart are entitled to different amounts of free early education while starting school at the same time and within the same school-cohort. We show that treatment is as good as randomly assigned close to the cut-off.

[2]The National Academy of Sciences Committee on Early Childhood Pedagogy advised that "Each group of children in an early childhood education and care program, should be assigned a teacher who has a bachelor's degree with specialized education related to early childhood" (National Research Council, 2001, p.13). The Tickell Review for the UK Department of Education (2011) states "an ambition for the sector to become fully graduate-led" (p.42).

[3]Similar policies are in place in the other nations of the UK, but we only have adminstrative data for England.

Moreover, we use an additional data set, the Family Resources Survey to test that eligibilty affects take-up of childcare, and find that it increases childcare use by 10-15 percentage points.[4] To assess the effect of an extra term spent in settings of different quality we control comprehensively for observed and unobserved differences across children attending different quality settings. As we shall discuss, our estimated quality effects can be given a causal interpretation if sorting is completely controlled for or the quality effect is homogenous across any remaining unobserved differences between children.

We find that eligibility for an extra term of childcare increases the probability that a child reaches the expected level of competencies after the first year in primary school by 0.6 percentage points (1 percent of the mean) compared to children not eligible for the extra term. This is an Intention-to-Treat-effect (ITT) which under some strong assumptions can be scaled up using our estimates of the effect of eligibility on take-up of childcare to give a Treatment-on-the-Treated effect of 4 percentage points, or 7 percent on the mean.[5] These results imply that the impact of ECEC in England on educational outcomes is smaller than the effects of early compulsory schooling in England found by Cornellisen and Dustmann (*forthcoming*) but of a similar magitude to the longer term effects in Havnes and Mogstad (2011) for Norway.

Further, we find that the effects of eligibility are seen on literacy scores and creative development but not on numeracy scores and personal development. Heterogeneity analysis reveals stronger effects on boys, a result also found in a different evaluation of the free entitlement policy (Blanden et al., 2016), but no significant pattern in relation to children's socio-economic characteristics. In line with much of the early child development literature we find that results fade out quickly and are no longer evident in outcomes measured at age 7.

The analysis of the impact of childcare quality first tests whether the effect of entitlement varies according to the presence of a carer with a degree-level qualification. We do not find any evidence that the effects of eligibility vary according to the qualification level of staff working with 3-4 year olds. This is the case for different ways of capturing formal staff qualifications. Next we use the rating of the setting as recorded by the national regulator. Attending an Outstanding setting brings an additional benefit from eligiblity to an extra term; it increases the probability of working at or above the expected levels of achievement by 1.3 percentage points compared to children in lower quality settings. Probing further we find that of the finer measures we have available, Leadership and Teaching ratings matter most. These results imply that although eligibility interacted with formal staff qualifications does not show a systematic association with child attainment, the quality of teaching is, nonetheless, important.

In assessing the impact of ECEC quality on children's educational outcomes this paper contributes to a number of literatures. First, it enhances the evaluation of preschool programmes by investigating not just whether programmes make a difference, but *why* some programmes might be more successful than others. Few studies within economics currently question whether variation in setting quality within the same childcare system can be linked to variations in child

[4]Children frequently attend ECEC before they are eligible for subsidized provision, or more rarely start later or not at all (Campbell et al. 2018).

[5]This is derived by dividing the effect by the increase in take-up triggered by eligibility, 0.006/.15=0.4.

outcomes (Duncan and Magnuson, 2013). We speak to this question. The administrative data that we have allows us to explore policy-relevant measures that capture both structural and process concepts of quality. Structural regulations on staff qualifications are more straightforward to put in place than expensive nationwide inspections, so understanding the relative merits of each is of crucial relevance for policy making. Our results imply that inspection judgements can contain valuable information and encourages both researchers and policy makers to continue to analyse and develop them.

More broadly, our results on structural and process quality add to the literature on school and teacher effectiveness which so far has focused on the school years (Walters, 2015, on Head Start is one exception). This literature generally finds big differences in teacher effectiveness but struggles to identify measurable teacher characteristics such as education and experience which drive these (Rockoff, 2004, Rivkin et al. 2005). In constrast, Dobbie and Fryer (2011) are able to identify particular practices (processes) which drive the differing success of charter schools but confirm that staff qualifications and pupil-teacher ratios do not have predictive power. In this vein, our results show that in the pre-school context, staff quality matters as it is recognised within the school inspection regime, but is not adequately proxied by staff qualifications.

The next Section discusses the literature on childcare quality. Section 3 describes the institutional background to the English education and childcare sector, and more specifically, the free entitlement. Section 4 lays out our empirical strategy, based on a regression discontinuity design, while Section 5 provides information on the data used. Section 6 describes our results, and Section 7 concludes.

2 Previous literature on childcare quality

Within the economics literature the conclusion that pre-school quality matters is generally reached by comparing the features of programmes which show substantive benefits such as those in Norway (Havnes and Mogstad, 2011), Spain (Felfe et al., 2015) Oklahoma and Georgia (Cascio and Schazenbach, 2013), with those with no benefits such as Quebec (Baker et al., 2008, Haeck et al., 2017), and Denmark (Datta-Gupta, 2010). In this spirit Cascio (2017) compares a number of US states and while she demonstrates that compared to contemporary targeted programmes universal systems have much greater benefits for disadvantaged children, it is not possible to identify the precise features of programmes that lead to success.

The ideal research design to assess the impact of quality would randomly assign quality to settings or children, although this is rarely feasible.[6] Education researchers relating quality to children's outcomes commonly try to account for the selection of children into settings by adopting a value-added specification, controlling for differences in potential between children by

[6] Exceptions are Araujo et al. (2016) who consider the random allocation of teachers to classes in the first year of school in Equador and Jensen et al. (2013) who consider the random assignment of childcare settings to a quality improvement programme. Both find substantial effects.

using baseline tests (Ulferts and Anders, 2016.) Our approach, in contrast, combines quasi-experimental variation in time spent in childcare with comprehensive controls for selection into different quality settings. Walters (2015) is perhaps the study with the closest approach to our own. This makes use of random assignment to Head Start within settings and relates treatment effects to setting and child characteristics. Some aspects of practice are shown to make a difference (settings that offer full-time care and home visits are more effective) but, as in our study, having a teacher with a Bachelor's degree is not shown to have a significant positive effect.

Researchers often distinguish between the effects of structural quality and process quality. Structural quality consists of the more easily observable aspects of the setting such as space, staff-child ratios, group size and qualifications of staff. Process quality concerns the more proximal processes of children's everyday experiences and involves the social, emotional physical, and instructional aspects of staff-child and peer interactions while being involved in play, activities or routines. These are often measured using scales based on classroom observations. Examples are the ECERS-R (which rates the pre-school environment generally, Harms et al. 2014), the ECERS-E (which focuses on the educational curriculum, Sylva et al., 2006) and the CLASS (which is focused on the interactions that take place between adults and children, Pianta et al., 2007; and used in Araujo et al., 2016). While we do not have these detailed measures of process quality, the measures we do have are directly related to policy - both in the UK and other developed countries - as they capture the aspects of provision that are often regulated on in different countries.

Of the two types of measure, there is stronger descriptive evidence that measures of process quality are associated with outcomes, with higher ratings for pedagogic activities shown to be linked to child outcomes (see Ulferts and Anders, 2016 for a review). Peisner-Feinberg et al. (2001) find influences of process quality on cognitive skills and behaviour extending through to second-grade. US research has found measures of structural quality to be rather weakly correlated with children's outcomes (Mashburn et al., 2008). Early et al. (2007) review a number of studies and find no effect on children's school readiness of being taught by an instructor with a Bachelor's degree but positive effects of broader measures of staff qualifications on outcomes. Evidence from the UK suggests that staff qualified to graduate level are able to produce better process quality (Siraj-Blatchford et al., 2005, Mathers et al., 2011), although the direct link between higher-level qualifications among staff and children's development one year later is quite weak (Blanden at al. 2017).

Alongside staff qualifications we measure quality by the Ofsted ratings awarded to the nursery. These evaluations cover both structural aspects such as space and resources, and an evaluation of the activities the children take part in; this should therefore combine aspects of both structural and process quality (Mathers et al., 2012). Ofsted awards an overall grade ranging from 1, "Outstanding" to 4, "Inadequate" for overall effectiveness. The existing evidence from England on the relationship between Ofsted ratings, quality and children's outcomes is inconclusive. Mathers et al. (2012) find that, on average, settings graded as "Outstanding" by Ofsted achieve higher ECERS process quality scores than "Good" settings, which do better than settings graded as "Satisfactory". However, those graded as "Inadequate" do not have the lowest quality ratings

on average. Hopkin et al. (2010) examine the impact of childcare Ofsted ratings on a range of cognitive tests administered as part of the survey as well as teacher graded assessments of the children collected from their schools at age 5 and find no link. Neither of these studies are based on representative samples.

In the absence of a national regulator like Ofsted the vast majority of US states have adopted a voluntary QRIS (Quality Rating and Improvement System) in an effort to improve the information available to parents and drive up the quality of ECEC (Elicker et al., 2011, indicate that parents act on the QRIS, at least in Indiana). Schemes differ between states but the idea is to award settings or wider programmes a grade or star-rating, (usually on a one to four or one to five basis) based on measures of structural and process quality. Given that QRIS ratings incorporate cut-offs of structural measures which have been shown to have little influence on children's outcomes (such as group-size 20 or below) and process quality (where acceptable levels in the ECERS-E varies surprisingly widely across systems) it is hardly surprising that studies have shown only fairly weak, if any, relationships between QRIS ratings and children's outcomes (Zellman and Perlman, 2008; Sabol et al., 2013). This study uses a novel identification strategy to understand if the levers used for quality regulation in England have a clearer link to outcomes.

3 Institutional Background

Since 2004 all English Local Authorities (equivalent to school districts) have been funded to provide universal part-time early years education and care for children from the term after their third birthday.[7] For the cohorts we study here this was 12.5 hours for 38 weeks a year until 2010, extending to 15 hours per week from September 2010 onwards.[8] In further expansions beyond the period we study, disadvantaged two year olds have also been offered 15 hours of free care since 2013 and since September 2017 "working families"[9] have been entitled to 30 hours.

In England all children enter primary school in the academic year in which they turn 5 (the Reception year). In recent years most schools have adopted a unique intake date in September. This implies that irrespective of their date of birth, all children within a school-cohort (going from 1st September to 31st August) start formal schooling at the same time (but at a different age). In contrast, eligibility to free part-time pre-school care changes discontinuously across the year; children born between 1st September and 31st December are entitled to claim their free hours from the following January, children born between 1st January and 31st March from April, while those born between 1st April and 31st August are allowed to claim their entitlement only from

[7] The impact of the roll out of this policy on children's educational outcomes was studied in Blanden et al. (2016). Estimated effects are small in the short-run. This is largely explained by the fact that few families changed their behaviour when the policy was implemented. Effects faded out quickly even among those groups who took up childcare as a result of the free entitlement.

[8] This change will affect the last cohort in our sample and we investigate the impact of this in our sensitivity checks.

[9] Working families are defined as two parent families where both are working the equivalent to 16 hours a week at the National Living Wage (although they can work less if they earn more), and who both earn less than £100,000 each. A single parent will qualify if (s)he meets the working criteria applied to each dual parent (Department for Education, 2015).

September of the following school year. To the extent that children's participation is governed by their entitlement, children who experience more months in free ECEC will also start at a younger age. Our analysis considers only children born around the 31st December and 31st March cut-offs, who are different in respect of their eligibility for free early education and care but start school at the same time and belong to the same school cohort.

Around half of children are provided their free place in the public sector and the other half in the private sector, with eligibility rules being the same across both sectors, although insitutional arrangements are more flexible in the private than the public sector, leading to more variation between private sector settings. Whether children attend childcare in the public or private sector will depend on availability where they live, the preferences of parents and the hours of care required.[10] Opening hours in the public sector are restrictive, never exceeding school hours (about 6.5 hours a day) and sometimes children are offered only morning or afternoon sessions. In contrast the private sector can provide full-time care. Moreover, public sector settings must have a qualified teacher present, and the adult-child ratio is set to 1:13. Requirements for qualifications are lower in private settings, but if there is no qualified teacher or person with Early Years Professional Status (EYPS)[11] present then the ratio of adult per child is increased to 1:8 (Gambaro et al., 2015). Notably, the EYPS does not qualify individuals to work as a nursery teacher in the public sector, implying that the two qualifications are not universally viewed as comparable.

All providers who receive Government funding are required to follow a common curriculum, the Early Years Foundation Stage. The curriculum emphasises learning through play, ensures that a range of stimulating activities are provided and that children's development across a range of areas is encouraged and monitored. All settings are subject to inspection by the Government regulator Ofsted (Office for Standards in Education), roughly every four years. Ofsted states the purpose of its inspections is "to judge the overall quality and standards of the early years provision in line with the principles and requirements of the "Statutory framework for the Early Years Foundation Stage" (Ofsted, 2015). Inspection judgements for private settings are based on one-day visits which gather evidence by observation, discussion with staff and parents and by reviewing the development of example children (through discussion and observation of records, with a focus on the disadvantaged) (Ofsted, 2015). Public provision in nursery classes are inspected as part of the whole school's inspection, which leads to doubt about the acccuracy of judgements, providing a further rationale for our focus on private provision.

[10] Public sector nursery education was provided by some local authories in decades prior to the introduction of the free entitlement. The expansion of free ECEC occurred almost entirely in the private (or Private, Voluntary and Independent, PVI, sector). See Blanden et al (2016).

[11] Early Years Professional Status was created in 2006 as an alternative to Qualified Teacher Status for leaders in this field, and both qualifications are considered as degree-level qualifications. In order to be awarded EYPS individuals are required to demonstrate that they meet 38 professional standards when working with children from 0 to 5 years old. Training routes vary and accreditation can take from four months part-time to one year full-time depending on the experience of the individual (Mathers et al., 2012). Even the long route is considerably shorter than QTS training which usually takes three years full-time or one year via the post-graduate route. The qualification has now been replaced by Early Years Teacher Status.

4 Empirical Strategy

Our empirical strategy proceeds in two steps. First, we establish whether entitlement to an extra term of free part-time early education has a significant effect on child educational attainment. Then, we consider whether the effect of eligibility varies according to the quality of the pre-school setting. In this way we try to understand whether the quality of early education - as measured in our data - matters for children's outcomes.

Access to free part-time early education and care is based on strict date-of-birth rules. This enables us to pursue a sharp Regression Discontinuity Design (Imbens and Lemieux, 2008; Thistlethwaite and Campbell, 1960) to assess the impact of eligibility to an additional term of the free entitlement on educational outcomes.

We define an indicator variable T_i which takes value 1 when the child's date of birth a_i is before the cut-off date $\bar{a}$ which entitles them to a free place at the start of the following school term. Children whose birth date is after the cut-off, will have to wait another term before becoming eligible for the subsidy, so that for them T_i has a value of 0.

$$T_i = I\{a_i < \bar{a}\} \tag{1}$$

We want to relate child outcomes to eligibility for an extra term of childcare, which is a function of date of birth. All children take their assessments at the same time, so that date of birth determines age at test. Eligible children will be, by construction, older than non-eligible children, and owing to the well-documented positive relationship between age at test and test scores (Crawford et al., 2014; Leuven et al., 2010) they will have better outcomes. It is therefore essential that we control for a flexible function of date of birth. Although we assume that eligibility is unrelated to the child's observed and unobserved characteristics, all our specifications also control for X_i, a vector of individual-level characteristics; this also improves the precision of our estimates. Our estimation equation is thus:

$$Y_i = \beta T_i + f(a_i) + \Pi X_i + \varepsilon_i, \tag{2}$$

where the outcome of child i, Y_i, is a function of eligibilty T_i, date of birth a_i, a vector of child characteristics X_i and ε_i is a random error term. All our models contain school fixed-effects to allow for the fact that assessments may differ systematically across schools. Standard errors are clustered at the level of date of birth and school. The clustering by date of birth is particularly important as this is the variable which defines our treatment, i.e. eligibility to an extra term of childcare.

Following Altonji and Mansfield (2018), we augment the specification above using the averages of individual characteristics of children in different settings, $\overline{X_j}$. This is important in situations where there might be sorting of individuals into different groups (such as pre-schools), and the outcome is a function of individual as well as group characteristics. So, for each child i who

attends pre-school j our model becomes:

$$Y_{ij} = \beta T_{ij} + f(a_{ij}) + \Pi X_{ij} + \lambda \overline{X}_j + \varepsilon_{ij}. \tag{3}$$

The graphical analysis presented in the next section suggests that $f(\cdot)$ is a linear function of date of birth and we use this formulation in most of our analysis. We also run models where $f(\cdot)$ is specified as a quadratic function of date of birth or as a linear function whose slope is allowed to change at the cut-off. Our data has the advantage of a very large sample size which means we can restrict estimates to data very close to the discontinuity (births within four weeks either side of the cut-off), thereby minimising the sensitivity of our results to the specification of $f(\cdot)$. Related to this, we will show how our estimates change with the size of the data window around the cut-off.

In order for specification (3) to produce a valid estimate of the causal effect of eligbility to an extra term of childcare, the treatment must be as good as randomly assigned close to the cut-off. This assumption can be checked in two ways; by showing that eligibility is orthogonal with respect to the observed determinants of test scores, and by checking for changes in the density function of the running variable (date of birth) around the cut-off. If births are concentrated on one side of the eligibility cut-off, this might suggest that families can choose the date of birth of their children to take advantage of the policy, implying that those receiving the treatment have selected into it. We will provide evidence that there is no systematic association between observed characteristics and date of birth and that the frequency of births is smooth around the cut-off below.

Equation (3) allows us to estimate β, the effect of eligibility on child outcomes and is an intention-to-treat effect (ITT). We are also interested in the effect of ECEC attendance, that is the effect of treatment on the treated (TT). However, to achieve this we would need information on when precisely children start attending pre-school and our data does not provide this.[12] Instead, we use information from a separate dataset, the Family Resources Survey, to show the relationship between eligibility and attendance close to the eligibility threshold. We then use these estimates to give an idea of the effect of ECEC attendance on child outcomes or the TT. As eligibility might affect child outcomes through channels other than attendance, such as the number of hours of early education or family income, our main focus remains on the effect of eligibility or the ITT.

The second step in our analysis is to examine whether the quality of the pre-school setting influences the effect of eligibility. To do this, we include in our estimation the available measures of setting quality, Q_j, as well as interactions of $Q_j$and eligibility status. Our model thus becomes:

$$Y_{ij} = \beta T_{ij} + f(a_{ij}) + \psi Q_j + \phi T_{ij} * Q_j + \Pi X_{ij} + \lambda \overline{X}_j + \varepsilon_{ij}. \tag{4}$$

Here the coefficient ψ shows the association between the measure of setting quality and child outcomes. This is interesting per se, but it cannot be given a causal interpretation because even

[12]We only have information about the nursery the child attended in the January before they start school, i.e. several months after the child qualified for the extra term of free part-time early education.

though we use setting-level averages of child observable characteristics to control for sorting, we cannot totally exclude the possibility that children choose different pre-schools on the basis of their unobserved characteristics. In other words, the coefficient on Q_j captures unobserved differences across children that attend settings with different quality characteristics, as well as the true effects of quality on outcomes.

Our main interest is instead on the coefficient ϕ. This represents how the eligibility effect - and therefore ECEC attendance - varies with the quality of the setting. It may not, however, capture the causal effect of quality of pre-school education on children's outcomes if the returns to an extra term in pre-school are different in higher quality nurseries because of sorting by background characteristics such as socio-economic status, for example because more advantaged children benefit more. However, we can say that ϕ picks up a causal effect of quality if one of two conditions are met: either 1) sorting is completely controlled for, or 2) the effect of eligibility is not heterogeneous with respect to those individual characteristics we cannot control for, but that influence the sorting into settings. As well as controlling for sorting through $\overline{X_j}$, we control for it through Q_j - recall that this captures unobserved characteristics of children attending different quality settings - and we show that the eligibility effect does not vary with the observable characteristics of the child. We can therefore be confident that our novel approach comes very close to capturing the genuine effect of ECEC quality.

5 Data

5.1 National Pupil Database

Our analysis is based on data from the National Pupil Database (NPD). This is an administrative dataset containing information on the educational achievement of all children attending public (state) schools in England, and covering about 93 per cent of all pupils in the country. The dataset can be matched to the Pupil Census to add information on child characteristics including gender, eligibility for free School Meals (FSM), ethnicity, whether the first language spoken at home is English, and the level of income deprivation in the neighbourhood around the child's postcode of residence. Families entitled to FSM are usually in receipt of means tested benefits and/or have one if not both jobless parents. As is standard this indicator will be used to distinguish low- from medium to high-income families, and although it has its limtations it is a reasonable proxy (Hobbs and Vignoles, 2010). The dataset is longitudinal, in that it follows each child over the primary and secondary school years, and contains school and Local Education Authority (equivalent to districts in the US) identifiers.

We focus our analysis on educational attainment at the end of the Reception year, when children are approximately 5 years old, because this is where we expect to see the clearest evidence of the effect of the entitlement. At the end of their first year in school, all children are assessed by their teacher in the different areas of learning covered by the Foundation Stage Profile (FSP) curriculum (Department for Education, 2012a and 2012b). This consists of 13 assessment scales,

each with a range between 1 and 9, grouped into six areas: personal, social and emotional development; communication, language and literacy; problem solving, reasoning and numeracy; knowledge and understanding of the world; physical development, and creative development. Children who score 6 points or above in all 13 scales are defined as "working within the Early Learning Goals", implying they are at least meeting the expected level of achievement. We will define them as working *at or above the expected level*. Children with a score of 9 in at least one of the scales are deemed to be working "beyond the ELGs", so will be categorized as working *beyond the expected level*. Finally, those with a score of 1 to 3 in at least one of the assessment scales will be classified as working *towards the expected level*. The assessments can also be summed up to give a total score of up to 117 points, but we will mainly focus on the threshold measures because they allow us to capture effects at different points in the ability distribution.

We also report results for the assessments obtained two years later, when children are aged 7 and reach the end of a part of the curriculum called Key Stage 1 (KS1).[13] These scores are given by the child's teacher, although students sit some tests to provide more information to base them on. Results are provided for Reading, Writing, Maths and Science in terms of levels (0, 1, 2c, 2b, 2a and 3). Children of this age are expected to reach level 2b, while level 3 is regarded as exceeding expectations. The levels can also be transformed into a total KS1 point score using a standard scoring system. Both the FSP and KS1 assessments are moderated within the Local Education Authority, providing quality assurance.

We have access to NPD extracts including date of birth for several cohorts of children who start school between the academic years 2008/09 and 2011/12. We use data on mainstream schools only, i.e. our sample excludes schools which cater exclusively for children with special needs. Because of the confidential nature of the data, we only have information on a subsample of all children from each cohort, including children born up to four weeks before and after 31st December and 31st March cut-offs. This means that we have information on children born in 16 weeks of the year, for a total sample size of 688,006, as shown Table A1 in the Appendix. From this sample we exclude duplicate cases and observations with missing information on the FSP scores (less than 1 percent of the sample), children born on the first day of the cut-off, [14] and children that attend schools with staggered school-starting policies where school entry coincides with the eligibility cut-off.[15]

[13] We are able to match about 97 percent of children between age 5 and age 7, due to children changing school sector or moving abroad.

[14] Our initial checks show that the proportion of children from non-White British families and the share of children who speak English as an additional language is very high among those born on January 1st. We think this is because some children from immigrant families are inaccurately registered as having a January 1st date of birth. As these children on average score lower on standard educational tests, including them in our analysis would lead us to overstate the effects of eligibility. We therefore exclude any child born on January 1st from our analysis and for balance we also exclude children born on April 1st.

[15] Most children in England start school in September after they turn 4 irrespective of their exact date of birth, but in the past it was fairly common for children born later in the school year to start school in the second or third term (i.e. in January or April, respectively). Since the free entitlement was introduced staggered school starts are found in fewer schools but where they persist they confound the impact of eligibility for free early education with length of formal schooling. We therefore use information on date of birth and date of enrolment to identify schools where a significant proportion (more than 30 percent) of children start in January or April. The 10% of children attending these schools are excluded from our analysis but in any case our results prove robust to their

5.2 Early Years Census and Ofsted Data

We merge children's school outcomes contained in the NPD data to the Early Years Census (EYC) which uses the same child identifiers as the NPD and contains data from the year before they start school for all children receiving the free entitlement in the private sector (the focus of our analysis).[16] Our sample of children attending private pre-school settings includes 284,544 children; as shown in Table A1, they make up about 47 percent of the total sample of children for whom we have a record of pre-school attendance.

From the EYC we have information on *teaching qualifications* and group size. Specifically, for all children attending pre-school education in the private sector, we have information on the number of staff who are qualified teachers (QTS) and who have Early Years Professional Status (EYPS). Questions on qualifications are asked with respect to all staff (including managers) and also more specifically about those carers working with the children who receive the free entitlement, i.e. teaching staff. We mostly use the variables that refer to teaching staff working with 3 and 4 year olds, but results are robust to broader definitions.[17] We also make use of information on the total number of staff and children to construct a ratio of 3 and 4 year olds per member of teaching staff. As well as being of interest in itself, this variable is important to isolate the effect of teacher qualifications from group size, as the two are mechanically linked through policy guidelines (see section 3). Due to missing information on some of these variables and measurement issues, we exclude pre-school settings which are very large or small, or who have a very large or small pupil to teacher ratios (7 per cent of observations), leading to a final sample of 265,679 observations. Table A1 shows step by step how we construct our sample.

Further, we link information on Ofsted ratings to our data. We have data on all assessments of private settings carried out by the regulator between 2005 and 2011, and we match each child to the rating for their setting that is closest in time to their attendance.[18] As well as providing an 1-4 (Outstanding to Inadequate) rating of overall effectiveness, the same categorical judgements are given for different sub-areas. These vary over time, but we can generate six fairly consistent categories, including: safety, helping to be healthy, encouraging to make a positive contribution, achieve and enjoy, teaching develops skills and effectiveness of leadership and management.[19] We focus on the last three areas, as they are more directly associated with educational attainment, but will explore variation in all six to generate a continuous measure of the Ofsted ratings which

in/exclusion.

[16] We we are able to successfully match over 93 percent of children observed in the first year of school to their EYC records (or, for those attending ECEC in the public sector their preschool NPD records) the year before. The remaining 6.5 percent of children for whom we have no record of pre-school attendance are in most cases children who never attended pre-school education.

[17] It is not clear a priori which variable should matter more, a graduate manager can set the tone for the whole setting, while a graduate in the room might affect the child's experience more directly.

[18] Our use of the Ofsted data is complicated by a change in the inspection regime in 2008. In the 2005-2008 cycle childcare settings that delivered the free entitlement were inspected against the criteria in the Curriculum Guidance for the Foundation Stage. Post-2008 all settings where judged on their delivery of the EYFS. Previous analysis in Blanden et al. (2017) indicates that this change does not matter, and we control for it in our analysis.

[19] Other sub-areas rated by Ofsted are not consistently considered across the years, and generally have a less obvious relationship with child development, examples are 'partnership with parents', 'safeguarding' and 'ability to deliver continuous improvement'.

ranges from 6 to 24, where 24 points implies an Outstanding judgement across all areas. We can match Ofsted ratings to 80 percent of children who attend pre-school in private sector settings. We include observations for which Ofsted ratings are missing in our analysis and use a dummy to distinguish them from the rest.

5.3 Descriptive Statistics

Our analysis considers the effect of eligibility on the educational outcomes of children who attend early education in the private sector. As explained above, this is because we do not have good measures of the quality of early education in public sector settings and within this sector there is generally less variability in teaching qualifications and ratios due to stricter policy rules. So, it is important to understand the representativeness of our sample with respect of the total population of children attending pre-school education. In Table A2 we present summary statistics for all the observable characteristics of children in our sample. In the first column we report results for all children attending ECEC the year before entering school, while the second column restricts the sample to children attending nurseries in the private sector. The main differences are by family and social background, with low-income children being less represented in the private sector. For instance, we observe 10 percent of the children in private settings are eligible for free school meals, while this percentage is 17.2 across both private and state sectors. Similarly, among children attending private settings, we have a lower percentage of pupils who speak English as an additional language and a higher percentage who are from a white British background, than in the general population.

Table A3 shows child outcomes at age 5. Here we also disaggregate our private settings sample by gender, as there are significant differences. We report the raw and standardised (using the overall year mean and standard deviation) total FSP scores, and the percentage of children working at or above, towards and beyond the expected level of achievement. We also add the percentage of children working at or above the expected level in literacy and numeracy, as these are core parts of the curriculum. Finally, we show the standardised sub-scores for the six areas of learning. Children who attend nurseries in the private sector have higher scores, on average, than all children attending pre-school. For example, the standardised total FSP score is 0.21 for children from private sector nurseries and 0.10 among all children in ECEC. However, this is not a clear indication that private sector nurseries are higher quality though, as we saw earlier that children attending these nurseries tend to come from less disadvanaged family backgrounds. As we would expect at age 5, girls out-perform boys in all outcome measures, with the gap being generally smaller in numeracy than in literacy.

Table A4 focuses on children attending private sector nurseries and provides information about our measures of quality. There is substantial variation across the dimensions of quality we consider, i.e. in terms of teacher qualifications and Ofsted ratings. The proportion of children in private settings with at least one Qualified Teacher is low, at 22 percent, while 12 percent have an Early Years Professional in the setting. This compares unfavourably with public sector settings, where settings always require a Qualified Teacher, and with the situation in most other countries

(Gambaro et al. 2014). In terms of the Ofsted ratings, 13 percent of children attend a setting rated Outstanding, with the majority of children being in settings rated Good (55 percent), 15 percent are in settings rated Satisfactory, and only 1.5 percent in settings rated Inadequate. In our analysis we therefore focus on the consequence of attending an Outstanding or Good setting compared to the combination of the other two categories. We will also show results using a continuous score as described at the end of section 5.2 (Ofsted overall score). Finally, we report the proportion of children in settings rated Outstanding in one of three sub-areas of interest. Here the variation is limited, with the proportions ranging from 13.8 to 16.4, and the subscores are highly collinear (e.g. 85.6 percent of settings rated Outstanding in Achievement are also rated Outstanding in Teaching), implying that it might be hard to distinguish exactly which aspect of the overall score is most significant.

5.4 Is Eligibility Randomly Assigned?

As is standard in RDD analyses, we need to check that date of birth in proximity of the cut-off is as good as randomly assigned. We start by plotting the distribution of our running variable (date of birth) either side of the two cut-off dates (31st December and 31st March). This is to investigate whether the policy determining entitlement to free part-time early education had any effect on the day on which a child was born. Parents who are aware of the importance of the eligibility rule (because they are well-informed or because they have an older child) might time the birth of their child to receive more free part-time child care. If so, we would see relatively more births in the days preceding the cut-off dates, and fewer births in the first few days afterwards. As noted in Section 4 this could invalidate the identification strategy as date of birth would be correlated with outcomes for reasons other than eligibility.

The first panel of Figure 1 plots the relationship between date of birth and number of children born on each day for the eight weeks around the December cut-off. The bold line shows the raw number of births on each day. While there is no apparent bunching of births before the cut-off we do see some non-random patterns. In particular there is a clear weekly pattern in the number of births with fewer occurring at weekends, and a sharp drop at Christmas. These patterns are likely to be driven primarily by the timing of planned caesarean sections and inductions away from weekends and holidays. We therefore plot residuals from a regression of the number of births on separate sets of dummies for being born on each day of the week, bank holidays and festivities (e.g. Christmas), and their interactions. The pattern of births is now much smoother over time with no relationship between the number of births and the cut-off. The same is true for the March eligibility cut-off shown in the second panel, where the smoothed line includes controls for Easter. In the remainder of the analysis we join the data for the two cut-offs and show how our results change without and with controls for the day of the week, bank holiday and festivities.

A second important check is whether observed individual characteristics are correlated with eligibility status. If births around the cut-off are randomly assigned, this should not be the case. We run regressions testing for the presence of a discontinuity in observable characteristics either side of the cut-off, using a specification similar to the one in equation (2), but where X_i is the

dependent variable and among the vector of controls we have only day of the week, bank holiday and festivity dummies and their interactions. We vary the way we control for date of birth, using different functional forms and show results including and excluding average setting-level characteristics. Results (shown in Table A5) reassure us that there is no significant effect of eligibility on child observable characteristics.

6 Eligibility Rules and Childcare Participation

In this section we provide evidence about the extent to which eligibility for the free entitlement leads parents to take up early education. We can only expect eligibility to affect educational outcomes if it leads to changes in behaviours. The size of the relationship between eligibility and childcare participation will help us to interpret our results on the impact of eligibility on educational outcomes.

We use the Family Resources Survey which is an annual cross-sectional survey of UK households with interviews running continuously throughout the year.[20] We use data from 2005-06 to 2012-13 and select children living in England. In the Family Resources Survey we can observe participation in early education at different points in time between birth and entry into school, and how nursery attendance varies by the rules governing eligibility, i.e. by time of birth. We observe the date of interview and the month of birth of the child, so that we can only define the child's age in months (rather than days). The fact that we do not know the child's precise date of birth and we have a much smaller sample size means we cannot use the same RD design we adopt for our main analysis, and instead rely on a difference-in-difference approach.

As shown in Equation (5), we model children's participation in ECEC (defined by the parent reporting they are cared for in a day nursery or pre-school in the reference week) as a function of their term of birth (Autumn, Spring or Summer, denoted by TOB_i) and their eligibility (T_{it}), where the latter is defined by the age of the child at interview (eligibility takes value 1 if the child is observed after becoming eligible for the free entitlement and 0 otherwise). We then construct interactions between term of birth and eligibility. The coefficient on these interactions (γ) represents the impact of the free entitlement on participation for each group of children - as defined by their term of birth – when they are old enough to benefit. In all regressions we also control for date of interview (month and year) and some family characteristics, see notes to Table 1. As in our main analysis, it is important that we control for a flexible function of age at interview ($f(a_{it})$), as children will be more likely to attend ECEC as they become older, independent of their eligibility status. This means that if we use a very short window of data (say children between 30 and 40 months of age) our eligibility variable might simply capture the effect of age at interview. In order to disentangle the effect of eligibility from the effect of age, we include in our regression children from a wide age spectrum (i.e. from 12 to 59 months) and control for a flexible function of the child's age in months. Our basic results explore the difference

[20] The National Pupil Database and Early Years Census provide information on children's attendance at nursery on a census day in the academic year before children start school (i.e. age 4), but does not show how it varies as children become eligible.

made by changing both these margins, so we use samples of children observed at variable age-intervals from 12 to 59 months, and control for age in both a quadratic and cubic function for each sample.

$$P_{it} = \alpha + \beta TOB_i + \gamma TOB_i * T_{it} + f(a_{it}) + \Pi X_{it} + \psi Month_{it} + \varphi Year_{it} + u_{it} \quad (5)$$

Table 1 shows our main results. Eligibility for the free entitlement increases the use of childcare by 11 to 17 percentage points for the Spring-born, 10 to 18 percentage points for the Summer-born, and 10 to 16 percentage points for the Autumn-born (these coefficients are not statistically different from each other). Our specifications also include a dummy for the term before a child becomes eligible to capture anticipation effects. It is possible that families are prepared to enter their child into an early education setting a few months before the child becomes eligible, perhaps in order to take advantage of available spaces. We expect this effect to be larger for children born in the Autumn term, who become eligible in January but might anticipate this by attending in the September of the year before, at the start of the academic year. Indeed, we find that Autumn-born children experience an increase of about 6-11 percentage points in ECEC attendance the term before eligibility. This implies that the treatment effect for these children may be as much as two terms of additional early education and care.

We might expect families to respond to eligibility in ways other than a change in attendance at the extensive margin. Table A6 provides estimates of the impact of eligibility on weekly hours in subsidisable childcare, time spent in informal care and on weekly spend. Eligibility increases the average hours used per week by about 2 (higher for Summer-borns and no effect for Spring-borns) while decreasing the time spent in informal care by a similar amount. It decreases expenditure on childcare by £8 per week for the Spring and Summer born, indicating that the effects are quite small.[21]

7 Regression Discontinuity Results

7.1 The impact of an additional term of eligibility

As a first piece of descriptive evidence about the impact of eligibility for an extra term of ECEC on educational attainment at age 5, Figure 2 plots measures derived from Foundation Stage Profile (FSP) scores either side of the eligibility cut-off (note that we pool the December and March cut-offs), adjusted for day of the week, bank holiday, and festivity effects and for average differences across schools (this is particularly important as assessments are conducted by teachers). For each

[21] We also investigate whether the effect of eligibility on attendance at ECEC is heterogeneous according to family income or maternal educational qualification. We find some suggestive evidence that the effects are slightly larger for lower income families and less educated mothers, a result that is consistent with the evidence in Blanden et al. (2016). However, the precision of our estimates, based on a much smaller sample size than the one available through the NPD, is such that we cannot exclude that the effects are similar across all groups. Results are available on request.

outcome we plot the average value of the residual outcome measure (solid dots) for all children and interpolate these points, allowing the slopes of the lines to be different before and after the cut-off.

The Figure shows that a linear association between date of birth and outcomes matches the data well in this small window around the cut-off, although we will check for non-linearities in our regression analyses. Discontinuities at the cut-off are visible for the total FSP score, the categorical variable which measures whether children are working at or above the expected level overall, and (most clearly) the categorical variable which measures whether children are working at or above the expected level in literacy. These effects appear to be small, however.

In our regression analysis we run five specifications of our main model for each outcome. First we build up the set of controls. All models include school fixed effects, individual Xs, the number of children in the nursery attended the year before starting school, and control for a linear function of age. We then add controls for being born on each week day, a bank holiday or a festive day and their interaction (our equation (2)). Last, we include the mean characteristics of the other children in the setting (our equation (3)). We then check the sensitivity of the results to controlling for more flexible functions of date of birth using a quadratic term and a linear term which is allowed to change at the cut-off point. Level differences in outcomes between children born around the 31st of December and children born around the 31st March are captured by a dummy in all models.[22]

The results for the outcomes in Figure 2 are shown in Table 2. Estimation is by linear regression for continuous variables (such as the FSP standardised score), while for the categorical variables we run linear probability models. Standard errors are clustered by date of birth and school.[23] As we can see, the estimates are slightly sensitive to the controls included, but not at all sensitive to the functional form used to control for age. Evidence from the range of specifications shown indicates small but significant positive effects of eligibility on the probability that children are working at or above the expected level overall and at or above the expected level in literacy. The effects on literacy are slightly stronger. Eligibility to an extra term of free entitlement raises the probability of achieving the expected level in literacy by just under 1 percentage point. Table 3 runs the same specifications for other FSP outcomes and confirms the positive and (weakly) statistically significant result for literacy using a continuous measure. Positive and slightly larger effects are also found for the creative development scale.

Figure 3 shows the sensitivity of our estimates to varying the data window around the cut-off. Our sample includes children born 4 weeks either side of December 31st and March 31st, and we show in this figure how the point estimate (bold line) and confidence intervals (lighter line) vary when using data from 1 to 4 weeks, adding one day at a time. The figure shows that the estimate

[22] In analysis not shown we interact day of birth with the 31st March dummy, but this interaction is never statistially significant. We also interacted the eligibility effect with the 31st March dummy to check whether the effects are different for children becoming eligible at different points in time over the school year, but we could not generally reject the hypothesis that the effects are the same.

[23] We do not adjust our standard errors for multiple hypothesis testing as all the dependent variables are measures of the same underlying outcome.

for the impact of eligibility on achieving the expected level in the overall FSP or the expected level in literacy vary quite substantially when using data on children born only a few days after the cut-off. These estimates are also generally larger and the confidence intervals are wider. This suggests that it would be very hard to be precise about the effect of eligibility by using a very short data window around the cut-off due to the difficulty of disentangling age and eligibility effects with very few data points. The figure, however, makes it clear that the estimates become much more stable and robust when using at least 2 weeks of data, and do not change much at all after 3 weeks.

Table A8 in the appendix shows results for a range of outcomes at the end of Key Stage 1 (age 7).[24] Here the magnitude of the coefficients on eligibility is much smaller and no statistically significant effects are found. This is in line with much of the literature (Schweinhart et al. 2005, Deming, 2009, Garces et al., 2002, Elango et al., 2016) which finds a rapid fade-out of early years' interventions. Two possible caveats are in order here though. First, we need to point out that the assessments provided at the end of the Reception year, the FSP scores, take into account a broad range of skills, including creative thinking and social and emotional development. By contrast, the assessments carried out at age 7, the KS1 scores, are more narrowly focused on the academic subjects Mathematics, English and Science. Second, a recent literature argues that the effects of early interventions develop over time and may become clearer towards the adult years, so an insignificant result at age 7 (i.e. 3-4 years after the treatment) may not tell us much about the long-term effects (Elango et al., 2016).

The parameters we report in our tables are all intention-to-treat effects. Combining these results with those from the previous section on the impact of eligibility on participation suggests that *attending* ECEC for an additional term as a consequence of the policy leads to around a 3.4-6.2 percentage point increase in the probability of working at or above the expected level for the overall FSP score (compared to the mean of 60 per cent), and between a 5.1 and 9.3 percentage point increase in the probability of meeting the expected level in literacy (compared to the mean of 68 percent).[25] Compared to the results in Cornelissen and Dustmann (*forthcoming*), our estimates are undoubtedly smaller. They estimate the effect on age 5 educational achievement of each month of full-time education at age 4 to be be in the order of 6-9 percent of a standard deviation which would suggest effects in the order of 20-30 percent of a standard deviation for our 3.5 month treatment. The difference in results could be explained in at least two ways; one possibility is that the benefits of part-time attendance are not the same as those of full-time attendance, or alternatively the quality of ECEC is not comparable to the quality of compulsory education. This should not be surprising, given the pay and status differential between staff in nurseries and those in schools (Gambaro et al., 2014, Bonetti, 2019).

Table 4 shows results where we consider meeting the expected level overall and add interactions

[24] Table A7 presents descriptive statistics for these outcomes.

[25] These estimates are obtained by dividing estimates of the impact of eligibility on outcomes from column (3) in Table 1 (0.006 for at or above the overall expectation and 0.009 for the threshold in literacy) by the proportion of children who participate as a result of becoming eligible. This is obtained from Table 1. We use the lower and upper bound estimates of 0.097 and 0.176.

with the child characteristics that are available in the NPD (gender, free school meals eligibility, deprivation of the neighborhood in tertiles, language spoken at home, and ethnicity). There are striking results for gender: for all outcomes considered the benefits of attending an additional term are entirely experienced by boys with no significant effects for girls. This is in contrast to evidence from early targeted interventions that finds larger effects for girls (Garcia et al. 2018, Elango et al. 2016, Havnes and Mogstad 2011) but consistent with newer evidence for universal programmes (Blanden et al. 2016, Cornelissen and Dustmann, forthcoming, Cornelissen et al., 2018 and Leuven et al. 2010). Also, there is no evidence that an additional term spent in childcare is more beneficial for children from disadvantaged backgrounds as measured by free lunch eligibility and deprivation in the neighbourhood of residence (results in Blanden et al, 2016, indicate that effects of the policy roll out are slightly larger for disadvantaged families, but not statistically different).

The results in Table 4 are also relevant to our strategy to assess the effect of quality. As previously noted, if the effect of eligbility varied by social background we might confound this with variations in effects of elibility by setting quality, casting doubt on the causal interpretation of the results that follow. There is limited evidence that this is the case, which is reassuring.

7.2 Does attending a nursery of higher quality have larger benefits?

We now turn to the key question of this paper, that is whether there is a significant interaction between eligibility for an additional term in early pre-school education and the quality of the setting attended. Our regression models follow equation (4). We start by looking at staff qualifications, which can be considered measures of *structural* quality, and focus initially on working *at or above the expected level* in the overall FSP score as the main outcome of interest, although we will also show results for other outcomes in the Appendix.

First, we look at the share of graduates working with 3 and 4 year olds within a setting. This includes Qualified Teachers and Early Years Professionals and continues to be cited in policy circles as a key quality criterion (Department for Education, 2017, Nursery World, 2018). Note that when adding this variable to our model we must control for the number of 3 and 4 year olds per teaching staff (group size) to isolate the effect of qualifications, because regulations permit lower staff-child ratios when there is more highly qualified staff.

Table 5 shows our baseline results in column (1). Column (2) adds the share of graduates to the estimation. There is a positive association between the share of graduates and children working at or above the expected level in the FSP, but this is not statistically different from zero. To evaluate whether the share of graduates has an impact on the benefit of an extra term in childcare we interact this variable with our indicator for eligibility. Under the assumption that sorting into settings of different quality is controlled for, this interaction gives the causal effect of spending the additional term in early education in a setting with a higher share of graduate staff. That is, it measures whether the quality of the setting increases (or reduces) the overall benefit of the extra term. Results from the interaction are displayed in column (3) and show a negative point estimate which is not statistically different from zero, suggesting that there is no additional

benefit of being entitled to an extra term of part-time early education in a setting with a higher proportion of graduate staff.

The absence of any effect of better staff quality on children's outcomes is unlikely to be driven by sorting into settings based on unobservable characteristics. If anything, we would expect positive sorting into settings by quality, with children that have unobserved characteristics that make them high achievers sorting into better quality settings. Any such sorting would bias our estimates upwards, not downwards.

In Section 3 we explained that the group of graduates is quite diverse, with qualified teachers (QTS) benefiting from much longer training than Early Years Professionals (EYP), who can obtain their qualification in as little as four months. We therefore look separately at these groups. Column (4) shows that there is a positive and statistically significant association between the share of QTS in a setting and our outcome, but this is not so for the share of EYP. To investigate whether there is a causal effect of entitlement to an extra term in nurseries with higher shares of QTS and EYP, respectively, we again enter interactions with eligibility in our estimation. As can be seen in column (5), there is no effect of either QTS or EYP interactions on working at or above the expected level in the overall FSP. We check whether results are different if we use a binary indicator for whether any member of staff working with 3 and 4 year olds has that qualification instead of shares of staff with a certain qualification. Results are shown in column (6) and again reveal no benefit of the extra term being spent in nurseries with higher quality in terms of staff qualifications. In Table A9 we show results that check for impacts on other outcomes, including the standardised total FSP score and the threshold measures used earlier.[26] Across all of these outcomes we find that higher staff qualifications do not have any effect on the benefit from being eligible for an additional term of childcare.

Next we turn to the effect of setting quality as measured by the rating awarded by the national regulator Ofsted, which is closer to a measure of *process* quality. As before, we first investigate the association between Ofsted rating and children working at or above the expected FSP level at age 5. Table 6 shows this in column (2), whereas column (1) reports our baseline finding. The coefficients on the indicators for Outstanding and Good ratings show that there is a positive and statistically significant association between ratings and child outcomes, where Satisfactory and Inadequate are the combined omitted category. Adding an interaction between Ofsted rating and eligibility in column (3) shows that those children who attend Outstanding settings have an additional benefit from eligiblity to an extra term; it increases their probability of working at the expected FSP levels by 1.3 percentage points (2.3 per cent of the mean) compared to children in lower quality settings. This effect is 2-3 times larger than the baseline effect of the extra term (0.005 in this column) and, if we are prepared to assume that any observed and unobserved selection into settings of a different quality is taken into account by the average characteristics of children in the setting, $\overline{X_j}$, and the association between Ofsted rating and outcomes, Q_j, then this effect gives us a causal estimate of how the eligibility effect varies with the quality of pre-

[26] The impact of eligbility on creative development is only present for the continuous version of this score, so we focus on the threshold for literacy and numeracy which deliver more robust findings throughout.

school education. The sum of the coefficients on eligibility and its interaction with quality (that is, coefficients β and ϕ in equation (4)) gives the total effect of receiving the extra term in a setting of a particular quality. This effect is 1.8 percentage points in settings rated Outstanding, that is a 3 percent increase from the mean across all children in private settings. In contrast to Outstanding nurseries, there is no additional effect of being eligible to attend nurseries rated Good for an extra term.[27]

We repeat the analysis for another way of measuring Ofsted ratings, using the continuous measure which adds up scores on the six sub-areas (see Section 5). Columns (4) and (5) of Table 6 confirm that children attending a setting with a better rating do better, and there is (weak) evidence that spending more time in a better setting is beneficial.

Table A10 shows results for the other FSP outcomes. Here it is interesting to see that the interaction effect between quality (Outstanding) and eligibility is also statistically significant for high level attainment (working beyond the level expected in the FSP), but not for the other measures considered here.

Given that we find sizeable effects of attending childcare settings rated Outstanding for longer, we want to explore what is driving these results. We do this by looking at sub-area ratings that are potentially closely related to child cognitive outcomes, including 'children achieve and enjoy' (Achievement), 'teaching develops skills' (Teaching) and 'effectiveness of leadership and management' (Leadership), see Section 5.

In Table 7 we show results of regressions where we enter indicator variables for settings achieving Outstanding in each of the sub-area ratings, then we introduce interactions of these with the eligibility dummy, and finally we show whether these results are sensitive to adding additional controls for teacher qualifications and group size, i.e. measures of *structural* quality.[28] Columns (1), (4) and (7) of Table 7 show the results for entering the sub-area Outstanding dummies. For all three sub-areas an Outstanding rating is positively associated with children working at or above expected FSP levels. Interactions with the eligibilty indicator, shown in columns (2), (5) and (8), reveal that there is a statistically significant effect of Outstanding Leadership and (more weakly) Teaching on the outcomes for children eligible for an extra term, but the effect of eligibility is not larger in settings rated Outstanding in Achievement. These results remain unaffected by controlling for staff qualifications (see columns (3), (6) and (9)), suggesting that our measures of structural quality are not driven by differences in staff qualifications. Table A11 shows results for other outcomes at age 5 and focuses on the Teaching scores. We see here that the impact of spending an extra term in an Outstanding setting for Teaching also extends to *working beyond the expected leve*l, similar to our results for the headline Ofsted grade.

[27] Altonji and Mansfield (2018) note that controlling for mean characteristics of children at the setting level might result in a lower bound on the setting characteristics estimates if families sort according to these characteristics. We have therefore checked the robustness of our results to leaving out these controls, and find that the estimates change only marginally. For example, when mean characteristics are excluded, the coefficient on the interaction term between being in an Outstanding nursery and eligibility is 0.014 (instead of 0.013) for achieving the expected level in the FSP, and is identical to the one shown in Table A10 column 3 for achieving a score beyond the expected level. Similarly, the results in Table 6 for structural quality indicators change only marginally if at all.

[28] We enter the Ofsted sub-area ratings one by one because they are collinear.

Our findings in Table 4 show that all impacts of eligibility are found for boys. We might therefore suspect that the impact of being eligible to an extra term in an Outstanding nursery might also be restricted to boys. To investigate this we include in our model triple interactions between gender of the child, quality and eligibility. This obviously reduces the number of observations per cell which are used for the identification of the effects, so we expect some loss of precision. Results in Table A12 show that the magnitude of the effect of being entitled to an extra term of part-time early education in an Outstanding setting are larger for boys than girls, with a coefficient of 0.016 and 0.02 for achieving a score at or above the expected level for the overall FSP and for literacy, respectively. While these coefficients are not statistically significantly different from zero, they are by far the largest effects observed in any of our models and point out once again that the gains of an extra term are not gender-neutral.[29]

7.3 Robustness and sensitivity checks

In this section we present robustness checks and check the sensitivity of our results to sample restrictions. We focus on the main estimates of interest, i.e. those showing the interaction between eligibility status and quality of the setting as measured by Ofsted scores, which indicate (as shown in Table 8 column 1) that children eligible for an additional term in part-time early education are 1.3 percentage points more likely to score at or above the expected level for the overall FSP and 1.5 percentage points more likely to go beyond the expected level if they attend a high quality setting. First, we run a placebo test where we use an arbitrary cut-off date to define eligibility status to check whether our results are unique to eligibility cut-off dates, and cannot be found at other arbitrary dates. This is shown in column (2) of Table 8, where we have set the cut-off to January 15th and April 15th, and use observations on children born 2 weeks around these dates. The point estimate for achieving a score at or above the expected FSP level is now 0.002 and this is not statistically different from zero. Similarly, the coefficient for achieving a score beyond the expected level is 0.004, less than a third of the size of our baseline estimate, and again not statistically significant. This indicates that the cut-off associated with eligibility has explanatory power for educational outcomes that is not shared by other, arbitrary, dates.

However, to be able to attribute the effect we observe to the entitlement at age 3 we need to make sure it is not confounded by school starting dates. As explained previously, term of birth also affects some children's date of entry into compulsory schooling. We exclude from our main sample all schools where a significant minority of children start school in January or April, i.e. schools which appear to have different starting dates for children with different dates of birth.[30] Although starting date policies vary at the level of school and not district, in column (3) we more conservatively exclude all local authorities (or school districts) where a significant proportion of children (10 percent) start school during different terms over the year. This leads us to exclude

[29] A stronger effect of quality on boys is consistent with results found in Bauchmüller et al. (2014), Datta Gupta and Simonsen (2010) and Dynarski et al., (2013).

[30] Specifically we exclude schools where more than 30 percent of children born in January start school during the second term. In preliminary checks we did not notice any sensitivity of the results to using slightly different criteria.

a further 9.1 percent of observations from our sample. Our results are robust to this restriction and - if anything - the point estimates become larger.

Next we check whether our results could be contaminated by another policy implemented at the same time. During the period analysed here, the government introduced a new subsidy for the poorest 2 year olds in some pilot areas of the country (Smith et al., 2009). This intervention was also made available in the term after the child's birthday, so positive effects might be counfounding the impact of the three year old entitlement. To check for this we introduce in our regression a variable indicating the amount spent by each district on the 2-year-old subsidy, normalized by the number of children in the district to control for the effect of the pilot for 2 year olds.[31] Column (4) of Table 8 shows that this makes little to no difference to our estimates.

The remaining two columns of Table 8 restrict the estimation sample in two ways. Column (5) presents estimates of the interaction between eligibility and setting quality when excluding the last cohort of children from our sample who were exposed to a slightly higher number of hours of free early education (15 as opposed to 12.5) and more flexibility for parents when to take these hours (e.g. could choose to have them all in 2 days rather than spread them over the week). Coefficients are very similar to those shown in the main results, albeit a bit less precisely estimated. Column (6) excludes London from our analysis as educational attainment has followed a different trend from that seen in other parts of the country in the last decade (Blanden et al., 2015). Effects are not driven by London.

8 Conclusion

This study moves beyond standard evaluations of universal ECEC programmes to provide evidence on the impact of the quality of pre-school setting on children's educational outcomes. This is crucial given the widespread conviction that high quality programmes are necessary for child development, and the relatively scant evidence on what a high quality setting looks like and how it can be achieved. Specifically, we consider whether the benefits of an additional term of entitlement to ECEC are larger in higher quality settings, focusing on measures of quality that are part of the existing regulatory framework in the UK. In counterpoint to much of the policy discussion about quality we find no evidence that extra time spent in nurseries with highly qualified workers is beneficial for any of the outcomes we observe. However, we find that spending more time in a setting rated highly by the national regulator Ofsted improves children's chances of achieving both expected and higher levels of attainment.

These findings resonate with the literature on school quality which emphasises that teacher practice matters but finds it difficult to demonstrate the observable characteristics of teachers which lead to better student outcomes. Similarly, our finding that the effectiveness of leadership and management in pre-school matters for later achievement mirrors results found for school management practices (Bloom et al., 2015). Few other papers have found a significant link be-

[31] We do not have information on the number of places offered to 2 year olds, only about district spending. We also used a dummy for the districts that implemented the pilot, and the results were very similar to the ones shown here.

tween the effectiveness of pre-school and the drivers of quality in compulsory schooling. The two are generally treated rather separately, and our work implies that further consideration of these parallels might be beneficial to both literatures. Our research indicates that regulating staff qualification (especially at higher levels) might not provide a clear route to improving quality. However, the evidence about the value of inspections is more encouraging. The Ofsted 'Outstanding' rating that we focus on is rather a black box, meaning that further investigation is necessary to more fully understand the ingredients that drive quality. However, our results imply that, with careful consideration, countries should be able to provide regulation and inspection regimes which support the high quality provision that children need to flourish.

References

[1] Altonji, J. G., & Mansfield, R. K. (2018). Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects. *American Economic Review,* 108(10), 2902-46.

[2] Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo and N. Schady. (2016). Teacher Quality and Learning Outcomes in Kindergarten. *Quarterly Journal of Economics,* 1415–1453.

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

[4] Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, 25-50.

[5] Bauchmüller, R., Gortzb, M., Rasmussen, A. (2014). Long-run benefits from universal high-quality preschooling. *Early Childhood Research Quarterly,* 29, 457–470

[6] Blanden J., Del Bono E, McNally S, Rabe B. (2016). Universal pre-school education: the case of public funding with private provision. *The Economic Journal*, 126 (May), 682-723.

[7] Blanden, J., Greaves, E., Gregg, P., Macmillan, L. and Sibieta, L. (2015). Understanding the improved performance of disadvantaged pupils in London schools. Social Policy in a Cold Climate Working Paper No. 21, Centre for the Analysis of Social Exclusion, London School of Economics.

[8] Blanden, J., Hansen, K. and McNally, S. (2017). Quality in Early Years Settings and Children's School Achievement. Centre for Economic Performance (CEP) Discussion Paper No.1468, London School of Economics.

[9] Bloom, N., Lemos, R. Sadun, R. and Van Reenen, J. (2015). Does Management Matter in Schools? *The Economic Journal,* 125(May), 647-674.

[10] Bonetti, S. (2019). The Early Years Workforce in England. Education Policy Institute. Accessed on 3rd April 2019: https://epi.org.uk/publications-and-research/the-early-years-workforce-in-england/.

[11] Campbell, T., Gambaro, L. and Stewart, K. (2018). 'Universal' early education: Who benefits? Patterns in the take-up of the entitlement to free early education for three-year-olds in England. *British Educational Research Journal,* 44(3), 515-538.

[12] Cascio, E. (2015). The promises and pitfalls of universal early education. IZA World of Labor, Bloomsbury, London. Accessed on 3rd April: https://wol.iza.org/articles/promises-and-pitfalls-of-universal-early-education/long.

[13] Cascio, E. (2017). Does Universal Preschool Hit the Target? Program Access and Preschool Inputs. NBER Working Paper 23215.

[14] Cascio, E., and Schanzenbach, D. W. (2013). The Impacts of Expanding Access to High-Quality Preschool Education. *Brookings Papers on Economic Activity*, 2013(2), 127-192.

[15] Conti. G., Heckman, J. and Pinto, R. (2016). The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour. *The Economic Journal*, 126(596), F28-F65

[16] Cornelissen, T. and Dustmann, C. (Forthcoming). Early School Exposure, Test Scores and Noncognitive Outcomes. *American Economic Journal: Applied Economics.*

[17] Cornelissen, T., Dustmann, C., Raute, A. and Schoenburg. U. (2018). Who benefits from universal childcare? Estimating marginal returns from early childcare attendance. *Journal of Political Economy,* 126(6) 2356-2409.

[18] Crawford, C., Dearden, L. and Greaves, E. (2014). The drivers of month-of-birth differences in children's cognitive and non-cognitive skills. *Journal of the Royal Statistical Society Series A*, 177(4), 829–860.

[19] Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1 (3), 111– 26 134.

[20] Dobie, W. and Fryer, R. (2013) Getting behind the veil of Effective Schools: Evidence from New York City *American Economic Journal: Applied Economics* 5(4), 28-60.

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

[22] Department for Education (2011) The Tickell Review of the Early Years Foundation Stage. London: Department for Education.

[23] Department for Education (2012a) Statutory Framework for the Early Years Foundation Stage, Nottingham.

[24] Department for Education (2012b) Early Years Foundation Stage Profile Results in England 2011-2012Statistical First Release 23, Nottingham.

[25] Department for Education (2013). More Great Childcare: Raising Quality and Giving Parents More Choice. London: Department for Education.

[26] Department for Education (2015) Childcare Bill: Policy Statement. London: Department for Education.

[27] Department for Education (2017). Early Years Workforce Strategy. London: Department for Education.

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

[29] Early, D., K. Maxwell M. Burchinal S. Alva, R. H. Bender, D. Bryant, K. Cai, R. Clifford, C. Ebanks, James A. Griffin G. Henry, C. Howes, J. Iriondo-Perez, H-J. Jeon, A. Mashburn, E. PeisnerâĂŘFeinberg, R. Pianta N. Vandergrift, N. Zill (2007).Teachers' Education, Classroom Quality, and Young Children's Academic Skills: Results From Seven Studies of Preschool Programs. *Child Development,* 78(2), 558-580.

[30] Elango, S., J. L Garcia, J. Heckman, A. Hojman (2016). Early Childhood Education. In *Economics of Means-Tested Transfer Programs in the United States*, Volume II ed. R. Moffitt. The University of Chicago Press, Chicago.

[31] Elicker, J., Langill, C., & colleagues. (2011). Evaluation of Paths to QUALITY, Indiana's Child Care Quality Rating and Improvement System: Final Report. West Lafayette, IN: Purdue University.

[32] Felfe, C., Nollenberger, N., and Rodriguez-Planas, N. (2015). Can't buy mommy's love? Universal childcare and children's long-term cognitive development. *Journal of Population Economics*, 28(2), 393-422.

[33] Fitzpatrick, M. (2008). Starting School at Four: The Effect of Universal Pre-Kindergarten on Children's Academic Achievement. *Journal of Economic Analysis and Policy*, vol. 8(1), 1-40.

[34] Gambaro, Ludovica and Stewart, Kitty and Waldfogel, Jane, eds. (2014) An equal start? Providing high quality early childhood education and care for disadvantaged children. CASE Studies on Poverty, Place and Policy. Policy Press, Bristol, UK.

[35] Gambaro, L., Stewart, K., and Waldfogel, J. (2015). A question of quality: do children from disadvantaged backgrounds receive lower quality early childhood education and care? *British Educational Research Journal*, 41(4), 553-574.

[36] Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. American Economic Review, 92 (4), 999–1012.

[37] Garcia, J. L., Heckman, J and A. Ziff (2018). Gender differences in the benefits of an influential early childhood program. *European Economic Review,* 109, 9-22.

[38] Gormley, W. T., and Gayer, T. (2005). Promoting school readiness in Oklahoma an evaluation of Tulsa's pre-k program. *Journal of Human Resources*, 40(3), 533-558.

[39] Haeck, C., L. Lebihan and P. Merrigan (2018) Universal Child Care and Long-Term Effects on Child Well-Being: Evidence from Canada *Journal of Human Capital,* 2(1), 38-98.

[40] Harms, T., Clifford, R. M., and Cryer, D. (2014). Early childhood environment rating scale. Teachers College Press.

[41] Havnes, T., and Mogstad, M. (2011). No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes, *American Economic Journal: Economic Policy*, 3(2), 97–129.

[42] Heckman, J., Moon, S., Pinto, R., Savelyev, P., and Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1), 114-128.

[43] Herbst C. and Tekin E. (2010). Child Care Subsidies and Child Development.*Economics of Educatio Review,* 29, 618–638.

[44] Hobbs, G. and Vignoles, A. (2010). Is free school meal eligibility a good proxy for family income?, *British Educational Research Journal*, 36(4), 673–90.

[45] Hopkin, R., Stokes, L., & Wilkinson, D. (2010). Quality, outcomes and costs in early years education. National Institute of Economic and Social Research. Mimeo.

[46] Imbens, G. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615-635.

[47] Jensen, B., Holm, A., and Bremberg, S. (2013). Effectiveness of a Danish early year preschool program: A randomized trial. *International Journal of Educational Research*, 62, 115–128.

[48] Karoly, L. A., Kilburn, M. and Cannon, J. (2006). Early childhood interventions: Proven results, future promise. Rand Corporation.

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

[50] Mathers, S., Ranns, H., Karemaker, A., Moody, A., Silva, K., Graham, J. and Siraj-Blatchford, I. (2011). Evaluation of the Graduate Leader Fund: Final report. London: Department for Education (DfE).

[51] Mathers, S., and Smees, R. (2014). Quality and Inequality: Do three-and four-year olds in deprived areas experience lower quality early years provision. Nuffield Foundation.

[52] Mathers, S., Sylva, K. and Joshi, H. (2007). Quality of Childcare Settings in the Millennium Cohort Study. Research Report SSU/2007/FR/025. Retrieved from DfES website: http://www.dfes.gov.uk/research/data/uploadfiles/SSU2007FR025 per centREV.pdf.

[53] Mathers, S., Singler, R. and Karemaker, A. (2012). Improving quality in the early years: a comparison of perspectives and measures. University of Oxford.

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

[55] National Audit Office (2016). Entitlement to Free Education and Childcare. London: The Stationery Office.

[56] National Research Council (2001). Eager to Learn: Educating Our Preschoolers. Committee on Early Childhood Pedagogy. B. Bowman, M.S. Donovan, and M.S. Burns, editors. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

[57] Nursery World (2018). Alarm over Government axing of early years graduate plans. 24th July 2018.Accessed 3rd April 2019: https://www.nurseryworld.co.uk/nursery-world/news/1165111/alarm-over-government-axing-of-early-years-graduate-plans.

[58] OECD (2015). *Starting Strong IV: Monitoring Quality in Early Childhood Education and Care.*Accessed 3rd April 2019: http://www.oecd.org/publications/starting-strong-iv-9789264233515-en.htm.

[59] Ofsted (2015) Early Years Inspection Handbook. Manchester: Ofsted.

[60] Pianta, R., LaParo K., and Hamre B. (2007). *Classroom Assessment Scoring System—CLASS.* Baltimore: Brookes.

[61] Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child-care quality to children's cognitive and social development trajectories through second grade. *Child Development*, 72, 1534–1553.

[62] Rivkin, S., Hanushek, E. and Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73, 417–458.

[63] Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94, 247–252.

[64] Sabol, T., Soliday Hong, S., Pianta, R. and Burchinal, M. (2013) Can Rating Pre-K Programs Predict Children's Learning? *Science,* 341, 845-846.

[65] Schweinhart, L.J., Montie, J., Xiang, Z., Barnett, W.S., Belfield, C.R., & Nores, M. (2005). Lifetime effects: The High/Scope Perry Preschool study through age 40. Ypsilanti, MI: High/Scope Press.

[66] Siraj-Blatchford, I, Sylva, K., Laugharne, J., Milton, E. and Charles, F (2005). Monitoring and Evaluation of the Effective Implementation of the Foundation Phase (MEEIFP) Project across Wales. Final Report of Year 1 Pilot - Roll Out Age 3-5 Years. Available from here http://dera.ioe.ac.uk/6281/1/meeifp.pdf last accessed 24th August 2018.

[67] Smith, R., Purdon, S., Schneider, V., La Valle, I., Wollny, I., Owen, R. Bryson, C., Mathers, S., Slyva, K., and Lloyd, E. (2009). *Early Education Pilot for Two Year Old Children: Evaluation*. Department of Children Schools and Families Research Report RR134.

[68] Slyva, K., I. Siraj-Blatchford, B. Taggart, E. Melhuish, K. Elliot,V. Totsika. (2006). Capturing Quality in Early Years through Environmental Rating Scales *Early Childhood Research Quarterly,* 21, 76-92.

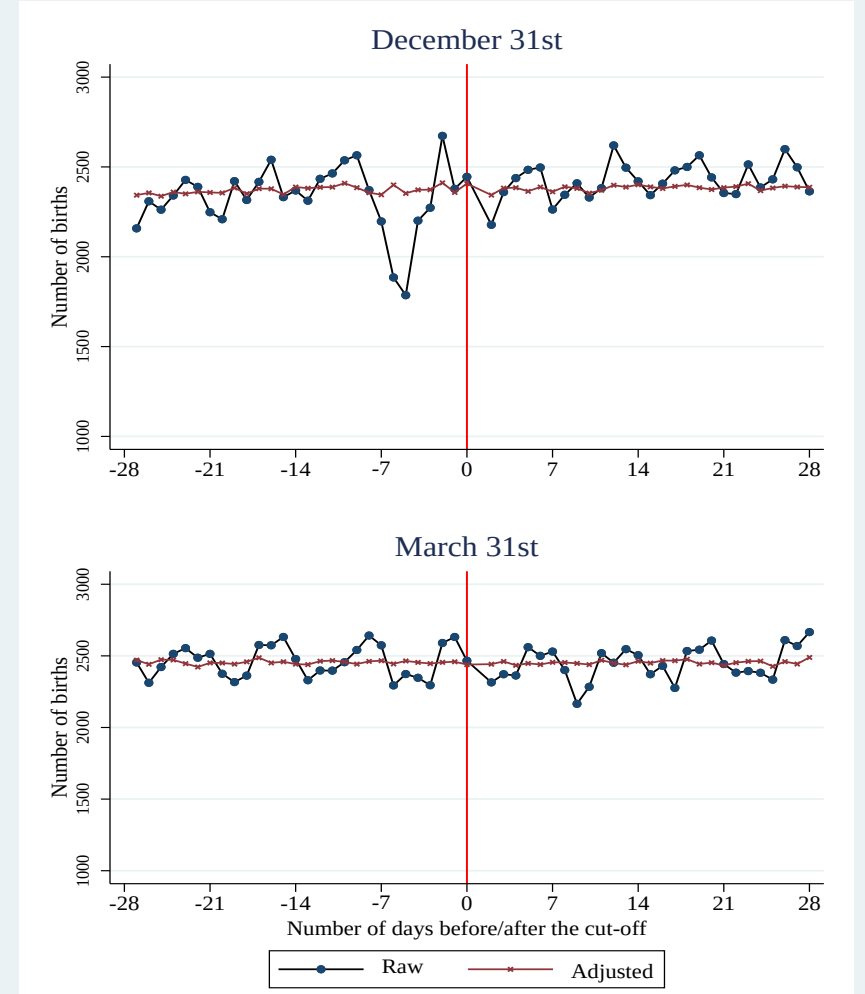
[69] Thistlethwaite, D. and Campbell, D. (1960). Regression-discontinuity analysis: an alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), 309-317.

[70] Ulferts, H., and Anders, Y. (2016). Effects of ECEC on academic outcomes in literacy and mathematics: Meta-analysis of European longitudinal studies. Accessed on 6th March 2019: http://ecec-care.org/resources/publications/#c30167.

[71] Walters, C. (2015). Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start. *American Economic Journal: Applied Economics*, 7(4), 76-104.

[72] Zellman, G. and Perlman, M. (2008). Child-care quality rating and improvement systems in five pioneer states: Implementation issues and lessons learned. Santa Monica, CA: RAND Corporation.

Figure 1: Distribution of births around the cut-off



Notes: Each point represents the total number of children born on each day before or after the relevant cut-off (31st December and 31st March). The line with a circular marker represents the unadjusted total number of births, the line with a cross marker represents the residual of a regression of number of births on separate dummies for days of the week, bank holidays, and festivities (e.g. Christmas), as well as interactions between days of the week and bank holidays and between days of the week and festivities.

Figure 2: Effect of eligibility on Foundation Stage Outcomes

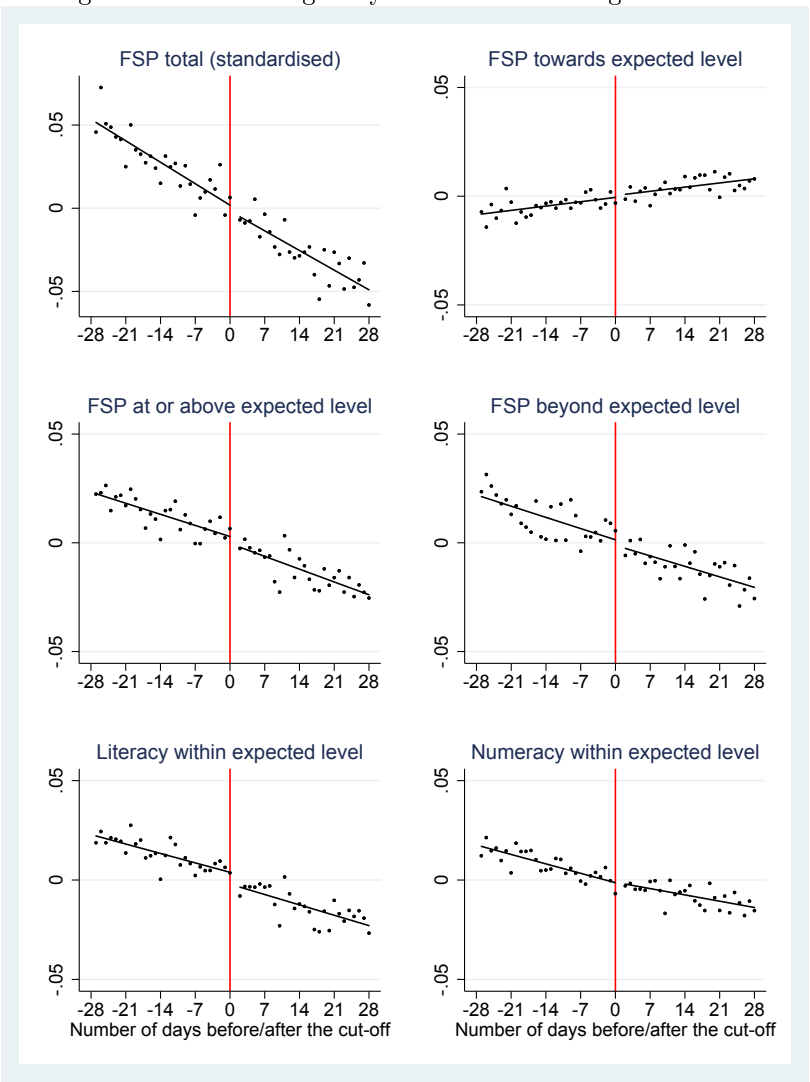


Figure 3: Estimates by size of the data window around the cut-off

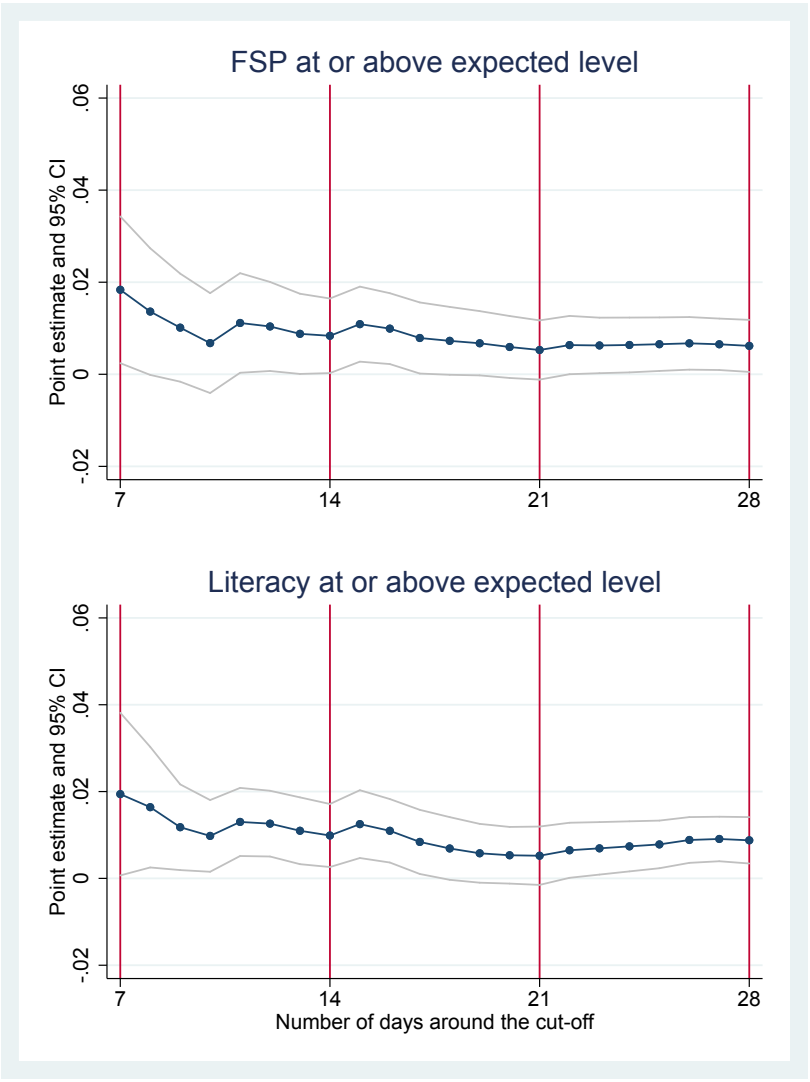


Table 1: Impact of eligibility: use of subsidisable childcare

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age range	12-59 months	12-59 months	24-59 months	24-59 months	12-47 months	12-47 months	24-47 months	24-47 months
Age (in months) controls	Quadratic	Cubic	Quadratic	Cubic	Quadratic	Cubic	Quadratic	Cubic
Summer born	0.010	0.010	0.011	0.011	0.009	0.010	0.011	0.011
	(0.013)	(0.013)	(0.018)	(0.018)	(0.013)	(0.013)	(0.020)	(0.020)
Autumn born	-0.003	-0.002	-0.014	-0.014	-0.004	-0.003	-0.016	-0.016
	(0.013)	(0.013)	(0.019)	(0.019)	(0.014)	(0.014)	(0.020)	(0.020)
Spring born - eligible	0.169**	0.109**	0.107**	0.113**	0.141**	0.118**	0.121**	0.121**
	(0.022)	(0.027)	(0.028)	(0.030)	(0.030)	(0.031)	(0.033)	(0.042)
Summer born - eligible	0.166**	0.105**	0.0967**	0.103**	0.176**	0.155**	0.144**	0.144**
	(0.019)	(0.026)	(0.026)	(0.028)	(0.029)	(0.030)	(0.031)	(0.041)
Autumn born - eligible	0.163**	0.107**	0.107**	0.112**	0.125**	0.102**	0.116**	0.116**
	(0.021)	(0.026)	(0.027)	(0.029)	(0.029)	(0.030)	(0.031)	(0.041)
Spring born - anticipation	0.023	-0.010	-0.038	-0.037	0.029	-0.002	-0.022	-0.021
	(0.031)	(0.033)	(0.034)	(0.034)	(0.034)	(0.035)	(0.037)	(0.040)
Summer born - anticipation	0.034+	-0.002	0.007	0.008	0.018	-0.015	0.005	0.005
	(0.020)	(0.023)	(0.024)	(0.024)	(0.023)	(0.025)	(0.026)	(0.030)
Autumn born - anticipation	0.109**	0.0788**	0.0569*	0.0566*	0.111**	0.0793**	0.0695*	0.0696*
	(0.025)	(0.026)	(0.028)	(0.028)	(0.027)	(0.028)	(0.031)	(0.034)
Observations	15275	15275	11377	11377	11497	11497	7599	7599

Source: Family Resources Survey. Notes: Coefficents from a linear regression of use of subsidisable childcare on eligibility to free part-time education and care. Eligibility is defined as a dummy with value 1 if the interview takes place when the child is observed in the school term after turning 3 years old. Anticipation is captured by a dummy which assumes value 1 the term before the child becomes eligible to free pre-school education. Other controls not shown include indicators for interview month and survey year, age a polynomial in the child's age in months (quadratic or cubic), educational qualifications of the main carer and partner, if present, number of siblings and whether the child is the youngest child in the family. Heteroskedasticy robust standard errors in parenthesis. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table 2: Impact of eligibility: discontinuity on educational outcomes at age 5

	(1)	(2)	(3)	(4)	(5)
Functional form for day of birth	Linear	Linear	Linear	Quadratic	Kinked
School FE	√	√	√	√	√
Individual controls	√	√	√	√	√
Day of the week, bank holiday and festivity controls		√	√	√	√
Mean of individual controls at setting level			√	√	√
FSP total score (standardised)	0.007	0.008	0.008	0.008	0.006
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
FSP working **at or above** the expected level	0.004	0.006+	0.006*	0.006*	0.005+
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
FSP working **beyond** the expected level	0.004	0.004	0.004	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
FSP working **towards** the expected level	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
FSP working **at or above** the expected level in **literacy**	0.007**	0.008**	0.009**	0.009**	0.008**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
FSP working **at or above** the expected level in **numeracy**	0.001	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Observations	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: Each coefficient represents the effect of being entitled to an extra term of free part-time education and care on the outcome of interest and is obtained from a separate regression. FSP total score is the sum of scores in all the 13 areas of assessment reported as part of the Foundation Stage Profile (see text). Literacy is shorthand for Communication, Language and Literacy and is the sum of scores in these three areas of the FSP assessment. Numeracy is shorthand for Problem solving, Reasoning and Numeracy and is the sum of scores in these three areas of the FSP assessment. Children are classified as working within the expected level if they score 6-9 in all the areas of the related FSP assessment. Children are classified as working beyond the expected level if they achieve 9 in at least one of the FSP assessment areas. Children are classified as working towards the expected level if they have a score between 1 and 3 in at least one of the FSP assessment areas. Day of birth, bank holiday and festivity controls include interactions between the "festivity" and the day of the week it falls on. Additional controls include sex of the child, free school meal status, ethnicity, whether they speak English as an additional language, the deprivation of the area where they live measured by the decile of the neighbourhood of residence on the Income Deprivation Affecting Children (IDACI) scale, the average number of children in the pre-school setting, and a dummy for being part of the March subsample. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table 3: Impact of eligibility: discontinuity on educational outcomes at age 5; FSP subscales

	(1)	(2)	(3)	(4)	(5)
Functional form for day of birth	Linear	Linear	Linear	Quadratic	Kinked
School FE	√	√	√	√	√
Individual controls	√	√	√	√	√
Day of the week, bank holiday and festivity controls		√	√	√	√
Mean of individual controls at setting level			√	√	√
Standardised sub-scores:					
FSP literacy	0.010+	0.010+	0.011+	0.011+	0.009
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
FSP numeracy	0.008	0.007	0.007	0.007	0.005
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
FSP social and emotional development	-0.002	0.000	0.000	0.000	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
FSP knowledge and understanding of the world	0.009+	0.010+	0.009	0.009	0.008
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
FSP creative development	0.014*	0.014*	0.014*	0.014*	0.013*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
FSP physical development	-0.000	0.000	0.001	0.001	-0.001
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: See notes to Table 2. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table 4: Heterogeneity in the impact of eligibility on being at or above the expected level in the FSP at age 5

	Boys	FSM	Deprivation level of local area		English as an additional language	Non-white
	(1)	(2)	(3)		(4)	(5)
			Middle deprivation areas	High deprivation areas		
Eligibility	0.002	0.006*	0.007*		0.006+	0.006*
	(0.004)	(0.003)	(0.003)		(0.003)	(0.003)
Eligibility* Interaction	0.008*	0.002	-0.001	-0.001	-0.000	-0.005
	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)	(0.004)
Observations	265679	265679	265679		265679	265679

Source: National Pupil Database and Early Year Census. Notes: The coefficients shown represent the effect of eligibility and eligibility interacted with the related individual characteristics (each group of coefficients represents a separate regression) on being at or above the expected level in the FSP. We also allow for an interaction between the dummy for being part of the March subsample (not shown) with observable individual characteristics. The model includes a linear function of day of birth, school fixed-effects, and the mean of individual pupil characteristics at the setting level. Additional individual controls as in Table 2 column 3. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table 5: Impact of eligibility on working at or above the expected level at 5: interactions with staffing variables

	FSP working at or above the expected level					
	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Share of staff with specified qualification working with 3 and 4 year olds				Any staff with specified qualification working with 3 and 4 year olds
Eligibility	0.006*	0.006*	0.007*	0.006*	0.007*	0.008*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Graduates (QTS + EYP)		0.008	0.011			
		(0.006)	(0.008)			
Eligibility*Graduates			-0.006			
			(0.009)			
Qualified Teachers (QTS)				0.017*	0.023*	0.008*
				(0.007)	(0.010)	(0.004)
Eligibility*QTS					-0.012	-0.006
					(0.012)	(0.004)
Early Years Professionals (EYP)				0.001	-0.002	-0.003
				(0.008)	(0.010)	(0.005)
Eligibility*EYP					0.005	0.000
					(0.013)	(0.006)
Number of 3 and 4 year olds per teaching staff		√	√	√	√	√
Observations	265679	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: Each column reports the results of a regression of the outcome (top) on the independent variables shown (left). The model includes a linear function of day of birth, school fixed-effects, and the mean of individual pupil characteristics at the setting level. Additional controls as in Table 2 column 3. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table 6: Impact of eligibility on working at or above the expected level at age 5: interactions with quality ratings

	FSP working at or above the expected level				
	(1)	(2)	(3)	(4)	(5)
Eligibility	0.006*	0.006*	0.005	0.006*	-0.004
	(0.003)	(0.003)	(0.005)	(0.003)	(0.007)
Ofsted Outstanding		0.021**	0.014*		
		(0.004)	(0.006)		
Eligibility * Outstanding			0.013*		
			(0.006)		
Ofsted Good		0.008**	0.006		
		(0.003)	(0.004)		
Eligibility *Good			0.003		
			(0.005)		
Ofsted continuous score				0.002**	0.001**
				(0.000)	(0.000)
Eligibility*Ofsted continuous score					0.001+
					(0.001)
Ofsted rating missing		√	√	√	√
Ofsted new regime		√	√	√	√
Observations	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: Each column reports the results of a regression of the outcome (top) on the independent variables shown (left). We also control for a dummy for Ofsted rating missing and its interaction with eligibility, plus a dummy to indicate whether the Ofsted inspection took place under the new regime and its interaction with eligibility (not shown). The model includes a linear function of day of birth, school fixed-effects, and the mean of individual pupil characteristics at the setting level. Additional controls as in Table 2 column 3. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table 7: Impact of eligbility on working at or above the expected level at age 5: interactions with more detailed quality ratings

	FSP working at or above the expected level								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Eligibility	0.006*	0.007+	0.008+	0.006*	0.007+	0.008+	0.006*	0.007+	0.008+
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Ofsted Outstanding in Achievement	0.015**	0.011**	0.011**						
	(0.003)	(0.004)	(0.004)						
Eligibility*Outstanding in Achievement		0.007	0.007						
		(0.004)	(0.004)						
Ofsted Outstanding in Leadership				0.012**	0.007+	0.007+			
				(0.003)	(0.004)	(0.004)			
Eligibility*Outstanding in Leadership					0.011*	0.011*			
					(0.005)	(0.005)			
Ofsted Outstanding in Teaching							0.017**	0.012**	0.012**
							(0.003)	(0.004)	(0.004)
Eligibility*Outstanding in Teaching								0.009+	0.009+
								(0.005)	(0.005)
Share of Qualified Teachers (QTS)			√			√			√
Eligibility*Share of QTS			√			√			√
Share of Early Years Professionals			√			√			√
Eligibility*Share of EYP			√			√			√
Number of 3 and 4 year olds per teaching staff			√			√			√
Ofsted rating missing	√	√	√	√	√	√	√	√	√
Ofsted new regime	√	√	√	√	√	√	√	√	√
Observations	265679	265679	265679	265679	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: Each column reports the results of a regression of the outcome (top) on the independent variables shown (left). We also control for a dummy for Ofsted rating missing and its interaction with eligibility, plus a dummy to indicate whether the Ofsted inspection took place under the new regime and its interaction with eligibility (not shown). The model includes a linear function of day of birth, school fixed-effects, and the mean of individual pupil characteristics at the setting level. Additional controls as in Table 2 column 3. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table 8: Robustness checks

	Baseline results for Ofsted Outstanding interaction	Placebo: 2 weeks either side 15th Jan/March	Excluding LEAs with different school starting dates	Controlling for pilot on 2 year olds	Excluding last year of data; school starters in 2011-12	Excluding London
	(1)	(2)	(3)	(4)	(5)	(6)
FSP total score (standardised)	0.002	0.008	0.002	0.002	0.000	0.004
	(0.010)	(0.017)	(0.010)	(0.010)	(0.012)	(0.011)
FSP working **within** the expected level	0.013*	0.002	0.015*	0.013*	0.014+	0.013+
	(0.006)	(0.009)	(0.006)	(0.006)	(0.007)	(0.007)
FSP working **beyond** the expected level	0.015*	0.004	0.016*	0.015*	0.011+	0.017**
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
FSP working **towards** the expected level	0.001	0.001	0.001	0.001	-0.002	0.002
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
FSP working **within** the expected level in **literacy**	0.005	0.004	0.005	0.005	0.007	0.004
	(0.006)	(0.010)	(0.006)	(0.006)	(0.007)	(0.006)
FSP working **within** the expected level in **numeracy**	-0.001	0.012+	-0.001	-0.001	0.001	-0.003
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	265679	130499	265679	(0.006)	196962	243142

Source: National Pupil Database and Early Year Census. Each column reports the results of a regression of the outcome (left) on a model including Ofsted ratings dummies (Outstanding and Good) and their interactions with eligbility to an extra term of free part-time education and care. The coefficient shown is the coefficient of the interaction between Ofsted rating Outstanding and eligibility. All other variables are as specified in Table 6. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Additional Tables

Table A1: Sample selection

	% sample excluded	Observations remaining
National Pupil Database		
All children born 4 weeks before and after 31st December or 31st March		688,006
Excluding children born on 1st January and 1st April	1.74	676,039
Excluding observations with duplicates	0.00	676,021
Excluding observations with missing information on FSP scores	0.00	675,368
Excluding children in schools which may admit children based on date of birth	10.00	607,735
Merging with the Early Year Census		
Excluding children not in a pre-school setting the year before starting school	6.52	568,135
Excluding children in state sector	53.18	284,544
Excluding missing observations for nursery characteristics, very large or small nuseries, or nuseries with very large or very small child/staff ratios	6.63	265,679

Notes: Sample selection criteria used in determining the sample used in our analysis and obtained through merging the National Pupil Database and the Early Year Census.

Table A2: Descriptive statistics for child characteristics

	All children observed in ECEC	**Children in private settings**
Male	0.511	0.513
Free School Meals (FSM)	0.172	0.100
Not FSM	0.826	0.897
FSM missing	0.003	0.002
White British	0.661	0.750
Non White British	0.229	0.151
Ethnicity missing	0.110	0.099
EAL (English as Additional Language)	0.138	0.067
Not EAL	0.723	0.784
EAL missing	0.138	0.149
Least deprived third by area	0.335	0.464
Middle deprived third by area	0.326	0.341
Most deprived third by area	0.312	0.163
Deprivation missing	0.026	0.032
Number of 3 and 4 years old in the setting	n.a.	33.260
		(14.29)
Observations	568135	265679

Source: National Pupil Database and Early Year Census. Notes: Means and standard deviations (for continuous outcomes only) of individual characteristics.

Table A3: Descriptive statistics for outcome variables

	All children observed in ECEC	**Children in private settings**	**Girls in private settings**	**Boys in private settings**
FSP total score	88.92	90.90	93.37	88.55
	(15.43)	(14.53)	(13.22)	(15.31)
FSP total score (standardised)	0.097	0.212	0.357	0.0745
	(0.900)	(0.850)	(0.773)	(0.896)
FSP working **at or above** the expected level	0.560	0.600	0.695	0.509
FSP working **beyond** the expected level	0.262	0.298	0.346	0.252
FSP working **towards** the expected level	0.109	0.0875	0.0587	0.115
FSP working **at or above** the expected level in **literacy**	0.642	0.678	0.768	0.592
FSP working **at or above** the expected level in **numeracy**	0.770	0.810	0.837	0.784
FSP sub-scores (standardised)				
FSP literacy	0.096	0.200	0.366	0.0423
	(0.918)	(0.877)	(0.801)	(0.916)
FSP numeracy	0.094	0.204	0.253	0.158
	(0.896)	(0.839)	(0.773)	(0.895)
FSP social and emotional development	0.071	0.169	0.340	0.00736
	(0.930)	(0.895)	(0.805)	(0.945)
FSP knowledge and understanding of the world	0.089	0.208	0.218	0.198
	(0.916)	(0.854)	(0.806)	(0.896)
FSP creative development	0.077	0.182	0.430	-0.0535
	(0.926)	(0.894)	(0.791)	(0.922)
FSP physical development	0.096	0.181	0.321	0.0481
	(0.918)	(0.881)	(0.763)	(0.962)
Observations	568135	265679	129398	136281

Source: National Pupil Database and Early Year Census. Notes: Means and standard deviations (for continuous outcomes only) of educational outcomes at age 5 (FSP) and age 7 (KS1). FSP total score is the sum of scores in all the 13 areas of assessment reported as part of the Foundation Stage Profile (see text). Literacy is shorthand for Communication, Language and Literacy and is the sum of scores in these three areas of the FSP assessment. Numeracy is shorthand for Problem solving, Reasoning and Numeracy and is the sum of scores in these three areas of the FSP assessment. Children are classified as working within the expected level if they score 6-9 in all the areas of the related FSP assessment. Children are classified as working beyond the expected level if they achieve 9 in at least one of the FSP assessment areas. Children are classified as working towards the expected level if they have a score between 1 and 3 in at least one of the FSP assessment areas.

Table A4: Setting quality

	Children in private settings	
	mean	st. dev.
Staff Qualification		
Share of graduates among teaching staff	0.091	0.201
Share of Qualified Teachers among teaching staff	0.058	0.153
Share of Early Years Professionals among teaching staff	0.036	0.135
Any Qualified Teachers among teaching staff	0.218	
Any Early Years Professionals among teaching staff	0.120	
Number of 3 and 4 year olds per teaching staff	6.007	4.866
Ofsted Rating		
Ofsted rating: Outstanding	0.130	
Ofsted rating: Good	0.548	
Ofsted rating: Satisfactory	0.153	
Ofsted rating: Inadequate	0.013	
Ofsted rating: Missing	0.156	
Ofsted overall score	10.988	5.645
Ofsted Subscores		
Ofsted Outstanding in Achievement	0.164	
Ofsted Outstanding in Leadership	0.138	
Ofsted Outstanding in Teaching	0.146	
Observations	265679	

Source: National Pupil Database and Early Year Census. Notes: Means and standard deviations (for continuous outcomes only) of setting characterics.

Table A5: Impact of eligibility on observable characteristics

	(1) Linear	(2) Quadratic	(3) Kinked	(4) Linear	(5) Quadratic	(6) Kinked
Including mean of children characteristics at setting level				√	√	√
Male	-0.001	-0.001	-0.001	-0.002	-0.002	-0.002
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)
Free school meals	0.005+	0.005+	0.005+	0.005+	0.005+	0.005+
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Non White British	-0.003	-0.003	-0.003	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
EAL (English as Additional Language)	-0.002	-0.002	-0.002	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
First decile of deprivation	0.000	0.000	0.000	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Fifth decile of deprivation	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Tenth decile of deprivation	0.000	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Number of 3 and 4 years old in the setting	-0.278+	-0.277+	-0.284+	-0.226	-0.225	-0.231
	(0.141)	(0.141)	(0.146)	(0.254)	(0.258)	(0.261)
Share of graduates among teaching staff	0.002	0.002	0.002	0.002	0.002	0.002
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Share of Qualified Teachers among teaching staff	0.002	0.002	0.002	0.002+	0.002+	0.002+
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of Early Years Professionals among teaching staff	0.000	0.000	0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number of 3 and 4 year olds per teaching staff	-0.068+	-0.067+	-0.069+	-0.062	-0.061	-0.063
	(0.037)	(0.036)	(0.037)	(0.039)	(0.041)	(0.042)
Ofsted rating: Outstanding	-0.006+	-0.006+	-0.006+	-0.005	-0.006+	-0.006+
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Ofsted rating: Good	0.004	0.004	0.004	0.001	0.001	0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Ofsted rating: Satisfactory	0.001	0.001	0.001	0.001	0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Ofsted rating: Inadequate	-0.001	-0.001	-0.001	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ofsted overall score	-0.060	-0.061	-0.065	-0.083	-0.083	-0.087
	(0.068)	(0.068)	(0.069)	(0.067)	(0.066)	(0.067)
Observations	265679	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: Each coefficient represents the effect of being entitled to an extra term of free part-time education and care on the outcome of interest and is obtained from a separate regression. Controls are included for a linear trend in day of birth, day of week of birth, being born on a festivity day and the interaction between these. Individual controls included are sex of the child, free school meal status, ethnicity, whether they speak English as an additional language, the deprivation of the area where they live measured by the decile of the neighbourhood of residence on the Income Deprivation Affecting Children (IDACI) scale, the average number of children in the pre-school setting, and a dummy for being part of the March subsample. The average of these characteristics within each setting are included where indicated. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table A6: Impact of eligibility: other outcomes

	Weekly hours in subsidisable childcare	Weekly hours in informal childcare	Weekly spend on subsidisable childcare
	(1)	(2)	(3)
Summer born	0.148	-0.509	-1.082
	(0.436)	(0.494)	(1.974)
Autumn born	-0.256	-0.506	-1.181
	(0.449)	(0.508)	(2.030)
Spring born - eligible	0.860	-2.996**	-8.260+
	(0.937)	(1.062)	(4.240)
Summer born - eligible	2.138*	-1.975+	-8.501*
	(0.910)	(1.031)	(4.115)
Autumn born - eligible	1.729+	-2.426*	-4.630
	(0.910)	(1.031)	(4.117)
Spring born - anticipation	-0.719	-1.929+	-1.806
	(0.884)	(1.001)	(3.998)
Summer born - anticipation	-0.662	-0.744	0.988
	(0.674)	(0.764)	(3.048)
Autumn born - anticipation	-0.124	-0.660	0.110
	(0.752)	(0.853)	(3.405)
Observations	7599	7599	7599

Source: Family Resources Survey. Notes: See notes to Table 1. Specification as in Table 1 column (8). Heteroskedasticy robust standard errors in parenthesis. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table A7: Descriptive statistics for Key Stage 1 outcomes

	All children observed in ECEC	**Children in private settings**	**Girls in private settings**	**Boys in private settings**
KS1 total score	63.10	65.34	65.31	63.43
	(12.02)	(11.74)	(11.19)	(12.18)
KS1 total score (standardised)	0.057	0.160	0.241	0.0834
	(1.005)	(0.982)	(0.936)	(1.019)
KS1 working **at or above** the expected level	0.655	0.690	0.759	0.625
KS1 working **beyond** the expected level	0.587	0.629	0.689	0.572
KS1 working **towards** the expected level	0.345	0.310	0.241	0.375
KS1 working **at or above** the expected level in **reading**	0.807	0.833	0.877	0.794
KS1 working **at or above** the expected level in **mathematics**	0.805	0.833	0.846	0.820
Observations	551125	257475	125549	131926

Source: National Pupil Database and Early Year Census. The KS1 total score is obtained summing up the points achieved in each of the KS1 subscores for Reading, Writing, Mathematics and Science. Children are said to be meeting the expected level if they achieve a level 2/2b in all areas, they are working towards the expected level if they do not achieve a 2/2b across the board. Children are defined as working beyond the expected level if they do better than level 2/2b in at least one assessment.

Table A8: Impacts of eligibility: Key Stage 1 outcomes

	(1)	(2)	(3)	(4)	(5)
Functional form for day of birth	Linear	Linear	Linear	Quadratic	Kinked
School FE	√	√	√	√	√
Individual controls	√	√	√	√	√
Day of the week, bank holiday and festivity controls		√	√	√	√
Mean of individual controls at setting level			√	√	√
KS1 total score (standardised)	0.006	0.005	0.006	0.006	0.005
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
KS1 working **at or above** the expected level	-0.001	-0.000	0.000	0.000	-0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
KS1 working **beyond** the expected level	-0.004	-0.002	-0.001	-0.001	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
KS1 working **towards** the expected level	0.001	0.000	-0.000	-0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
KS1 working **at or above** the expected level in **reading**	0.001	0.001	0.002	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
KS1 working **at or above** the expected level in **mathematics**	-0.004	-0.004	-0.003	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	257475	257475	257475	257475	257475

Source: National Pupil Database and Early Year Census. Notes: Each coefficient represents the effect of being entitled to an extra term of free part-time education and care on the outcome of interest and is obtained from a separate regression. The KS1 total score is obtained summing up the points achieved in each of the KS1 subscores for Reading, Writing, Mathematics and Science. Children are said to be meeting the expected level if they achieve a level 2/2b in all areas, they are working towards the expected level if they do not achieve a 2/2b across the board. Children are defined as working beyond the expected level if they do better than level 2/2b in at least one assessment. Additional controls are as in Table 2 column 3. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table A9: Impact of eligibility on other FSP outcomes: interactions with staffing variables

	FSP total score (standardised)	FSP working at or above the expected level	FSP working beyond the expected level	FSP working towards the expected level	FSP working at or above the expected level in literacy	FSP working at or above the expected level in numeracy
	(1)	(2)	(3)	(4)	(5)	(6)
Eligibility	0.009	0.007*	0.005	-0.003	0.009**	0.001
	(0.006)	(0.003)	(0.004)	(0.002)	(0.003)	(0.002)
Share of Qualified Teachers (QTS)	0.043*	0.023*	0.024**	-0.009	0.020*	0.018*
	(0.017)	(0.010)	(0.009)	(0.007)	(0.009)	(0.008)
Eligibility*Share of QTS	-0.006	-0.012	-0.007	0.001	-0.008	-0.011
	(0.022)	(0.012)	(0.011)	(0.008)	(0.012)	(0.010)
Share of Early Years Professionals (EYP)	0.019	-0.002	-0.003	0.002	-0.005	-0.013+
	(0.018)	(0.010)	(0.009)	(0.007)	(0.010)	(0.007)
Eligibility*Share of EYP	-0.017	0.005	0.004	-0.000	0.003	0.015
	(0.023)	(0.013)	(0.011)	(0.009)	(0.013)	(0.011)
Number of 3 and 4 year olds per teaching staff	√	√	√	√	√	√
Observations	265679	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: Each column reports the results of a regression of the outcome (top) on the independent variables shown (left). The model includes a linear function of day of birth, school fixed-effects, and the mean of individual pupil characteristics at the setting level. Additional controls as in Table 2 column 3. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table A10: Impact of eligibility on other FSP outcomes: interactions with quality ratings

	FSP total score (standardised)	FSP working at or above the expected level	FSP working beyond the expected level	FSP working towards the expected level	FSP working at or above the expected level in literacy	FSP working at or above the expected level in numeracy
	(1)	(2)	(3)	(4)	(5)	(6)
Eligibility	0.006	0.005	0.001	-0.005	0.010*	0.002
	(0.011)	(0.005)	(0.006)	(0.004)	(0.005)	(0.005)
Ofsted Outstanding	0.037**	0.014*	0.017**	-0.004	0.017**	0.012**
	(0.009)	(0.006)	(0.004)	(0.003)	(0.006)	(0.004)
Eligibility * Outstanding	0.002	0.013*	0.015*	0.001	0.005	-0.001
	(0.010)	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)
Ofsted Good	0.016*	0.006	0.009**	-0.005*	0.007+	0.008*
	(0.006)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)
Eligibility *Good	0.001	0.003	0.005	0.002	0.001	-0.002
	(0.007)	(0.005)	(0.005)	(0.003)	(0.005)	(0.004)
Ofsted rating missing	√	√	√	√	√	√
Ofsted new regime	√	√	√	√	√	√
Observations	265679	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: Each column reports the results of a regression of the outcome (top) on the independent variables shown (left). We also control for a dummy for Ofsted rating missing and its interaction with eligibility, plus a dummy to indicate whether the Ofsted inspection took place under the new regime and its interaction with eligibility (not shown). The model includes a linear function of day of birth, school fixed-effects, and the mean of individual pupil characteristics at the setting level. Additional controls as in Table 2 column 3. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table A11: Impact of eligbility on other FSP outcomes: Ofsted sub-score for quality of teaching

	FSP total score (standardised)	FSP working at or above the expected level	FSP working beyond the expected level	FSP working towards the expected level	FSP working at or above the expected level in literacy	FSP working at or above the expected level in numeracy
	(1)	(2)	(3)	(4)	(5)	(6)
Eligibility	0.008	0.008+	0.005	-0.004	0.012**	0.001
	(0.008)	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)
Ofsted Outstanding in Teaching	0.031**	0.012**	0.010*	-0.002	0.014**	0.008**
	(0.007)	(0.004)	(0.004)	(0.002)	(0.004)	(0.003)
Eligibility*Outstanding in Teaching	-0.003	0.009+	0.010*	0.001	0.001	-0.001
	(0.008)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
Share of Qualified Teachers (QTS)	√	√	√	√	√	√
Eligibility*Share of QTS	√	√	√	√	√	√
Share of Early Years Professionals (EYP)	√	√	√	√	√	√
Eligibility*Share of EYP	√	√	√	√	√	√
Number of 3 and 4 year olds per teaching staff	√	√	√	√	√	√
Ofsted rating missing	√	√	√	√	√	√
Ofsted new regime	√	√	√	√	√	√
Observations	265679	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: Each column reports the results of a regression of the outcome (top) on the independent variables shown (left). We also control for a dummy for Ofsted rating missing and its interaction with eligibility, plus a dummy to indicate whether the Ofsted inspection took place under the new regime and its interaction with eligibility (not shown). The model includes a linear function of day of birth, school fixed-effects, and the mean of individual pupil characteristics at the setting level. Additional controls as in Table 2 column 3. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.

Table A12: Impact of eligibility: interactions with Ofsted ratings and gender

	FSP total score (standardised)	FSP working at or above the expected level	FSP working beyond the expected level	FSP working towards the expected level	FSP working at or above the expected level in literacy	FSP working at or above the expected level in numeracy
	(1)	(2)	(3)	(4)	(5)	(6)
Eligibility	0.000	0.005	-0.002	-0.005	0.008	0.003
	(0.013)	(0.008)	(0.007)	(0.004)	(0.008)	(0.007)
Eligibility*Male	0.010	-0.000	0.006	-0.001	0.005	-0.002
	(0.017)	(0.010)	(0.009)	(0.006)	(0.010)	(0.009)
Ofsted Outstanding	0.027*	0.020**	0.019**	0.003	0.018**	0.004
	(0.011)	(0.007)	(0.007)	(0.004)	(0.006)	(0.006)
Ofsted Good	0.009	0.009+	0.012*	-0.002	0.006	0.001
	(0.008)	(0.005)	(0.005)	(0.003)	(0.005)	(0.004)
Ofsted Outstanding * Male	0.021	-0.012	-0.003	-0.013*	-0.003	0.016*
	(0.016)	(0.009)	(0.009)	(0.006)	(0.009)	(0.008)
Ofsted Good * Male	0.014	-0.005	-0.005	-0.006	0.001	0.013*
	(0.012)	(0.007)	(0.007)	(0.005)	(0.007)	(0.006)
Eligibility * Ofsted Outstanding	0.003	0.003	0.018*	0.001	-0.004	0.001
	(0.015)	(0.009)	(0.009)	(0.005)	(0.009)	(0.007)
Eligibility * Ofsted Good	0.005	0.004	0.013	0.006	-0.004	-0.002
	(0.015)	(0.010)	(0.009)	(0.005)	(0.009)	(0.008)
Eligibility * Ofsted Outstanding * Male	-0.003	0.020	-0.007	0.001	0.016	-0.005
	(0.022)	(0.013)	(0.012)	(0.008)	(0.013)	(0.011)
Eligibility * Ofsted Good * Male	-0.005	0.012	-0.002	-0.005	0.003	-0.000
	(0.017)	(0.010)	(0.009)	(0.006)	(0.010)	(0.009)
Observations	265679	265679	265679	265679	265679	265679

Source: National Pupil Database and Early Year Census. Notes: Each column reports the results of a regression of the outcome (top) on the independent variables shown (left). We also control for a dummy for Ofsted rating missing and its interaction with eligibility and gender, plus a dummy to indicate whether the Ofsted inspection took place under the new regime and its interaction with eligibility and gender (not shown). The model includes a linear function of day of birth, school fixed-effects, and the mean of individual pupil characteristics at the setting level. Additional controls as in Table 2 column 3. Standard errors are clustered by day of birth and school. Symbols: + indicates p<.10, * indicates p<0.05, ** indicates p<0.01.