2012

The impact of pre-school on adolescents' outcomes: evidence from a recent english cohort

Patricia Apps
University of Sydney

Silvia Mendolia
University of Wollongong, smendoli@uow.edu.au

Ian Walker
Lancaster University

Publication Details

Apps, P., Mendolia, S., Walker I.,(2012). The Impact of Pre-school on Adolescents' Outcomes: Evidence from a Recent English Cohort. IZA Discussion Paper No. 6971, 30.

The impact of pre-school on adolescents' outcomes: evidence from a recent english cohort

Abstract
This paper investigates the relationship between attendance at nursery school and children's outcomes in adolescence. In particular, we are interested in child cognitive development at ages 11, 14 and 16, intentions towards tertiary education, economic activity in early adulthood, and in a group of non-cognitive outcomes, such as risky health behaviours (smoking, early pregnancy, use of cannabis) and personality traits (feelings and commitments about school; psychological well-being). Using matching methods to control for a very rich set of child's and family's characteristics, we find that pre-school childcare largely improves results in cognitive tests at age 11 and 14 and 16, and has a positive effect on intentions towards further education and economic activity at age 19-20. Positive effects are especially noticeable for children coming from disadvantaged socio-economic backgrounds. Results on non-cognitive outcomes are more mixed: we do not find any evidence of improvement in psychological well-being, but we do find some positive effects on health behaviours.

Disciplines
Business | Social and Behavioral Sciences

Publication Details
Apps, P., Mendolia, S., Walker I.,(2012). The Impact of Pre-school on Adolescents' Outcomes: Evidence from a Recent English Cohort. IZA Discussion Paper No. 6971, 30.

The Impact of Pre-school on Adolescents' Outcomes: Evidence from a Recent English Cohort

Patricia Apps
University of Sydney and IZA

Silvia Mendolia
University of Wollongong

Ian Walker
Lancaster University and IZA

Discussion Paper No. 6971
October 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6971
October 2012

ABSTRACT

The Impact of Pre-school on Adolescents' Outcomes: Evidence from a Recent English Cohort*

This paper investigates the relationship between attendance at nursery school and children's outcomes in adolescence. In particular, we are interested in child cognitive development at ages 11, 14 and 16, intentions towards tertiary education, economic activity in early adulthood, and in a group of non-cognitive outcomes, such as risky health behaviours (smoking, early pregnancy, use of cannabis) and personality traits (feelings and commitments about school; psychological well-being). Using matching methods to control for a very rich set of child's and family's characteristics, we find that pre-school childcare largely improves results in cognitive tests at age 11 and 14 and 16, and has a positive effect on intentions towards further education and economic activity at age 19-20. Positive effects are especially noticeable for children coming from disadvantaged socio-economic backgrounds. Results on non-cognitive outcomes are more mixed: we do not find any evidence of improvement in psychological well-being, but we do find some positive effects on health behaviours.

JEL Classification: J13, I21

Keywords: childcare, child outcomes

Corresponding author:

Ian Walker
Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
United Kingdom
E-mail: ian.walker@lancaster.ac.uk

* We are grateful to the Economic and Social Data Service at the University of Essex and the Department for Education for providing the data. The data can be obtained via http://www.esds.ac.uk/longitudinal/access/lsype/faq.asp and our STATA code is available on request.

1. Introduction

The objective of this paper is to analyse the role of pre-school education on child outcomes in later life. This research question has received increasing attention among economists, especially in the United States, and is motivated by the focus on childcare policy in many Western countries. Governments have devoted a significant amount of resources towards developing early-childhood policies, with the objective of improving the well-being of children, either through the direct effects of early education on children, or through the impact of childcare on maternal employment and income, or both. This paper sheds some light on the effect of formal childcare, provided prior to compulsory education, on various outcomes, both cognitive and non-cognitive. We add to the existing literature, by providing an analysis of a recent and very rich data-set of English adolescents, and by providing evidence of long-lasting effects in various areas of the child's life.

The effect of parental time and home inputs on child development has been widely analysed by psychologists and sociologists (see for example McCartney, 1984 and Lamb, 1996). A substantial body of literature has focused on maternal time vs. alternative care time and on the effect of household goods and income on children's outcomes. Economists are also interested in these topics, especially because some recent literature has showed that long term labour market outcomes, such as wages and employment, that determine lifecycle incomes, largely depend on factors and skills that are already in place by adolescence (see, for example, Cunha *et al*., 2006 and Keane and Wolpin, 2001 and 2006). Extensive research has showed that early cognitive achievements are strong predictors of later educational and labour market outcomes. For example, Bernal and Keane (2008) show that test scores at ages 4 and 6 are strongly correlated with the completed education for children of US single mothers. Indeed, Heckman and Masterov (2007) make a compelling case for government subsidies for intervention at an early age, for example with high quality childcare, because of the social benefits from lower crime etc. Thus, it is particularly important to analyse the role that childcare may have in later life.

Most of the economics literature on the effect of childcare is based on UK and US cohort data, much of it quite dated. It is hard to justify extrapolating from estimates of the effect of pre-compulsory education based on data from the early Sixties or Seventies, especially given the changes over the recent decades in early educational practices and policies. Some recent UK research has relied on the Early Provision of Pre-School Education (EPPE) study of three thousand children who attended childcare institutions in the late

1990's. This research has been important in underpinning the extension of free part-time childcare in the UK to age 4, and now to age 3, and to support the implementation of the more general SureStart[1] policy. To date results from the official SureStart evaluation are only available for outcomes up to age 7 (NESS, 2010). EPPE research is available up to the age of 14 (see Sylva *et al* 2012). Our research, is for a cohort only a few years earlier than the EPPE children, but we adopt a matching methodology and we extend their work for outcomes at 14 to include outcomes at 16 and beyond.

Our analysis here is based on the Longitudinal Study of Young People in England, a very rich study of a cohort of English children, mostly born in 1990, selected through their schools and interviewed for the first time in 2004. A great deal of information is collected about the child and her/his family and seven waves of data are now available. Important cognitive outcomes are merged into the data from national administrative records. A major problem with much of the existing literature is that it measures the effect on short term outcomes and there is a fear that such effects might "fade". The contribution of the paper derives from its ability to investigate long term outcomes. In particular, we investigate the effect of childcare on adolescent outcomes that are effectively permanent (like educational achievement) rather than on test scores per se. Moreover, such outcomes are known to have important effects on lifecycle income.

In this paper, we consider the effect of attending nursery school (before the child entered primary school) on a variety of outcomes in adolescence and early adulthood, controlling for a very rich set of child's and family's characteristics. We begin our analysis by looking at cognitive achievements and test scores at age 11, 14 and 16 and we exploit the richness of our data-set, analyzing the test results in various subjects. Then, we examine children's intentions to apply for university at age 17 and children's economic activity at age 19-20. We look into school outcomes more carefully, by exploiting information available on children's attitudes and efforts in school work, and general happiness of the child at school. We also look at some health behaviours between 14 and 20, such as smoking, use of cannabis, teenage pregnancy and psychological well-being.

[1] SureStart was an area-based initiative, implemented by the British Government since 1998, primarily in England with slightly different versions in Wales, Scotland and Northern Ireland. The initiative had the aim of "giving children the best possible start in life" through improvement of childcare, early education, health and family support, with an emphasis on outreach and community development. It has increasingly focused on childcare.

Ordinary Least Squares, to control for observable confounders, is well known to lead to biased estimates of the causal effects because of neglected heterogeneity. It is not clear what direction this bias might be: childcare users might be better or worse parents in unobservable ways. A partial solution to this problem is to attempt to better match users and non-users in observational data. Thus, we adopt Propensity Score Matching to estimate the effects of attending nursery school on children's outcomes and we compare the magnitude of these effects with those of other important variables, such as maternal education, marital status at birth, month of birth of the child, etc. Propensity Score Matching has been used in various recent papers investigating the determinants of children well-being in recent economics literature (see for example Ruhm, 2008; Berger *et al.*, 2005; and Goodman and Sianesi, 2005).

Following the most recent literature on the effect of childcare, we analyse the different impact of early education on children from various socio-economic backgrounds. In particular, we look at disadvantaged families, where disadvantage is defined in a variety of ways, and we analyse the effect of attending nursery school for them, compared to their advantaged peers. The results broadly support the idea that childcare prior to compulsory education is particularly beneficial for children who come from disadvantaged socio-economic backgrounds while the effect on advantaged children is less clear over the various specifications of our model.

Our paper contributes to the existing literature in two principal ways. First, we produce new evidence, based on a large and recent dataset, consisting of a cohort of children born in 1990 and followed for seven years, starting in 2004. Second, we take into consideration a variety of outcomes, including cognitive and non-cognitive development, and we are able to follow the children in our sample until the age of 21.

The rest of this paper is organized as follows. Section 2 provides an overview of the existing literature, Section 3 analyses the data and briefly presents well-being indicators. Section 4 discusses the estimation methods and Section 5 presents the main results. Section 6 concludes.

2. Overview of existing literature

Melhuish (2003) provides an excellent survey of a wide variety of literature for the UK and elsewhere. He focusses, in particular, on experimental and quasi-experimental work. He finds that the existing evidence on childcare (aged 0-2) is equivocal, while the evidence

on pre-school nursery education (age 3-4) points to a beneficial effect across the population with notable effects of both months of use and the quality of provision and a larger than average effect for disadvantaged groups.

These conclusions have largely been substantiated in more recent literature. A substantial amount of this research has looked at the effects of early childhood education on children's outcome in the United States: this literature is particularly focused on the evaluation of specific programs, targeting children from disadvantaged socio-economic backgrounds, such as single-parent families, ethnic minorities, low income groups etc. These programs include small scale high quality models (such as the Perry Pre-School Program) as well as more general large scale programs, such as Headstart. It has generally been found that intensive, high-quality targeted interventions are generally beneficial in producing short and long term positive outcomes for disadvantaged children, in terms of educational and labour market outcomes (see for example Carneiro and Heckman, 2003; Blau and Currie, 2006 and Currie, 2001). A few papers also examine the effect of typical preschool or kindergarten programs on school readiness and behavioural problems using specific data sets, such as the Early Childhood Longitudinal Study–Kindergarten Class of 1998–1999 and generally showing positive effects on short run outcomes, such as reading and maths skills, especially for disadvantaged children (see Loeb et al., 2007 and Magnuson et al., 2007). Table 1 summarises the main findings in the relevant literature on the effect of child care on children outcomes.

Recent literature (see Bernal and Keane, 2010 and 2011) has used the National Longitudinal Survey of Youth (NLSY) and exploited welfare reforms around the mid 90's to provide instruments for childcare use to analyse the effect of childcare on cognitive development and to show that formal centre-based early education has positive effects on children of US single mothers.

The literature on the effect of early education on children more generally has produced more controversial results, especially because of the difficulty in estimating these effects using non-experimental data where only a few regressors are available to attempt to control for selection issues. Bernal and Keane (2010) provide a thoughtful discussion of the most problematic issues in the estimation of childcare effects on children's outcome and present the different approaches that have been taken by previous literature using American data, including family fixed-effects (see for example Blau, 1999) and instrumental variable estimation (see Bernal and Keane, 2011). Recently, Berlinski *et al.* (2009) have exploited the

variation introduced by the expansion of universal pre-school education over time that generated differences in exposure by cohort and municipality in Argentina, showing that pre-primary attendance improved school test scores and also had a positive effect on behaviours such as attention, effort, class participation etc.

A small number of studies have used rich data-sets, specifically collected on pre-school-age children such as the Effective Provision of Pre-School Education (EPPE) data in the UK and the National Institute of Child Health and Human Development (NICHD) in the US. Most of the rest of the literature has relied on cohort studies such as the large British cohort studies such as the 1958 National Child Development Study (NCDS) and the 1970 British Cohort Study (BCS) for the UK, and the US National Longitudinal Survey of Youth.

The first strand of literature is based on some ad hoc studies, where pre-school children were recruited and followed for a number of years. In the US, the NICHD Study of Early Child Care was initiated in 1991 and included 1,300 children followed up to their seventh year of school. Children in higher quality childcare centres were found to have better results on test scores than children in lower quality childcare arrangements but children who spent more time in childcare were also found to have more behavioural problems (see Waldfogel, 1999). More recent research by Vandell *et al* (2010) looks at outcomes at age 15 from the NICHD study and find significant positive effects on educational attainment of high quality care, but not childcare *per se* - although these outcomes were not national tests where one might be able to claim that there would then be an established causal effect on lifecycle income. The British EPPE study was launched in 1997 and included a sample of 3,000 children from various socio-economic backgrounds, who attended a range of different pre-schools. The study also included around 300 'home' children with no pre-school experience at all. Children were followed until age 7 and pre-school was found to have a positive impact on cognitive and social development with a particularly positive effect on children from disadvantaged backgrounds and a stronger effect for nursery schools with a strong educational focus (see Sylva et al., 2004). At age 11 EPPE children who had been in high quality childcare were found to have higher literacy and numeracy levels (around 0.2 and 0.4 of a standard deviation respectively). The results did not seem to vary by parental social background. The EPPE team are continuing to track the children but the effect on long term outcomes is not yet available. Nonetheless, one influential outcome of the EPPE research has been to show that the estimated effects of high quality childcare centres are much larger than more informal forms of childcare and this motivated the UK Department of Education to

concentrate Surestart childcare resources into such centres. Surestart is loosely targeted on disadvantaged areas although all families within an area can access the local centre. Thus, Surestart, which was originally quite heterogeneous because of the local nature of the investments, has become, on average, a high quality childcare treatment for, on average, relatively deprived children. The evaluation work reported in NESS (2010) suggests that Surestart children had lower BMIs and better general health at age 5, but more recent outcomes are not yet available.

The second noticeable body of literature used British Cohort studies to analyse the effect of childcare on various children outcomes. The 1970 BCS was analysed by Osborn and Milibank (1987) who performed an analysis of variance and controlled for a wide set of characteristics, such as social class, family size, neighbourhood, gender, mother's age, mental state and employment, type of family, ethnic origin and the presence of handicaps and found that children with pre-school education have better results in cognitive tests at age 5 and 10 than their peers who did not go to pre-school, with a slightly greater advantage for children from disadvantaged backgrounds.

An analysis of the 1970 BCS and the 1958 NCDS was conducted by Feinstein *et al*. (1998), by constructing a model of child care choice and using the price of child care as an instrument for the amount of hours of childcare. The authors showed that pre-school has a positive effect on cognitive abilities up to the age of 11, using the 1958 cohort. On the other hand, this analysis of the 1970 cohort suggested that pre-school has a negative effect on vocabulary when the children were 5, and reading skills when the children were 11. A recent paper by Goodman and Sianesi (2005) analyses the 1958 NCDS looking at the effect of any early education (specifically, early entry into primary school, as well as attendance of nursery schools and playgroups) on a wide range of outcomes, including cognitive achievements at age 7 through to 16, socialisation and outcomes such as wage and employment at age 33. This paper aims at estimating the total policy effect of early education, using Ordinary Least Squares, Fully Interacted Model and Propensity Score Matching. The authors show a positive effect of pre-school education on test scores, diminishing in size as the children grow, while the effect on socialisation was more mixed. In adulthood, pre-compulsory education was found to increase the probabilities of obtaining qualifications and of being employed at age 33. However, most of the effects found in this paper were attributed to *any* pre-compulsory education, including early school entry, nursery and playgroup (or any combination of these). The effects of pre-school were more mixed and not clearly identifiable.

Table 1 Summary of main findings in the literature on the effects of childcare on outcomes (in alphabetical order)

Paper	Data	Method	Main findings
Berlinski (2009)	Argentina data on universal childcare	IV	One year of pre-primary school increases average third grade test scores by 23% of a sd of the distribution of test scores.
Bernal and Keane (2010)	US, NSLY, 1979	ML	One year of full-time maternal work and childcare use reduces test scores (age 3-6) by 15% of a sd
Bernal and Keane (2011)	US, NSLY,1979	IV	A year of informal childcare reduces child test score (age 4, 5 and 6) by 11% of a sd
Dumas and Lefranc (2010)	French panel DEPP and FQP	IV	Delayed pre-school enrollment leads to test score (age 16) that are 10% of a sd lower
ECCRN and Duncan (2002)	US, NICHD, 1991	OLS	Two standard deviation improvement in child care quality in early childhood is associated with a one-sixth to one-seventh of a sd increase in cognitive functioning at age 24
Fenstein *et al.* 1998	UK BCS 1970 and NCDS 1958	IV	For NCDS cohort, no effects on social adjustment and positive effects on cognitive tests up to age 11. For the BCS70 cohort, evidence of marginally worse social adjustment and reduced vocabulary at 5, worse reading skills at 11 and no effects on maths skills
Goodman and Sianesi (2005)	UK, NCDS 1958	PSM	Pre-school education increased test scores at 7 of 9% of a sd; test scores at 11 of 7% of a sd; test scores at 16 of 5% of a sd. Positive effect on socialisation only at 7 (5% of a sd) Negative effect on self-control at 7; no effect on interpersonal skills at 7; Positive but weak effect on higher education, employment and wages at 33
Loeb et al. (2007)	US, ECLS 1998	OLS, PSM and IV	Center-based care raises reading and math scores (11212% of a sd), but has a negative effect for socio-behavioral measures. However, for English-proficient Hispanic children, the academic gains are considerably higher and the socio-behavioral effects are neutral.
Magnuson et al. (2007)	US, ECLS 1998	PSM and IV	Prekindergarten is associated with higher reading and mathematics skills (around 38% of a sd) at school entry, but also higher levels of behavior problems.
NESS (2010)	UK SureStart and MCS	PSM	SSLP reduced BMI of approx.. 12% of a sd and improves physical health of appr. 10% of a sd.
Osborn and Millibank (1987)	1970 BCS	OLS	Positive effect on cognitive development at 5-10, especially in vocabulary expansion. An average deviation from mean attainment equivalent to one-third of a sd was predicted if the child was in some form of pre-school care.
Vendell *et al* (2010)	US, NICHD, 2010	OLS	Significant effects of child care quality (about 20% of a sd) but not centre care per se, on age 15 academic attainment in tests.
Sylva *et al*. 2004	EPPE	OLS	Children who had been in high quality childcare were found to have higher literacy and numeracy levels (around 20% and 40% of a sd)

The major limitation of all the studies using the 1958 NCDS and the 1970 BCS is that they rely on relatively old data, and have only a relatively small percentage of children attending nursery schools or other formal childcare centres. The early childhood industry has rapidly evolved over the last three decades and many changes have taken place, in terms of pedagogy, teachers' qualifications, and focus on cognitive development. These changes make it difficult to rely on conclusions on the effects of child care based on these data, so the policy implications of these studies are limited. Therefore, our analysis fills a gap between papers based on 1958 NCDS and 1970 BCS, such as Goodman and Sianesi (2005), that consider long lasting outcomes based on old data, and more recent studies, such as the\ EPPE study, that use recent data but only look at the impact of pre-compulsory education on primary school results. Our outcomes span a reasonably long period of time (15 years after the end of pre-school education) while still looking at relatively recent child-care provision (1993-94 births). Our results are consistent with those presented by Goodman and Sianesi (2005), but are focused on pre-school childcare only, rather than both pre-school and early entry into primary school.

3. Data

This paper uses data from the Longitudinal Study of Young People in England (LSYPE), which is a large scale panel survey of English adolescents, interviewed for the first time when they were in school year 9 in 2004 at the age of 14[2]. The study is managed by the Department of Education and covers wide range of social policy issues. The questionnaires cover a variety of topics, including academic achievements, family relationships, attitudes toward school, family and labour market, and some sensitive or challenging issues, such as risky health behaviours (smoking, alcohol drinking, drug taking), personal relationships, etc.

In the first wave, selected to be representative of the young people in England, around 15,500 young people were interviewed. In the first four waves, parents/guardians were also interviewed. In addition, LSYPE can be linked to the National Pupil Database (NPD), a pupil level administrative database which matches pupil and school characteristics data to pupil level attainment and contains detailed information on test scores (Key Stage 2, Key Stage 3 and Key Stage 4 or GCSE) for all pupils in England and Wales. It also contains limited data about the pupil - such as free school meal eligibility and Special Education Needs status. Retrospective information about the LSYPE child and the family was asked of the parent at

[2] Note that grade repetition is very rare in the British school system.

waves 2 and 3 and we use this to determine whether the child attended nursery school, and to provide information on the child's and family's situation at birth. Our final sample includes around 11,000 observations of children with non-missing information on test scores, early education and other essential information on the child's birth and family background.

Outcomes

We are interested in analysing the impact of pre-school education on a variety of outcomes. The LSYPE includes information on various outcomes, measured from adolescence to early adulthood. This allows us to attain a very complete picture of the skills and behaviours that are affected by early education. Table 2 lists the outcomes we use in our analysis and the age of the child for each outcome. We focus on five groups of outcomes:

Cognitive development and test scores. We analyse data from the NPD on children's results in tests following the National Curriculum, and particularly Key Stage 2 (age 11), Key Stage 3 (age 14) and Key Stage 4 (GCSE at age 16). Key Stage 2 consists of national curriculum tests in English (reading, and writing, including handwriting and spelling) and mathematics, together with teacher assessments in English, mathematics and science. Key Stage 3 consists of teacher assessments only, in all national curriculum subjects. Schools have to submit results for English, maths and science. At the end of Key Stage 4, pupils generally take the national public examinations known as GCSE in most subjects studied – often in as many as 10 subjects. GCSE grades range from A* to G. The dependent variables in our analysis are: Key Stage 2 and Key Stage 3 scores (comprehensive and in single subjects); the number of subjects with grade from A* to C in GCSE exams; and two binary variables equal to 1 if the child attained A* to C in English and Mathematics.

Attitudinal variables and socialisation. Recent literature has showed that test scores do not necessarily capture individual behaviours very well. Rather, personality traits may have an important role in determining the future success of individuals (see Heckman, 2012). Therefore, we extended our analysis and take into consideration some additional outcomes such as: the intensity of effort at school; whether the child likes her/his school and teachers; the number of close friends the child has; and the

psychological well-being of the child (measured through the General Health Questionnaire score[3]).

Economic activity and education at age 20-21.

Health risky behaviours such as smoking, use of cannabis, and early pregnancy.

Problematic behaviours such as being involved in fighting, being suspended from school, vandalism, shoplifting, being in contact with the police.

Pre-school

We are interested in investigating the effect of pre-school on various children outcomes, including cognitive development. In order to address these research questions, we use the information recorded in the LSYPE history data file where the parents are asked whether the child went to nursery school. Unfortunately, we don't have any information on how many days/hours the child spent in nursery school. Around 80% of the children in the estimation sample went to nursery school[4]. This is a very high proportion and it is likely to include children who received part-time, as well as full-time pre-school. The percentage of children attending nursery school in LSYPE is much higher than the proportion of children receiving pre-compulsory education in the NCDS (83% vs. 15%) and, as a consequence, our sample of children attending nursery school is much bigger than the one utilised by Goodman and Sianesi (2005).

There is no a clear definition of nursery school in LSYPE data, but we believe that the self-reported "nursery school" is likely to include various forms of care in centre-based institutions, such as day care centres, local authority and independent nurseries, and even play-groups. Some will provide a formal curriculum, most will focus on play. Some will provide full time care, most will be for just for three hours most weekdays, and most will be run by professionally trained staff with little reliance on parental help.

[3] The GHQ Caseness score is constructed from the responses to 12 questions covering feelings of strain, depression, inability to cope, anxiety-based insomnia and lack of confidence. The twelve answers are combined into a total GHQ score that indicates the level of mental distress, giving a scale running from 0 (the least distressed) to 12 (the most distressed)

[4] We checked this proportion with data from the Family Resource Survey. The FRS 9394 only includes 750 observations for children aged 3-5 with valid childcare answers. In this sample, around 50% of the children attend some form of structured child care and the rest is looked after by a relative or friend. We believe that this percentage is smaller than what we find in LSYPE because it does not include children who are looked after by a family member (stay at home mother, grandparent, etc.) *and* attend a pre-school (or nursery school) for a couple of hours per day (as the parents are likely to record the first source of child care).

Table 2 Outcomes

Cognitive development	***Test scores***
Overall cognitive development at 11	KS2 Score (points)
Overall cognitive development at 14	KS3 Score (points)
Overall cognitive development at 16	GCSE – N subjects pupil achieved A*-C
Language skills at 11	KS2 Score in English (points)
Maths skills at 11	KS2 Score in Maths (points)
Science skills at 11	KS2 Score in Science (points)
Language skills at 14	KS3 Score in English (points)
Maths skills at 14	KS3 Score in Maths (points)
Science skills at 14	KS3 Score in Science (points)
Language skills at 16	=1 if pupil attained lever A*-C in GCSE English
skills at 16	=1 if pupil attained lever A*-C in GCSE Maths
Attitudinal variables and socialisation	
Effort in school work	=1 if child agrees with the following statement: at school I work as hard as I can (at wave 1)
Like school	=1 if child agrees with the following statement: On the whole, I like being at school (at wave 1)
Like teachers	=1 if the child declares she/he likes all or most of her/his teachers (at wave 1)
N. of friends	N. of close friends the child has (only recorded at wave 6 or 7)
Psychological well-being	General Health Questionnaire (from 0=not distressed to 12 = most distressed) at wave 1
Economic activity and Education	
Intentions towards tertiary education	= 1 if the child has applied to go to university in wave 4 (age 17)
In University/Education at 20-21	=1 if child is attending university (or other educational course) at wave 6 or 7 (age 19)
Working at 20-21	=1 if the child is working at wave 6 or 7
NEET (Not in Employment, Education or Training)	=1 if child not working, attending university, or other training course at wave 6 or 7
Health risky behaviours	
Smoking	=1 if child ever smokes cigarettes wave 1)
Smoking cannabis	=1 if child ever tried cannabis (wave 1)
Pregnant	=1 if child has ever been pregnant (recorded wave 6 or 7)
Problematic behaviours	
Suspended from school	=1 if child has ever been suspended from school (at wave 1)
Vandalism, Shoplifting, Graffiti	=1 if child has ever shop lifted, vandalised a public property or graffitied (at wave 1)

Children in LSYPE were born in 1989-1990, so they are likely to have attended nursery schools between 1992 and 1995. Before 1997, there were no by-law requirements for Local Authority in terms of educational provision for children under compulsory school age, and the decision of whether to provide free nursery places (and if so how many to provide) was left to each individual Local Authority, leading to substantial variation in provision. According to Brewer *et al*. (2005) provision across the country ranged from zero free pre-compulsory education places provided, to a high of 27.5 places per 100 children and this variation in access to nursery education persisted into the 1990s (see Dickson, 2008 for a discussion of changes to the provision of nursery places in the late 1990s).

Other explanatory variables

This analysis exploits the extensive information available in LSYPE. We estimate three versions of our model, progressively increasing the set of independent variables. As we will discuss in greater detail in section 4, we try to capture all factors that determine early education attendance and child outcomes. All of the variables we control for are, arguably, pre-determined variables – that is, not themselves influenced by pre-school education. Inputs in children's outcomes include individual mental and physical endowments, parental and family inputs (such as income, time, size of the family and number of siblings), and local area characteristics.

Our first, most parsimonious, model only includes at-birth characteristics such as: birth-weight; whether the child was premature; ethnic background; sex of the child; month of birth; and family characteristics such as marital status and age of the mother at birth. In the second model we include other family's characteristics (measured at wave 1, which are unlikely to have changed since the child's birth) such as: main language of the family, maternal education; child's and mother's disability; grandparents' education and older siblings. In the last model we include some characteristics at wave 1 such as: younger siblings; maternal working and marital status; household income; family size; local authority binary variables and whether the child has ever been in care. Table 3 lists the explanatory variables used in the empirical model.

We explore some potential heterogeneity in the effect of nursery school on children with different socio-economic backgrounds and we construct various indicators of socio-economic disadvantage. First, we follow Ruhm (2008) and construct a multivariate indicator of socioeconomic status by regressing total family income on mother's age at birth, education

marital status. Youths are then ordered by predicted status and classified as "advantaged" ("disadvantaged") if in the upper (lower) half of the distribution. This SES index simultaneously accounts for a larger number of determinants than simple income and possibly reduces the endogeneity problem. Secondly, we also divide youths by maternal marital status at birth, maternal employment when the child was 5, maternal ethnic background, and education.

Table 3 Control variables

Model 1 Observable characteristics at birth	
Child	
Birth-weight	In kg
Month of birth	Omitted: September
Premature birth	=1 if the child at least 3 weeks early
Sex of the child	Boy=1
Ethnic background:	White (omitted), black, Asian, mixed
Mother	
Young mother	=1 if mother was <21 ys old at child's birth
Single mother	=1 if mother was not married at child's birth
Model 2 Observable characteristics at birth (as in Model 1)	
Child	
Child's disability	=1 if the Child has a disability or long standing illness
Mother	
Maternal education:	Degree (omitted); Higher education (not degree); junior high school graduate (GCSE a*-c); No qualification
Family	
English	=1 if English main language spoken in hh
Older siblings	Number of older siblings
Grandparents' education	=1 if main parent's parent went to university
Main parent's disability	=1 if main parent disability or long standing illness
Model 3 Observable characteristics at birth (as Model 1). Other observable stable characteristics (as Model 2). Other characteristics at wave 1	
Child	
Care	=1 if the child has ever been in care
Mother	
Working status:	Employed (omitted); unemployed; out of the labour force
Single mother	=1 if the child's mother is single
Family	
Younger siblings	Number of younger siblings
Household yearly income.	3 groups: <£11,400; £11,400 to £31,200; >£31,200 (omitted)
Local Authority	Local Educational Authority dummy
Family size	Number of people in the household

Descriptive statistics

Table 4 presents the distribution of pre-school and all the independent variables included in our model, by socio-economic status. Not surprisingly, disadvantaged children are less likely to have attended nursery school. Their mothers are more likely to be poorly educated, out of the labour market, and single. Table 5 presents the distribution of outcomes, split by nursery school attendance. On average, children who went to pre-school perform better than those who didn't receive early education in all test-scores and they are more likely to be happy with their school and teachers and put lots of effort into school work. They also have an average higher number of friends, more likely to have applied for university at 17 and less likely to have ever tried smoking. However, no significance difference is found in terms of university attendance. Children who went to nursery seem less likely to be out of education or employment when they reach the age of 20-21 and less likely to be engaging in problematic behaviours.

Table 4 – Distribution of the independent variables

	All	**Disadvantaged**	**Advantaged**
Attendance of nursery school (%)	85%	83%	87%
Birth-weight (average in kg)	3.31	3.27	3.4
Premature birth (%)	10%	11%	10%
Boy (%)	51%	51%	51%
Single parent household at birth (%)	23%	37%	1%
Mother younger than 20 at birth (%)	6.5%	10%	1%
White (%)	76%	72%	83%
Black (%)	6%	6%	5%
Asian (%)	11%	14%	6%
Mixed ethnic background (%)	7%	8%	6%
Mother has a university degree (%)	10.5%	2%	23%
Mother has other tertiary qual (%)	13%	4%	26%
Mother has gcse a level	13%	7%	22%
Mother has gcse a-c	30%	35%	22%
Mother has qual level 1 or below (%)	9%	16%	0%
Mother has other qualification	2%	3%	0%
Mother has no educational qual (%)	19%	31%	0%
English main language in family (%)	95%	94%	98%
Number of younger siblings (average)	0.9	1	0.7
Number of older siblings (average)	0.9	0.93	0.9
Child has a disability (%)	14%	14%	13%
Main parent has a disability (%)	21%	23%	17%
Grandparents went to university (%)	9%	6%	12%
Mother is employed at wave 1 (%)	71%	63%	84%
Mother is unemployed at wave 1 (%)	1%	1.5%	0.8%
Mother is out of the labour force (%)	27%	35%	15%
Family income < 11,000 £ pa (%)	24%	31%	14%
Family income 11,000-31,200 pa (%)	44%	50%	37%
Family income >£31,200 pa (%)	31%	19%	49%
Child ever been in care (%)	0.4%	0.6%	0.1%
Family size (average)	4	4.4	4
Single mother (%)	22%	29%	12%

Note: all variables not related to characteristics at birth are measured at wave 1.

Table 5 – Distribution of outcome variables

	Whole sample	**Children attended nursery**	**Children not attend nursery**
KS2 Score (points) - Average	27.2 (3.9)	27.3 (3.8)	26.6 (4.1)
KS3 Score (points) - Average	34.4 (6.5)	34.6 (6.4)	33.2 (6.8)
# GCSE subject grade A*-C	6.1 (4.2)	6.2 (4.2)	5.5 (4.3)
KS2 Score in English (points)	26.7 (4.2)	26.8 (4.2)	26.0 (4.5)
KS2 Score in Maths (points)	26.7 (4.8)	26.8 (4.7)	26.2 (4.9)
KS2 Score in Science (points)	28.4 (3.6)	28.5 (3.6)	27.9 (3.7)
KS3 Score in English (points)	33.6 (6.0)	33.82 (6.0)	32.7 (6.2)
KS3 Score in Maths (points)	36.1 (7.8)	36.36 (7.7)	35.0 (8.3)
KS3 Score in Science (points)	33.5 (6.5)	33.73 (6.5)	32.5 (6.9)
GCSE A*-C in English(%)	62.3	63.6	54.9
GCSE A*-C in Maths (%)	57.2	58.2	51.6
Effort in school work (%)	81.5	81.8	80.2
Like school (%)	84.8	85.2	82.8
Like teachers (%)	42.6	43.1	39.8
N. of close friends (Average)	3.4 (8.6)	3.5 (8.1)	2.7 (11.7)
Psychological wellbeing GHQ	1.67 (2.51)	1.68 (2.51)	1.63 (2.55)
Applied to university age 17 (%)	63.1	63.8	59
University/Educ age 20-21 (%)	56.8	51.9	51.3
Working at age 20-21	39.2	36	34.2
NEET at age 20-21	8.8	8.4	11
Ever Smoking (%)	9.6	9.2	11.8
Ever tried cannabis (%)	8.6	8.5	9.4
Ever been pregnant (%)	4.5	4.4	5.3
Suspended (%)	9.4	9.1	11.0
Problematic behaviour	22.2	22.0	23.6

Note: Test scores are recorded at the appropriate age. All the other variables are measured at wave 1, unless differently specified. Problem behaviour examples are vandalism, shoplifting, graffiti, and having been in touch with the police. NEET= not in employment, education or training. Standard deviations in brackets.

4. Estimation

We begin our analysis by estimating a linear regression to examine the effect of pre-school education on children outcomes. The linear model can be written as:

$$C_i = \alpha + \beta_i N_i + \gamma_i X_i + \varepsilon_i,$$

where C_{it} represents a particular outcome, N_i is a binary variable equal to 1 if the child attended nursery school and X_i is a vector of child's and family's characteristics. We use linear probability models rather than nonlinear probit or similar.

The major challenge in this analysis is establishing causal connections between pre-school education and child outcomes, given that children did not receive early education through random assignment and the children who went to nursery school may have unobserved characteristics which also affect their cognitive development or other outcomes. Furthermore,

mothers who worked and decided to send their child to nursery school may be systematically different from those who did not, and their child's cognitive ability can itself influence mothers' decisions. However, a significant percentage of children in our sample went to nursery school (around 85%) so this group is likely to include children with working mothers (both part-time and full-time), as well some children with stay-at-home mothers. For these reasons, we think of OLS as providing an upper bound to the causal effects.

The effect of nursery school on children's outcomes may be estimated correctly through OLS if several assumptions are true. First, the "selection on observables" assumption must be satisfied (see Heckman, 1979). This means that all variables that predict both pre-school attendance and children's outcomes should be included in our model. These variables are sometimes called "confounding variables" because if they are not appropriately controlled for, their effect on the outcomes is confounded with the effect of the causing variable of interest (see Angrist and Krueger, 1999). Second, the model must be correctly specified: this is very problematic, as assumptions like linearity and additivity are really hard to verify when we include several independent variables. Further, if there is a lack of overlap in covariate distributions across children who went and did not go to nursery school – that is, if there are children who went to pre-school for whom there are no comparable children who did not go to nursery school- linear regression models extrapolate results over portions of the distribution where there is no support (comparing incomparable children).

We cannot, in this data, address the selection on unobservables problem. There is simply no quasi-experimental variation across our sample to exploit. However, we can go some way towards addressing the other problems. Firstly, we try to lower the upper bound provided by OLS estimation, through the inclusion of a progressively more detailed set of independent variables. Second, we exploit propensity score matching, that does not strongly rely on functional form assumptions and restricts inference to samples where we can find overlap in the distribution of covariates across the treatment (i.e. going to nursery school).

In order to estimate the effect of pre-school education on children's outcomes, we would ideally need to compare the average outcomes for children who went to nursery with the average outcomes for the *same* children had they not received pre-school education. However, the latter is an unobserved counterfactual and the evaluation problem is to provide unbiased estimates of this average counterfactual using appropriate methods and assumptions. The idea of propensity score matching is to find a group of children who did not go to nursery school that looks as similar as possible to our treatment group (children who went to nursery school). More specifically, firstly we estimate the conditional probability of

going to nursery school (being in the treatment group) for each child, given our covariates. This is called the propensity score. Then, estimated propensity scores are used to create a matched control group and for each treated child we find the comparison member with the closest propensity score. Non matched individuals are dropped from the analysis. Our analysis is performed using *psmatch2*[5] and appropriate tests have been run, in order to compare covariate distribution across our matched groups to ensure that adequate has been obtained. Results from the balancing tests are shown in the Appendix.

Matching is more robust than OLS because it does not restrict the way in which nursery school may affect child outcomes to be linear, and inference is limited to samples that are effectively comparable, based on the covariates distribution. Matching attaches appropriate weights to the observations in the control group, so that the distribution of their observable characteristics is realigned to the treatment group.

5. Results

Results from the estimation of the effect of pre-school education on child outcomes are presented in Tables 6 to 9. Appendix Table A2 presents the results in terms of percentage of standard deviations.

Table 6 presents results on the effect of pre-school education on cognitive development and test scores for the whole sample. Results are then split by sex of the child and by various indicators of socio-economic disadvantage in Tables 7 and 8. Pre-school education generally has a positive effect on test scores at age 11, 14 and 16 and the sizes of the effects are notable. The positive effects are found both for average test scores and separately for maths and language skills. In model 2 and model 3, nursery school attendance generally increases average Key Stage 2, Key Stage 3 and GCSE scores by about 6-7 per cent of a standard deviation. The size of these effects is comparable with some other important characteristics, such as birth-weight, sex of the child, number of older siblings or parental disability.

The estimation with propensity score matching yields similar results, with the exception of the effect on Key Stage 2 and Key Stage 3 scores, which become insignificantly different from zero. Interestingly, the magnitude of the effect is higher for language skills (between 8 and 10 per cent of a standard deviation) rather than mathematics or science skills

[5] Our approach is similar to Goodman and Sianesi (2005) and we use propensity score matching with the nearest neighbour method with replacement (as it has been shown to reduce bias relative to matching without replacement, see Dehejia and Wahba, 2002) and then used the option *common*, so that off-support observations are automatically dropped. Similar results were obtained with other methods.

Table 6 Effect of pre-school on cognitive development (Whole sample)

	Model 1		Model 2		Model 3
Age 11	**OLS**	**PSM**	**OLS**	**PSM**	**OLS**
KS2 Overall	0.582	0.600	0.274	0.278	N.A.
	(0.10)**	(0.15)**	(0.09)**	(0.14)+	
KS2 Maths	0.529	0.540	0.223	0.243	N.A.
	(0.12)**	(0.17)**	(0.12)+	(0.17)	
KS2 English	0.706	0.553	0.367	0.422	N.A.
	(0.11)**	(0.16)**	(0.10)**	(0.15)**	
KS2 Science	0.430	0.472	0.179	0.117	N.A.
	(0.09)**	(0.13)**	(0.08)*	(0.13)	
Age 14					
KS3 Overall	1.083	1.027	0.526	0.338	0.518
	(0.16)**	(0.243)**	(0.15)**	(0.24)	(0.17)**
KS3 Maths	1.065	0.732	0.427	0.387	0.459
	(0.20)**	(0.30)**	(0.14)**	(0.29)	(0.21)*
KS3 English	0.919	0.457	0.468	0.447	0.403
	(0.16)**	(0.22)*	(0.19)**	(0.22)*	(0.16)**
KS3 Science	0.929	0.988	0.400	0.211	0.420
	(0.17)**	(0.25)**	(0.16)**	(0.244)	(0.18)*
Age 16					
# GCSE at A*-C	0.627	0.678	0.298	0.346	0.299
	(0.11)**	(.150)**	(0.09)**	(0.15)**	(0.11)**
A*-C in English	0.083	0.093	0.047	0.052	0.044
	(0.01)**	(0.017)**	(0.011)**	(0.02)**	(0.01)**
A*-C in Maths	0.060	0.046	0.018	0.041	0.029
	(0.01)**	(.017)**	(0.01)**	(0.02)**	(0.01)*
Age 17-18					
Has applied to university	0.056	0.047	0.028	0.058	0.029
	(0.014)**	(0.019)**	(0.013)*	(0.019)**	(0.013)*
Age 19-20	0.0209	0.091	-0.010	0.102	NA
Is attending university	(0.011)+	(0.019)**	(0.0121)	(0.021)**	

Notes: Standard errors are in brackets + indicates that the underlying coefficient is significant at 10% level, * at 5% and **at 1%. Model 3 was not estimated for KS2 as KS2 is achieved at age 11 and Model 3 includes variables measured at wave 1 (age 14).

Table 7 Effect of pre-school on cognitive development (by sex of the child)

Model 2 PSM	**Girls**		**Boys**	
Age 11				
KS2 Overall	0.393	(0.195)*	0.254	(0.204)
Age 14				
KS3 Overall	0.750	(0.329)**	0.110	(0.343)
Age 16				
# GCSE at A*-C	0.415	(0.211)*	0.133	(0.206)

in Key Stage 2 and Key Stage 3 scores (around 4-5 per cent of a standard deviation). Note that Model 3 is not estimated for KS2 (age 11) because it includes variables at age 14. Nor are PSM estimates provided for Model 3 because it seems unwise to match according to data that was not observed at age 11.

Nursery attendance also increases the number of GCSE subjects with A*-C scores (around 8 per cent of a standard deviation) , as well as probability of getting A*-C in English (by around 5 percentage points) and Maths (by around 4 percentage points) in GCSE exams. Our results are consistent with Goodman and Sianesi (2005), who find that obtaining education before age 5 is associated with an increase of 7 per cent of a standard deviation in average test scores at age 11. Positive effect of nursery school can be mediated through a variety of factors, including early exposure to literacy and numeracy and socialisation. Also, the children may build up independence and self-confidence and get used to school routines and this is an advantage and enhance their learning when they are in formal education.

Interestingly, our results do not show any fading effect when the children grow up and nursery attendance seems to be beneficial in increasing the probability of applying to university at age 17-18 and being actually attending university at 19-20.

Results are then split by sex of the child in table 7. Nursery attendance seems to be particularly beneficial for girls, while the effects on boys' test scores are not significantly different from zero. In Table 8, we analyse the effect of nursery education by socio-economic status. Pre-school education is highly beneficial for children coming from disadvantaged socio-economic backgrounds. Key Stage 2 and Key Stage 3 average scores increase by around 10-11 per cent of a standard deviation for disadvantaged children who went to nursery school and by about 18% of a standard deviation for single mothers' children. Children with an unemployed mother also benefit highly from nursery attendance (GCSE results improve by around 14% of a standard deviation). These results are stable across subjects and different

estimation technique. The GCSE score increases by around 7 per cent of a standard deviation for disadvantaged children and the chance of getting A*-C in English or Maths increases by around 4-5 percentage points. The chances of applying and actually attending university are also positively affected by nursery attendance for children who come from a disadvantaged socio-economic background.

On the other hand, the effect on children from advantaged backgrounds is significantly different from zero only when we estimate model 1 and when we look at the probability of being in university at age 19-20. These results are consistent with other findings in the literature looking at the impact of child care on children's development. Pre-school education seems to be particularly positive for children experiencing a poor and disadvantaged environment at home, as they get significant benefits from the exposure to a positive learning environment. Also, they are more likely to get intellectual stimulation and early exposure to numeracy and literacy during their time at nursery. The size of these effects is considerable and does not vanish as these children get older.

We also find considerable evidence showing that pre-school education is particularly useful for children having siblings rather than only children (results are not presented for parsimony, but are available on request). This result might be driven by the fact that only children receive more attention within the family and do not derive as much advantage from pre-school.

Table 9 presents results from the estimation of the effect of pre-school education on non-cognitive outcomes. Results are estimated with OLS and PSM and independent variables from Model 2. The evidence on non-cognitive outcomes is more mixed than we found for test scores. Nursery attendance generally increases the number of friends and children's satisfaction and effort at school at age 14. It also decreases the risk of not being in education or employment at 20-21 by around 1 p.p. Children who went to nursery school also seem less likely to have engaged in health risky behaviours such smoking or cannabis use when they are 14. On the other hand, we do not find any significant effect on psychological well-being, problematic behaviours, or the risk of being suspended from school.

Appendix Table A1 presents results for the effect of other independent variables in Model 2 on Key Stage 2, Key Stage 3 and Number of GCSE subjects with grade A*-C. As expected, there is a strong education gradient so that children with more educated mothers (or with grandparents who went to university) are more likely to have higher test scores.

Table 8 Effect of pre-school on cognitive development (by socio-economic status)

Model 2 PSM	**Dis. youths**	**Adv. youths**	**Single mother at birth**	**Partnered mother at birth**	**Mother not employed when child 5**	**Mother employed when child 5**	**Non White Mother**	**White Mother**	**Low educated mother**	**High educated mother**
Age 11										
KS2 Overall	0.437	-0.110	0.735	0.271	0.419	0.217	0.216	0.392	0.051	-0.009
	(0.17)**	(.205)	(0.31)**	(0.27)	(0.19)*	(0.30)	(0.29)	(0.16)*	(0.24)	(0.29)
Age 14										
KS3 Overall	0.630	0.126	0.641	0.173	0.651	0.182	1.05	0.44	0.755	0.095
	(.28)**	(.374)	(0.452)	(0.15)	(0.31)*	(0.17)	(0.45)**	(0.28)	(0.37)	(0.50)
Age 16										
# GCSE at	0.295	0.166	0.428	0.111	0.605	0.307	0.201	0.345	0.253	0.283
A*-C	(0.18)+	(0.214)	(0.28)	(0.17)	(0.18)*	(0.20)	(0.29)	(0.16)*	(0.25)	(0.29)
Age 17										
Has applied	0.051	-.0110	0.097	0.022	0.0475	0.066	0.0014	0.0581	0.014	-0.004
to university	(0.02)+	(0.02)	(0.041)**	(0.021)	(0.029)	(0.027)*	(0.033)	(0.023)*	(0.034)	(0.031)
Age 19-20										
Is attending	0.047	0.078	-0.0161	-.0266	-0.041	0.058	-0.039	0.083	0.006	0.024
university	(0.026)+	(0.03)*	(0.064)	(0.036)	(0.049)	(0.045)	(0.046)	(0.038)	(0.032)	(0.046)

Notes: Standard errors are in brackets + indicates that the underlying coefficient is significant at 10% level, * at 5% and **at 1%.

Table 9 Effect of pre-school education on other outcomes- Whole sample

	Model 2			Model 2	
	OLS	**PSM**		**OLS**	**PSM**
Attitudinal variables and socialisation			**Health risky behaviours**		
Number of friends at wave 6 or 7	0.047	.089	Ever Smoked at wave 1	-0.024	-.0248
	(0.027)+	(0.05)+		(0.07)**	(0.011)*
Like her/his school at wave 1	0.028	.0307	Ever tried cannabis at wave 1	-.038	-.013
	(0.09)**	(0.013)**		(0.019)*	(0.010)
Like her/his teachers at wave 1	0.037	0.0487			
	(0.012)**	(0.016)**			
Puts lots of effort in school work at wave 1	0.022	0.0127			
	(0.01)*	(0.013)			

	Model 2			Model 2	
	OLS	**PSM**		**OLS**	**PSM**
Economic activity and education			**Problematic behaviours**		
Not in education or employment at wave 6 or 7	-0.011	-0.009	Ever suspended from school	-.009	-.0029
	(0.007)+	(0.013)		(0.007)	(0.01)
Is working at wave 6 or 7	0.010	-.0240	Shoplifting, vandalism, graffiti or contact with the police at wave 1	-.008	0.001
	(0.011)	(0.021)		(0.01)	(0.014)

Notes: Standard errors are in brackets + indicates that the underlying coefficient is significant at 10% level, * at 5% and **at 1%.

Children's and parents' disability decrease test scores, and the same effect is found for negative socio-economic conditions at birth, such as low birth-weight, single or very young mothers. Boys, ethnic minorities and younger children also generally show lower test scores.

6. Conclusion

In this paper, we have investigated the effect of pre-school education on various children's outcomes, including cognitive development and non-cognitive outcomes in adolescence and early adulthood. We have used the Longitudinal Study of Young People in England, which is a rich source of information on English teen-agers and can be linked to the National Pupil Database, in order to get detailed information on school outcomes. We find that pre-school education significantly increases test scores at age 11, 14 and 16 and is particularly beneficial for children coming from disadvantaged socio-economic backgrounds. The size of the effect is noticeable and is comparable to other important variables, such as birth-weight, sex of the child or parental disability.

The results on non-cognitive outcomes are more mixed. We do find some positive effects on socialisation, attitudes towards schooling and post-compulsory education, but we do not find any significant effect on mental well-being and problematic behaviours. On the other hand, we show that children who attended nursery school are less likely to have tried smoking or cannabis at age 14.

Our analysis is performed using Ordinary Least Squares and Propensity Score Matching. We make extensive use of the amount of information contained in LSYPE and gradually increase our set of independent variables, in order to control for all factors affecting both school outcomes and nursery attendance. Our results are stable over different specifications of our model. Propensity Score Matching allow us to find a group of children who did not go to nursery school that looks as similar as possible to our treatment group (children who went to nursery school), given our independent variables. We can then compare outcomes between these two groups and check whether there are any significant differences.

This analysis has some important policy implications and fills the gap in the existing literature, by providing new evidence on recent British data. Governments across various Western countries have been devoting a significant amount of resources to the improvement of life conditions in childhood and there has been considerable debate about the need of government intervention in increasing the provision of free early education places. This paper provides further evidence of the positive effects of pre-school education on children's

outcomes, especially for disadvantaged children. Further studies are needed to assess the stability of our results in the presence of selection on unobservables and to establish the transmission channels of the effects.

Appendix

Table A1 Effect of other independent variables on cognitive development

	KS 2 Overall Score	KS3 Overall Score	# GCSEs at A*-C
Premature birth	-0.001 (0.126)	0.022 (0.204)	0.098 (0.133)
Main parent has a disability	-0.135 (0.087)	-0.400 (0.14)**	-0.445 (0.091)**
Grandparents went to university	0.847 (0.128)**	1.718 (0.206)**	0.692 (0.132)**
English as main language	0.568 (0.192)**	0.815 (0.292)**	-0.029 (0.192)
Child has a disability	-1.251 (0.101)**	-1.996 (0.164)**	-1.198 (0.106)**
N. older siblings	-0.372 (0.033)**	-0.743 (0.053)**	-0.492 (0.034)**
Mother senior high school graduate	-1.624 (0.118)**	-3.485 (0.192)**	-1.824 (0.122)**
Mother junior high school graduate	-3.067 (0.150)**	-5.968 (0.243)**	-3.427 (0.156)**
Mother no qualification	-3.524 (0.140)**	-6.822 (0.225)**	-3.746 (0.145)**
Birth-weight in kg	0.468 (0.068)**	0.672 (0.109)**	0.305 (0.071)**
Sex - Male	-0.341 (0.070)**	-0.813 (0.113)**	-0.940 (0.074)**
Single parent household at birth	-0.964 (0.086)**	-2.070 (0.139)**	-1.496 (0.090)**
Mother<20 at birth	-1.303 (0.146)**	-2.489 (0.233)**	-1.796 (0.152)**
Month of birth			
October	-0.314 (0.173)+	-0.640 (0.28)**	-0.375 (0.182)*
November	-0.287 (0.175)	-0.509 (0.282)+	-0.377 (0.183)*
December	-0.658 (0.173)*	-0.987 (0.278)**	-0.501 (0.182)**
January	-0.562 (0.172)*	-0.788 (0.276)**	-0.382 (0.18)*
February	-0.900 (0.171)*	-1.152 (0.277)**	-0.698 (0.18)**
March	-0.673 (0.171)*	-0.772 (0.275)**	-0.417 (0.179)*
April	-1.323 (0.172)*	-1.527 (0.277)**	-0.864 (0.181)**
May	-1.290 (0.168)*	-1.537 (0.272)**	-0.719 (0.177)**
June	-1.275 (0.168)*	-1.497 (0.271)**	-0.695 (0.176)**
July	-1.362 (0.168)*	-1.500 (0.271)**	-0.577 (0.176)**
August	-1.629 (0.167)**	-2.092 (0.27)**	-0.992 (0.176)**
Ethnic background			
Black	-1.028 (0.153)**	-1.834 (0.235)**	-0.168 (0.155)
Asian	0.210 (0.13)	1.077 (0.207)**	1.621 (0.136)**
Mixed	0.119 (0.135)	0.257 (0.216)	0.511 .141)**

Table A2 - Effect of pre-school on cognitive outcomes. Results expressed as % of SD

	Model 1		Model 2		Model 3
	OLS	**PSM**	**OLS**	**PSM**	**OLS**
Age 11					
KS2 Overall	15%	15%	7%	7%	NA
KS2 Maths	12%	13%	5%	6%	NA
KS2 English	17%	13%	9%	10%	NA
KS2 Science	9%	10%	4%	2%	NA
Age 14					
KS3 Overall	16%	20%	8%	5%	8%
KS3 Maths	13%	9%	5%	5%	11%
KS3 English	15%	8%	8%	8%	7%
KS3 Science	14%	15%	6%	3%	6%
Age 16					
# GCSE Subjects with grade A*-C	15%	16%	7%	8%	7%

Table A3 shows results from balance tests of the estimation performed with Model 2 (Outcome: Number of GCSE with A*-C). The output shows two rows for each variable—unmatched and matched. In each row, it shows the mean of the variable for the treatment group and the mean for the control group. It also shows the "%bias," which is the standardized bias. This "bias" is defined as the difference of the mean values of the treatment group and the (not matched / matched) non treatment group, divided by the square root of the average sample variance in the treatment group and the not matched non treatment group. The table also shows the % reduction in bias, which is how much of this bias was eliminated by matching. In our example, we have very few variables exhibiting negative values for this column (meaning that the bias increased as a result of matching) and these are mostly cases in which the bias was already very low before matching.

To assess balance, one should look at both the bias and the mean differences between treatment and control in the matched sample. In our example, the bias is significantly reduced after matching (the mean goes from 5.38 to 1.65)

The last two columns presents results from a t-test on the hypothesis that the mean value of each variable is the same in the treatment group and the non-treatment group. It is done before and after matching. If $p>0.1$, the null hypothesis cannot be rejected on the 10% significance level. The null hypothesis that the mean values of the two groups do not differ after matching cannot be rejected for most of the variables included in our analysis. By matching, the differences between treatment group and non-treatment group are reduced considerably.

Table A3 Balance tests for Propensity Score Matching

Summary of the distribution of the absolute bias

Before matching: Mean = 5.38; SD=5.90 After matching: Mean = 1.65; SD=1.14

Variable	Sample	Mean - Treated	Mean Control	% Bias	% Bias Reduction	t	p
Birth-weight	Unmatched	3.3207	3.2891	5.3		2.07	0.038
	Matched	3.3205	3.3291	-1.4	72.8	-1.03	0.302
Premature	Unmatched	0.10412	0.11284	-2.8		-1.09	0.276
birth	Matched	0.10421	0.09116	4.2	-49.7	3.07	0.002
Main parent	Unmatched	0.20382	0.2228	-4.6		-1.80	0.072
has disability	Matched	0.20401	.20329	0.2	96.2	0.12	0.901
Grandparents	Unmatched	0.088	0.08693	0.4		0.14	0.885
university	Matched	0.08798	0.07729	3.8	-903.4	2.71	0.007
English main	Unmatched	0.95913	0.91883	16.9		7.33	0.000
language	Matched	0.9592	0.95642	1.2	93.1	0.96	0.336
Child has a	Unmatched	0.1378	0.14393	-1.8		-0.68	0.496
disability	Matched	0.13792	0.13392	1.2	34.6	0.82	0.415
N. older	Unmatched	0.90132	1.076	-15.4		-6.11	0.000
siblings	Matched	0.89969	0.90946	-0.9	94.4	-0.65	0.519
Mother HE	Unmatched	0.12784	0.12666	0.4		0.14	0.892
sub degree	Matched	0.12795	0.12806	-0.0	91.3	-0.02	0.983
Mother	Unmatched	0.13574	0.10708	8.8		3.26	0.001
A level	Matched	0.13587	0.14491	-2.8	68.4	-1.82	0.069
Mother	Unmatched	0.30496	0.25561	11.0		4.15	0.000
GCSE A-C	Matched	0.30514	0.30432	0.2	98.3	0.12	0.901
Mother	Unmatched	0.9601	0.9614	-0.0		-0.02	0.986
GCSE <C	Matched	0.9609	0.10925	-4.5	-9500.6	-3.02	0.003
Mother	Unmatched	0.01715	0.02245	-3.8		-1.53	0.125
Other qual	Matched	0.01716	0.01552	1.2	69.0	0.90	0.366
Mother	Unmatched	0.17497	0.2844	-26.2		-10.73	0.000
No qual	Matched	0.17503	0.16465	2.5	90.5	1.93	0.054
Mother	Unmatched	0.03296	0.02879	2.4		0.91	0.365
education -	Matched	0.03299	0.3083	1.2	48.3	0.86	0.392
missing							
Birth-weight	Unmatched	3.3207	3.2891	5.3		2.07	0.038
in kg	Matched	3.3205	3.3291	-1.4	72.8	-1.03	0.302
Sex - Male	Unmatched	0.51001	0.5141	-0.8		-0.31	0.753
	Matched	0.51028	0.52127	-2.2	-168.6	-1.53	0.125
Single parent	Unmatched	0.22292	0.6252	-9.2		-3.62	0.000
at birth	Matched	0.22292	0.21655	1.5	83.9	1.07	0.283
Mother < 20	Unmatched	0.06181	0.08751	-9.8		-3.98	0.000
at birth	Matched	0.06187	0.06413	-0.9	91.2	-0.65	0.516

Note: Month of birth results not reported.

Another way to check how the treatment and control groups differ is to graph the propensity scores for the two groups group. Figure 1 shows a histogram of the propensity scores while Figure 2 shows a kernel density estimate of propensity scores for treatment and control group.

Figure 1 - Histogram of propensity scores of treatment vs. control group

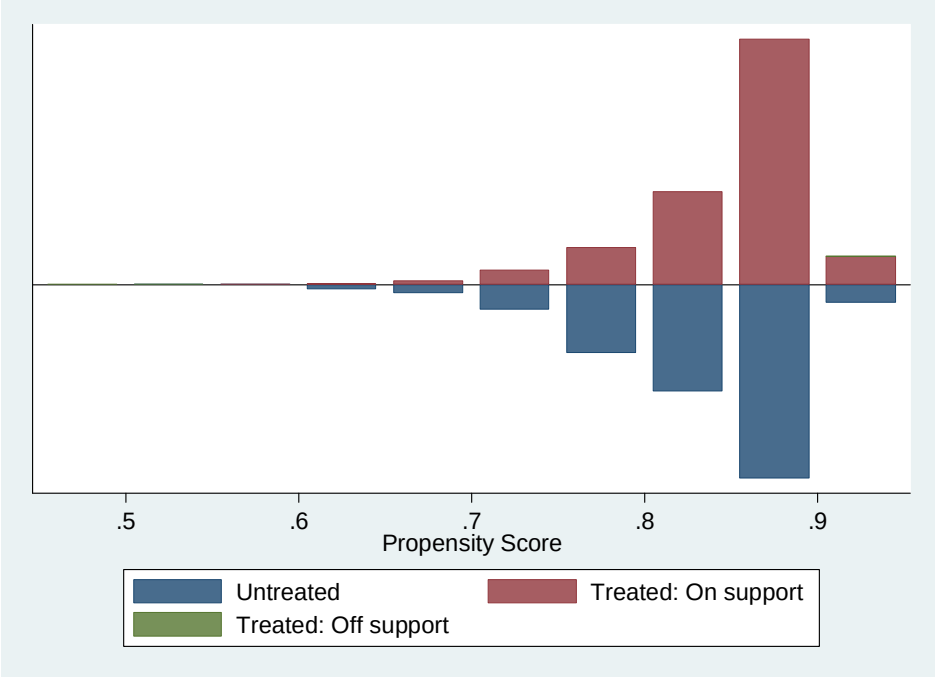
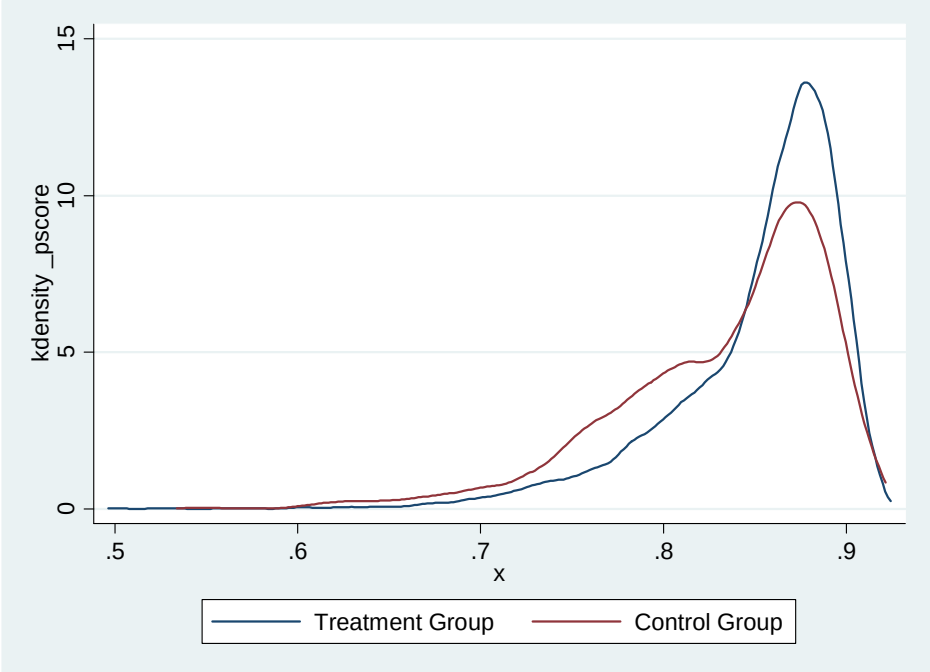


Figure 2 - Kernel graphs of propensity score for treated and control group



References

Berlinski S., Galiani S., and P. Gertler (2009), The Effect of Pre-Primary Education on Primary School Performance. *Journal of Public Economics* 93, 219-234.

Bernal, R. (2008), The Effect of Maternal Employment and Childcare on Children's Cognitive Development. *International Economic Review* 49, 1173-1209.

Bernal R. and, M.P. Keane (2010), Quasi-Structural Estimation of a Model of Childcare Choices and Child Cognitive Ability Production. *Journal of Econometrics*, 156, 154-189.

Bernal R., and M.P. Keane (2011), Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers. *Journal of Labor Economics*, 29, 459-512.

Berger L.M., Hill J., and J. Waldfogel (2005), Maternity Leave, Early Maternal Employment and Child Health and Development in the US. *Economic Journal*, 115, 29-47.

Blau, D., and J. Currie (2006), Pre-School, Day Care, and After School Care: Who's Minding the Kids? In: Hanushek, Eric, Welch, Finis (eds.), *Handbook of the Economics of Education*, Vol. 2. Elsevier, Amsterdam, 1163–1278

Blau, D., (1999), "The Effects of Child Care Characteristics on Child Development". *The Journal of Human Resources*, 34, 261-276.

Carneiro, P., and J. Heckman (2003), Human Capital Policy. In: Heckman, J., Krueger, A. (Eds.), *Inequality in America: What role for human capital policies?* MIT Press

Cunha, F., Heckman, J., Lochner, L., and D. Masterov (2006), Interpreting the Evidence on Life Cycle Skill Formation. In: Hanushek, E., and Welch, F. (eds.), *Handbook of the Economics of Education*, North-Holland, Amsterdam, 697- 812.

Currie, J. (2001), Early Childhood Education Programs. *Journal of Economic Perspectives*, 15, 213–38.

Dehejia, R. and S. Wahba (2002), Propensity Score Matching Methods for Non-Experimental Causal Studies. *Review of Economics and Statistics*, 84, 151–61.

Fenstein L., Robertson, D. and J. Symons (1998), Pre-School Education and Attainment in the NCDS and BCS, Centre for Economic Performance, Human Resources Programme Paper no. 382.

Goodman A., and B. Sianesi (2005), Early Education and Children's Outcomes: How Long Do the Impacts Last? *Fiscal Studies* 26, 513-548.

Heckman, J.J. and D.V. Masterov (2007), The Productivity Argument for Investing in Young Children, NBER Working Paper 13016.

Keane, M., and K. Wolpin, (2001), The Effect of Parental Transfers and Borrowing Constraints on Education Attainment. *International Economic Review* 42, 1051-1103.

Keane, M., and K. Wolpin (2006), The Role of Labor and Marriage Markets, Preference Heterogeneity and the Welfare System in the Life Cycle Decisions of Black, Hispanic and White Women, Working Paper, University of Pennsylvania.

Lamb, M. (1996), Effects of Nonparental Child Care on Child Development: An Update. *Canadian Journal of Psychiatry*, 41, 330-342.

Loeba,S., Bridges M., Bassoka, D.Fuller B. and R.W. Rumbergerd (2007), How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review* 26, 52-66.

Magnuson K., Ruhm C. and J. Waldfogel (2007), Does prekindergarten improve school preparation and performance. *Economics of Education Re*view 26, 33-51.

McCartney, K. (1984), The Effect of Quality of Day Care Environment upon Children's Language Development. *Developmental Psychology* 20, 244-260.

Melhuish, E.C. (2003), A Literature Review of the Impact of Early Years Provision on Young Children, with Emphasis Given to Children from Disadvantaged Backgrounds, National Audit Office.

NESS (2010), The impact of Sure Start Local Programmes on Five Year Olds and Their Families, The National Evaluation of Sure Start (NESS) Team, Institute for the Study of Children, Families and Social Issues, Birkbeck University of London. Available online at http://www.ness.bbk.ac.uk/impact/documents/RR067.pdf

Osborn, A. F., and J.E. Milbank (1987), *The Effects of Early Education: a Report from the Child Health and Education Study*. Oxford, Clarendon Press.

Ruhm, C.J. (2008), Maternal Employment and Adolescent Development. *Labour Economics* 15, 958-983.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., and B. Taggart (2004), The Final Report: Effective Pre-School Education. The Effective Provision of Pre-School Education (EPPE) Project, Technical Paper 12. London: DfES / Institute of Education, University of London.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., and B. Taggart (2012), Effective Pre-school, Primary and Secondary Education 3-14 Project (EPPSE 3-14) - Final Report from the Key Stage 3 Phase: Influences on Students' Development from Age 11-14, London: DfES / Institute of Education, University of London.

Vandell, D.L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., and NICHD Early Child Care Research Network (2010), Do Effects of Early Child Care Extend to Age 15 Years? Results From the NICHD Study of Early Child Care and Youth Development. *Child Development*, 81, 1467-8624.

Waldfogel, J. (1999), Early Childhood Interventions and Outcomes, CASE paper, no. 21.