Institute for Fiscal Studies

The drivers of month of birth differences in children's cognitive and non-cognitive skills: a regression discontinuity analysis

IFS Working Paper W13/08

Claire Crawford
Lorraine Dearden
Ellen Greaves

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

The drivers of month of birth differences in children's cognitive and non-cognitive skills: a regression discontinuity analysis[1]

Claire Crawford[2], Lorraine Dearden[3] and Ellen Greaves[4]

May 2013

This paper uses data from a rich UK birth cohort to estimate the differences in cognitive and non-cognitive skills between children born at the start and end of the academic year. It builds on the previous literature on this topic in England by using a more robust regression discontinuity design and is also able to provide new insight into the drivers of the differences in outcomes between children born in different months that we observe. Specifically, we compare differences in tests that are affected by all three of the potential drivers (age at test, age of starting school and relative age) with differences in tests sat at the same age (which are therefore not affected by the age at test effect) as a way of separately identifying the age at test effect. We find that age at test is the most important factor driving the difference between the oldest and youngest children in an academic cohort; highlighting that children born at the end of the academic year are at a disadvantage primarily because they are almost a year younger than those born at the start of the academic year when they take national achievement tests. An appropriate policy response in this case is to appropriately age-adjust these tests. However, we also find evidence that a child's view of their own scholastic competence differs significantly between those born at the start and end of the academic year, even when eliminating the age at test effect. This means that other policy responses may be required to correct for differences in outcomes amongst children born in different months, but not necessarily so: it may be that children's view of their scholastic competence would change in response to the introduction of appropriately age-adjusted tests, for example as a result of positive reinforcement.

Key words: Month of birth, regression discontinuity design

JEL classification: I21, J24

---

[1] The authors gratefully acknowledge funding from the Nuffield Foundation (grant number EDU/36559) and from the Economic and Social Research Council (ESRC) via the Centre for the Microeconomic Analysis of Public Policy at IFS (RES-544-28-0001) and RES-060-23-0011. The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org. The authors would like to extend particular thanks to Kate Northstone, part of the team running the Avon Longitudinal Study of Parents and Children based at the University of Bristol, for her invaluable support in providing access to exact date of birth for our analysis, and to Anna Vignoles for helpful comments and advice. All errors remain the responsibility of the authors.

[2] Institute for Fiscal Studies.

[3] Institute for Fiscal Studies and Institute of Education, University of London.

[4] Corresponding author: ellen_g@ifs.org.uk, Institute for Fiscal Studies, 7 Ridgmount Street, London, WC1E 7AE, UK.

1 Introduction

It is well known that children born at the start of the academic year tend to achieve better exam results, on average, than children born at the end of the academic year (Fredriksson & Ockert, 2005; Bedard & Dhuey, 2006; Datar, 2006; Puhani & Weber, 2007; Black, Devereux & Salvanes, 2008; Smith, 2010). In England, where the academic year runs from 1st September to 31st August, this means that children born in the autumn tend to outperform those born in the summer (Russell & Startup, 1986; Sharp, Hutchison & Whetton, 1994; Thomas, 1995; Alton & Massey, 1998). Our own previous research (Crawford, Dearden & Meghir, 2007), based on administrative data held by the UK Department for Education, used a linear regression model to show that August-born children score, on average, over half a standard deviation lower than their September-born counterparts in national achievement tests at age 7. This difference decreases over time, but is still significant at age 16, when young people are making decisions about whether to stay on for post-compulsory education. At just over 10% of a standard deviation, this gap translates into a 5.8 percentage point reduction in the likelihood that August-borns will reach the Government's expected level of 5 GCSEs at grades A*-C, which is usually regarded as the standard required to stay on. Crawford, Dearden & Meghir (2010) also find evidence that August-borns are 1.5 percentage points less likely to continue into higher education at age 18 or 19 than those born in September (see also HEFCE, 2005).

Given the importance of educational attainment in determining a range of later-life outcomes, these differences mean that the month in which individuals are born has the potential to affect them throughout their lives. Moreover, educational attainment is not the only outcome that might differ by month of birth: while there has been relatively less work investigating the effect of month of birth on other skills and behaviours, there are at least two reasons why such differences are likely to be important and of interest: first, because they may affect children's well-being in the short-term; for example, being amongst the youngest (and perhaps also the smallest) in your class may increase your chances of being bullied or lower your self-esteem. Second, because they may have potentially serious long-term consequences for children's lives, and not just via their effect on educational attainment. For example, if being amongst the youngest or smallest in your class affects your enjoyment of school and/or your motivation and determination to do well and/or you belief in your own academic ability, then the month in which you were born may have short- and longer-term consequences far beyond those captured by educational attainment alone.

To date, what little research there has been in this area has tended to focus on the likelihood of being classified as having special educational needs (Goodman et al., 2003; Crawford et al., 2007; Elder & Lubotsky, 2009; DfE, 2010; Dhuey & Lipscomb, 2010) and the likelihood of being bullied (DfE, 2010; Muhlenweg, 2010). Our own previous research (Crawford, Dearden & Greaves, 2012) built on this literature by investigating month of birth differences in a wide range of skills and behaviours, both cognitive and non-cognitive, using three UK cohort studies, but it did so by imposing parametric assumptions and was not able to investigate the source of the observed differences in outcomes, which is of paramount policy importance.

To identify the appropriate policy response, it is necessary to identify the drivers of the month of birth differences in skills and behaviours, which arise because of the interaction between a pupil's date of birth and the school admissions policy that they face. There are four main drivers of these differences: first, in a system in which exams are taken at a fixed date, as is the case in England, some children will be up to a year younger than others when they sit the test (referred to as the "absolute age" or "age of sitting the test" effect). Second, those born just before the discontinuity may be disadvantaged by the fact that they started school considerably younger than their peers (the "age of starting school" effect). Third, age relative to classroom or year group peers may adversely affect some children, for example if explicit comparison between children in the same class or year group negatively affects the self-belief of younger children (the "rank" or "relative age" effect). Finally, depending on the admissions system, some children born towards the end of the academic year may have attended school for fewer terms prior to the exam than those born towards the start of the academic year (the "length of schooling" effect).

While many of the papers outlined above report that the differences they observe are driven by one of these effects – usually the age of starting school or length of schooling effect – in many cases it is not clear how or indeed whether they have succeeded in separately identifying a particular effect. For example, Bell & Daniels (1990) report that the relative age effect explains the greatest proportion of the variance in test scores that they find, although it is not clear how they separate the relative age from the absolute age effect. Similarly, Fogelman & Gorbach (1978) combine regional variation in school admissions policies with data on a cohort of children born in Great Britain in a particular week in March 1958 (from the National Child Development Study) in an attempt to identify the impact of length of schooling on test scores at age 11. The fact that their sample are all born in the same week effectively enables them to eliminate the absolute and relative age effects. However, children who receive different amounts of schooling will also have started school at slightly different ages, hence their identification strategy would seem to lend itself to estimating a combined effect of the two.

Other studies have made more direct attempts to separate these effects. For example, Datar (2006) uses data from the Early Childhood Longitudinal Study in the US to look at the impact of expected school starting age on reading and maths test scores when children are in kindergarten, and then again two years later. Datar finds that the oldest starters score 0.8 standard deviations higher in maths and 0.6 standard deviations higher in reading than the youngest entrants in the kindergarten class. However, these estimates identify the combined impact of age of starting school and absolute age. To separately identify the age of starting school effect, she uses differences in test scores over time as her dependent variable. Under the assumption that the absolute age effect is linear – i.e. that the difference in test scores between two children who are six months apart in age is the same regardless of how old those children are – this approach should difference out the effect of absolute age on test scores, leaving only the age of starting school effect. Datar finds that the test scores of older entrants increase by 0.12 standard deviations over and above those of the youngest entrants over a two-year period, implying that it is better for children to start kindergarten when they are older.

Smith (2010) also gives serious consideration to this issue, using an identification strategy very similar to that adopted in our previous work (Crawford et al., 2007) and discussed in more detail below. Smith takes advantage of a temporary change in the school admissions policy in place in British Columbia to estimate an upper bound on the age at test effect and a lower bound on the age of starting school effect (neither of which can be separated from the length of schooling effect). Using administrative data on grade repetition at grade 3 (age 8-9) and literacy and numeracy scores at grade 10 (age 15-16), he finds relatively large age at test effects and relatively small age of starting school effects.

Crawford et al. (2007) take advantage of the fact that school admissions policies in England are set by local, rather than central, authorities, meaning that there is considerable regional and temporal variation in the age at which children born on a particular day start school (and hence the amount of schooling they receive prior to the tests). This identification strategy relies on making comparisons across areas, which requires large sample sizes and means that it is very important to account for any differences across areas that might affect test scores. Because the date on which children start school also dictates the number of terms of schooling they receive prior to the test, it is not possible to separate the effect of starting school younger from the effect of receiving an additional term of schooling, although it is possible to separate the combination of these two effects from the age at test (absolute age) effect, which they do by imposing parametric assumptions on their model.

In line with Smith (2010), Crawford et al. (2007) find that it is the age at test effect that matters most: at age 7, they find a small negative effect of starting school slightly older (and receiving one less term of schooling prior to the tests) of around 5% of a standard deviation; however, this effect is dwarfed by the age at test effect (which can be calculated – assuming linearity – by subtracting the combined age of starting school and length of schooling effects from the total effect) of around 55% of a standard deviation. Moreover, the age of starting school/length of schooling effect has disappeared completely by age 14.

This paper will add to the existing literature in this area in two key ways: first, it will use an alternative identification strategy – one which does not rely on comparisons across areas or the use of parametric regression models – to identify whether it is indeed the age at test effect that is the main driving force behind the month of birth differences in educational attainment that we observe in England. Specifically, we will take advantage of a rich cohort dataset covering pupils across three academic years to undertake a regression discontinuity analysis. This will not only enable us to estimate the differences in attainment between children born at the start and end of the academic year in a more precise fashion and using fewer assumptions than have been made in most previous analyses based on British data, but by taking advantage of the fact that these cohort members undertook a variety of similar tests at different points in time, we will be able to identify various combinations of the absolute age, age of starting school, length of schooling and relative age effects. Moreover, under certain assumptions regarding the comparability of these tests, we will be able use these combinations to identify a lower bound of the age at test effect. Second, it will make use of a similar strategy to identify the drivers of the month of birth differences in a range of non-cognitive skills first documented by Crawford et al (2011).

This evidence is vital in understanding how best to respond to the month of birth differences that we observe: if age at test is the main driving force behind the differences in outcomes, then a simple age-adjustment of the relevant tests may be an appropriate policy response. If other mechanisms also drive the differences we observe, however, then a more comprehensive policy response may be required to address the differences between children born in different months of the year.

This paper now proceeds as follows: Section 2 reviews the assumptions underlying the regression discontinuity design and discusses our identification strategy in more detail. Section 3 describes the data that we use. Section 4 presents our results. Section 5 concludes.

2 Methodology and identifying assumptions

Regression Discontinuity Design

Our aim is to identify the impact on a range of cognitive and non-cognitive skills of being born at the start rather than the end of the academic year. This problem can be thought of as an experiment, where the "treatment" is being the oldest in the academic year. Following standard notation we denote potential outcome variables under treatment and no treatment as Y_1 and Y_0 respectively. The evaluation problem arises because pupils are born at either the start or end of the academic year – they either do or do not receive the treatment – and hence it is impossible to observe both Y_1 and Y_0 for any given individual.

Many evaluation techniques have been developed to address this problem, which usually involve the construction of an appropriate control group whose outcomes represent the counterfactual outcomes for those in the treatment group (see Blundell and Costa Dias (2009) for a recent review). Regression discontinuity design (RDD) is often regarded as the quasi-experimental technique that comes closest to the experimental "gold standard" (the randomised experiment) in appropriate applications (Lee and Lemieux, 2010). RDD provides a way of identifying mean treatment effects for a subgroup of the population (close to the discontinuity) under minimal assumptions (Hahn et al, 2001). For example, parametric assumptions are not necessary, and the requirement to choose appropriate control variables (and their functional form) is removed. The limitation of RDD in some circumstances is that identification is relevant only for a sub-section of the population (close to the discontinuity), but in many cases, including this one, this identifies a policy relevant parameter.

RDD has the defining characteristic that the probability of receiving the treatment changes discontinuously as a function of one or more underlying variables. Under certain conditions, detailed below, the allocation of treatment on the basis of this underlying variable is analogous to assignment in a randomised experiment, and the causal effect of the treatment at the point of discontinuity is recovered. In our application, the probability of receiving treatment (being the oldest in the academic year) is determined by date of birth (*Z*) and varies discontinuously at 31st August / 1st September; those born on 31st August are the youngest in the academic year, while those born on 1st September are the oldest and hence receive the "treatment". We denote treatment status by the binary variable *T*, where *T=1* denotes treatment and *T=0* denotes no treatment.

Following Hahn et al. (2001), to identify the causal effect of the treatment using RDD we require the following conditions to be met:

a) ***The probability of treatment must vary discontinuously at some point with respect to Z.*** Formally: $P(T_i = 1|Z_{i-}) \neq P(T_i = 1|Z_{i+})$, where Z_{i-} refers to the region immediately below the discontinuity and Z_{i-} immediately above. We argue that this condition holds in our application, as parents in England have limited ability to change the date at which their child starts school; indeed, in a census of state school pupils in England in 2011, over 99% of pupils are in the "correct" academic year based on their age. This is the case for pupils born in all months of the year, although the proportion of pupils born in August in the "correct" academic year is lower than the proportion for those born in September (those born in August are much more likely to be in a lower academic year).

b) ***Pupils' characteristics (aside from date of birth) must be continuous at the point of discontinuity.*** Formally: $E(A_i|Z_i = Z)$ is continuous in Z at the discontinuity, where A_i represents all characteristics of pupils that affect the outcome of interest. This assumption ensures that other factors are not responsible for any differences in outcomes observed between the treatment and non-treatment groups. This assumption is often validated graphically, by comparing the characteristics of those either side of the discontinuity. We present a selection of evidence illustrating that there are no other obvious or significant discontinuities in other characteristics at the point of discontinuity in Appendix A.

c) ***Individuals do not select into the treatment on the basis of anticipated gains from treatment.*** While a pupil has no power to manipulate their date of birth, parents have some means to manipulate the month in which their child is born (either through conception or birth decisions). Some studies have found systematic differences in the number (e.g. Gans & Leigh, 2009) or family background characteristics (e.g. Buckles & Hungerman, 2008) of children born either side of the discontinuity, which might result from sorting of this kind. However, we find no evidence of such sorting in our sample; there are 927 pupils born in August and 932 born in September in the middle cohort in our data. Also, there are very few significant differences between individuals born either side of the discontinuity, i.e. that there is no evidence of systematic sorting that would invalidate this assumption for our sample; Table 1 presents the marginal effects from a probit regression model in which the outcome variable is a binary indicator equal to one if the individual was born in September and zero if they were born in August.

Table 1 Background characteristics of those born in August and September in ALSPAC: probit regression reporting marginal effects where dependent variable is 'August-born'

ALSPAC	*Child born in September relative to August*
Male	0.002 [0.001]
Lowest household income quintile	0.009 [0.007]
2nd household income quintile	0.004 [0.004]
3rd household income quintile	0.007 [0.005]
4th household income quintile	0.011 [0.006]
Child's ethnicity: White British	-0.008 [0.007]
Child speaks English as an additional language (EAL)	-0.005 [0.003]
Lone parent at 32 weeks' gestation	-0.006* [0.002]
Cohabiting at 32 weeks' gestation	-0.003 [0.002]
Household work: mother in work at age 3	0.001 [0.002]
Household work: father in work at age 3	-0.004 [0.005]
Mother's highest educational qualification (NVQ): CSE	0.004 [0.005]
Mother's NVQ: vocational	0 [0.004]
Mother's NVQ: O level	0.002 [0.004]
Mother's NVQ: A level	0.003 [0.004]
Father's highest educational qualification (NVQ): CSE	0.001 [0.003]
Father's NVQ: vocational	0.010 [0.007]
Father's NVQ: O level	-0.004 [0.002]
Father's NVQ: A level	-0.001 [0.003]
Mother's class: ii	-0.001 [0.004]
Mother's class: iii (non-manual)	-0.003 [0.004]
Mother's class: iii (manual)	-0.003 [0.003]
Mother's class: iv	0.005 [0.008]
Mother's class: v	-0.002 [0.005]
Father's class: ii	0.002 [0.003]

ALSPAC	*Child born in September relative to August*
Father's class: iii (non-manual)	0.003 [0.005]
Father's class: iii (manual)	-0 [0.003]
Father's class: iv	-0.003 [0.003]
Father's class: v	0 [0.006]
Mother's age at birth of child: 30–34	0.007 [0.004]
Mother's age at birth of child: 25–29	0.006 [0.004]
Mother's age at birth of child: 20–24	0.003 [0.004]
Mother's age at birth of child: under 20	-0.001 [0.004]
Ever lived in social housing	-0 [0.002]
Always owned/mortgaged home	0.001 [0.002]
Financial difficulties	0.004 [0.003]
Child was not breastfed	-0.001 [0.002]
Household smokes around child	0.002 [0.002]
Child's birth weight was low	-0.006** [0.002]
Child's birth weight was high	-0.004 [0.003]
Child was one of a multiple birth	0.006 [0.007]
Birth order within household: 2nd	-0.004* [0.002]
Birth order within household: 3rd	-0.005* [0.002]
Birth order within household: 4th or higher	-0.006** [0.002]
N	2,845
Joint significance test: gender	0.151
Joint significance test: income	0.086
Joint significance test: ethnicity	0.103
Joint significance test: EAL	0.192
Joint significance test: household status at birth	0.139
Joint significance test: working	0.605
Joint significance test: mother's NVQ	0.715
Joint significance test: father's NVQ	0.019
Joint significance test: mother's class	0.215
Joint significance test: father's class	0.538
Joint significance test: mother's age at birth of child	0.133
Joint significance test: social housing	0.910
Joint significance test: own/mortgage home	0.795
Joint significance test: financial circumstances	0.095
Joint significance test: breastfeeding	0.454
Joint significance test: household smoking	0.214
Joint significance test: birth weight	0.013
Joint significance test: multiple birth	0.303
Joint significance test: birth order	0.009

Notes: Marginal effects reported. Standard errors reported in brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ The reference categories are as follows: female, highest household income quintile, child's ethnicity is not White British, child does not speak English as an additional language, married at 32 weeks' gestation, mother and father born in work at age 3, mother has at least a degree level qualification, father's has at least a degree level qualification, mother has the highest social class ranking, father has the highest social class ranking, mother over 35 at birth, never lived in social housing, not always owned/had mortgage for home, no reported financial difficulties, child was breastfed, household doesn't smoke around the child, child had a normal birthweight, child was not a multiple birth, child is oldest of siblings in the household.

The average causal effect of being the oldest in the academic year can be found for those around the discontinuity at $\overline{Z}$ (under the assumptions above) and, when the discontinuity is "sharp", is given by:

$$E(\beta_i|Z_i = \overline{Z}) = E(Y_i|Z_i = \overline{Z_{i+}}) - E(Y_i|Z_i = \overline{Z_{i+}})$$

The necessary assumptions are more likely to hold in a small region around the discontinuity, $\overline{Z}$, suggesting that a small window should be used. Including observations from a larger region increases the sample size, however, which implies a trade-off between statistical power and calculating unbiased estimates.

The average causal effect can be estimated non-parametrically (for example using local linear non-parametric or kernel regression), or parametrically by approximating the continuous underlying function of Z using a polynomial of varying degrees. Our main specification controls parametrically for distance to the discontinuity using a quadratic polynomial which we allow to vary either side of the discontinuity.

Most outcomes we consider are continuous (standardised to have a mean of zero and a standard deviation of one). In each case, the average causal effect is interpreted as the effect of being born just after the discontinuity (in early September) relative to being born just before the discontinuity (in late August) in standard deviations. One outcome is binary, and the coefficient in this case is interpreted as the percentage point impact of being born to the right of the discontinuity. Our outcome variables are discussed in more detail below.

Our final specification has the following form:

$$Y_i = \beta_1 \mathrm{Z}_i + \beta_2 \mathrm{Z}_i^2 + \beta_3 T_i + \beta_4 \mathrm{Z}_i T_i + \beta_5 \mathrm{Z}_i^2 T_i + \beta_6 \boldsymbol{X}_i + \varepsilon_i$$

For each ALSPAC cohort member *i*, Y_i represents their outcome of interest. T_i is a binary variable that represents the "treatment" of being born after the discontinuity; it is equal to one if a child is among the oldest in the academic year, and zero if they are among the youngest. Z_i represents the distance from the discontinuity, the "assignment variable" referred to above. For example, Z_i is equal to zero if cohort member *i* was born on 1st September (in any cohort), one if the cohort member was born on 2nd September, and minus one if the cohort member was born on 31st August. The addition of the squared term Z_i^2 allows the impact of a child's day of birth to affect their outcome non-linearly. As stated above, we assume that child's date of birth has a smooth impact on their outcomes, and that in the absence of a discontinuity on 1st September there would be no "jump" in outcomes. The inclusion of interaction terms $\mathrm{Z}_i T_i$ and $\mathrm{Z}_i^2 T_i$ allows the impact of the "assignment variable" to vary either side of the discontinuity to increase the flexibility of the specification. $\boldsymbol{X}_i$ represents a vector of observable pupil and parent characteristics that may affect attainment (described in more detail below), and ε_i represents unobservable and random factors that may also affect attainment.

Our preferred window around the discontinuity is 30 days and, as described above, we choose to use a quadratic specification for the assignment variable, but our results are robust to alternative choices of window size and higher or lower order polynomials (see Appendix B). We also investigate the assumption that the assignment variable is smooth in the absence of a discontinuity by running a series of placebo tests (see Appendix C). Following Card & Lee (2008), we cluster standard errors by day of birth (as is also done by Berlinski et al., 2009; Brewer & Crawford, 2010; Fitzpatrick, 2010).

Identifying the drivers of the discontinuity effects

The treatment effect β_3, which represents the difference in test scores between those born just in September relative to late August, is a function of the following variables: age of starting school, length of schooling prior to the test, age at test, and relative age within cohort. The high correlation between age at test, age of starting school and relative age, and indeed the exact relationship between age of starting school, length of schooling and age at test prohibit this function being estimated by linear regression. Our strategy for identifying the main drivers of the month of birth differences in outcomes to overcome this problem is possible because the children in our dataset took a variety of similar tests at different points in time. (The dataset and specific tests used are described in more detail below.) Specifically, while those born at the start and end of the academic year all sit national achievement tests on the same date (and hence are almost a year apart in age, in addition to having started school at very different ages and sitting at very different ends of the relative age distribution within their school), they are subjected to some very similar tests as part of the survey, this time taken around the time of their birthday. This means that, in contrast to the national achievement tests, children born at the start and end of the academic year are very similar in age when they take these survey tests (shown in Table A2), but will have started school at very different ages and will be of different relative ages.

For national achievement tests β_3 is a function of the following variables: age of starting school (where children born in September start school older), age at test (where children born in September are older when assessed), and relative age within cohort (where children born in September are older relative to their peers). Note that β_3^{N} is not a function of length of schooling, as all children in our sample (in the Avon area) started school in the September in the academic year in which they turned five, and were assessed at the same point in time.

For survey tests β_3 is a function of the following variables: age of starting school (where children born in September start school older), length of schooling prior to the test (where children born in September have fewer terms of schooling before the test), and relative age within cohort (where children born in September are older relative to their peers). Note that β_3^{S} is not a function of age at test, as all children were assessed around the same age. Length of schooling is expected to have a negative effect on β_3^{S} (the premium attached to being born in early September), as length of schooling is hypothesised to have a positive impact on cognitive outcomes, and children born in September have had fewer terms of schooling when assessed in the survey.

Comparing the estimates of β_3^{N} and β_3^{S} for children born either side of the discontinuity therefore isolates the difference between the impact of age at test and length of schooling, as relative age and age of starting school effects cancel out (assuming separability of the four potential drivers). This means that, under certain assumptions regarding the comparability of the national achievement and survey tests (and that the length of schooling coefficient is positive), we are able use these combinations to identify a lower bound for the age at test effect; the difference between β_3^{N} and β_3^{S} is a lower bound under the assumption that length of schooling positively affects cognitive test outcomes and the four potential drivers are separable.

3 Data

We use data from the Avon Longitudinal Study of Parents and Children (ALSPAC), a longitudinal study that has followed the children of around 14,000 pregnant women whose expected date of delivery fell between 1 April 1991 and 31 December 1992, and who were resident in the Avon area of England at that time. This means that ALSPAC cohort members were born in one of three academic cohorts: 1990–91, 1991–92 and 1992–93. These three cohorts facilitate our regression discontinuity design as we observe pupils on either side of the discontinuity in two cohorts. To increase the power of the analysis, we pool these cohorts so that those immediately surrounding the discontinuity in all cohorts are combined.

ALSPAC cohort members and their families have been surveyed via high-frequency postal questionnaires from the time of pregnancy onwards, with information collected on a wide range of family background characteristics. The cohort members' cognitive and non-cognitive skills have also been assessed at various points throughout childhood via a series of clinics. In addition, cohort members have been linked to their scores in national achievement tests at ages 7, 11, 14 and 16 from administrative data held by the Department for Education (the National Pupil Database), which we standardise according to the cohort specific national distribution of scores.

Outcomes

We focus on the cohort members' scores from national achievement tests taken at ages 7 and 11 and compare these to a measure of the cohort members' IQ taken during a clinic session at age 8.

National achievement tests: pupils were assessed on the basis of reading, writing and maths at age 7 and on the basis of English, maths and science at age 11. We use the total points score at age 7 and the average points score across all three subjects at age 11. We standardise each score according to the cohort specific mean and standard deviation of that test. In this way we account for changes in the test across our three cohorts, as well as the potential selection of the ALSPAC cohort relative to the national population, as our derived scores are relative to the national mean and standard deviation for each child's cohort.

IQ: the measure of cognitive development computed by the survey is the third version of the Weschler Intelligence Scale for Children (WISC) (Wechsler, Golombok and Rust, 1992), designed as a measure of IQ for children between the ages of 6 and 16 and the most widely used cognitive test of its kind (Canivez and Watkins, 1998). WISC has five verbal subtests:

- Information (assessing the child's knowledge);
- Similarities (where similarities between things must be explained);
- Arithmetic (mental arithmetic questions);
- Vocabulary (ascertaining the child's understanding of the meaning of different words)
- Comprehension (where the child is asked questions about different situations, e.g. why are names in the telephone book in alphabetical order?),

and five performance subtests:

- Picture completion (the child must point out what is missing from a series of pictures);
- Coding (where shapes corresponding to different numbers must be copied as quickly as possible within a specified time limit);
- Picture arrangement (where pictures must be ordered to make a meaningful sequence);
- Block design (where pictures of specific patterns of blocks are copied with real blocks)
- Object assembly (which involves putting together puzzles).

We create a total score from these WISC components, which we standardise on the ALSPAC sample to have a mean of zero and a standard deviation of one.

Comparison between measures of cognitive development: as described above, our identification strategy depends on the similarity between cognitive measures taken from national achievement tests and those administered as part of the survey. Fortunately, the national achievement tests taken at age 7 (also known as Key Stage 1) and certain questions used to create the WISC score are very similar.

We report the correlation between scores for each component of Key Stage 1 (KS1) and WISC measures of cognitive development in Table 2, separately for children born in August and September. A high correlation indicates that both measures capture similar information about a child's development. By comparing scores for children born in the same month, we minimise the variation that may be caused by sitting the tests at a different age, for example. In practice, the correlation between components is similar for children born in August and September. The correlations between each KS1 component and WISC raw information, similarities, arithmetic, vocabulary and digit span score are high. While other components (such as the comprehension raw score) are less highly correlated with the national achievement tests, this does not seem to affect the total score, which has the highest correlation. This suggests that the WISC and KS1 scores contain similar information about a pupil's cognitive development, and therefore that our identification strategy should be valid.

Table 2 Correlation between WISC and KS1 scores

	August born children			
	KS1: reading	KS1: writing	KS1: maths	KS1: capped points score
WISC: information raw score	0.54	0.51	0.54	0.56
WISC: similarities raw score	0.41	0.37	0.41	0.42
WISC: arithmetic raw score	0.48	0.47	0.48	0.50
WISC: vocabulary raw score	0.40	0.41	0.40	0.42
WISC: comprehension raw score	0.18	0.17	0.18	0.19
WISC: digit span raw score	0.40	0.38	0.40	0.42
WISC: forwards digit span raw score	0.32	0.30	0.32	0.33
WISC: backwards digit span raw score	0.32	0.29	0.32	0.33
WISC: picture completion raw score	0.20	0.21	0.20	0.22
WISC: coding raw score	0.32	0.32	0.32	0.34
WISC: picture arrangement raw score	0.24	0.22	0.24	0.24
WISC: block design raw score	0.26	0.27	0.26	0.28
WISC: object assembly raw score	0.21	0.19	0.21	0.22
WISC: total score	**0.57**	**0.55**	**0.57**	**0.59**
	September born children			
	KS1: reading	KS1: writing	KS1: maths	KS1: capped points score
WISC: information raw score	0.52	0.48	0.52	0.53
WISC: similarities raw score	0.34	0.33	0.34	0.36
WISC: arithmetic raw score	0.46	0.40	0.45	0.46
WISC: vocabulary raw score	0.36	0.36	0.36	0.38
WISC: comprehension raw score	0.21	0.20	0.21	0.21
WISC: digit span raw score	0.35	0.35	0.36	0.38
WISC: forwards digit span raw score	0.27	0.27	0.27	0.29
WISC: backwards digit span raw score	0.28	0.28	0.29	0.30
WISC: picture completion raw score	0.20	0.21	0.20	0.22
WISC: coding raw score	0.32	0.35	0.32	0.35
WISC: picture arrangement raw score	0.20	0.16	0.21	0.20
WISC: block design raw score	0.29	0.29	0.29	0.31
WISC: object assembly raw score	0.20	0.20	0.20	0.21
WISC: total score	**0.56**	**0.54**	**0.56**	**0.59**

<u>Scholastic competence</u>: a total score created from six items from a shortened form of Harter's Self-Perception Profile for Children (Harter, 1985), which children were asked at age 8. They respond to questions by 'posting' whether the statement was true or not for them in a box. The score is standardised on the whole sample to have mean zero and standard deviation one.

<u>Likes school</u>: binary variable coded to equal one if the young person responds 'not much' or 'no' when asked whether they like school at age 8, and zero otherwise.

<u>Locus of control</u>: a total score from a shortened version of the Nowicki–Strickland Internal–External scale (Nowicki & Strickland, 1973) which makes use of self-reported responses amongst pre-school and primary age children, here administered at age 8. Locus of control captures the perception of the connection between one's actions and their consequences (Rotter, 1966), with a higher score indicating a more external locus of control (i.e. a lower belief that their own actions have consequences, and a stronger belief that fate or destiny is playing a role). The total score is standardised on the whole sample to have mean zero and standard deviation one.

<u>Global self-worth</u>: a total score created from six items from a shortened form of Harter's Self-Perception Profile for Children (Harter, 1985), which children were asked at age 8. They respond to questions by 'posting' whether the statement was true or not for them in a box. The score is then standardised on the whole sample to have mean zero and standard deviation one.

Summary statistics for our sample can be found in Appendix D.

Controls

To ensure that the individuals we are comparing are as similar as possible – as well as to improve the precision of our estimates – we include a variety of individual and family background characteristics in our models (although their inclusion does not fundamentally change our results). These variables are:

Cohort member: gender, ethnicity, first language, birthweight, whether they were breastfed, whether they were part of a multiple birth, birth order, whether they are around smokers at home.

Mother and father: highest educational qualification, social class during pregnancy, work status when the child is age 3, mother's age at birth of child.

Household: parents' marital status during pregnancy, income at age 3, ever lived in social housing, always owned own home (or had a mortgage), ever reported financial difficulties.

See Crawford et al (2011) for further discussion of these variables.

Frolich (2007) and others show that the identifying assumptions underlying regression discontinuity analysis hold in the presence of additional covariates.

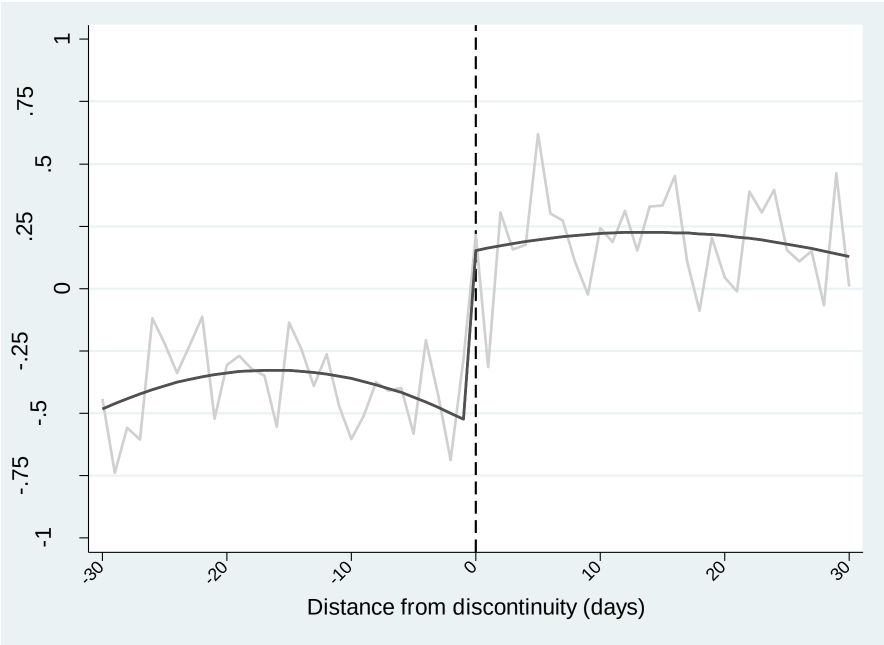
4 Results

Cognitive skills

This section uses the regression discontinuity approach described above to document differences in national achievement test scores at ages 7 and 11, and compares these results to differences in cognitive skills measured using the WISC in the ALSPAC survey in order to better understand the drivers of the differences that we observe.

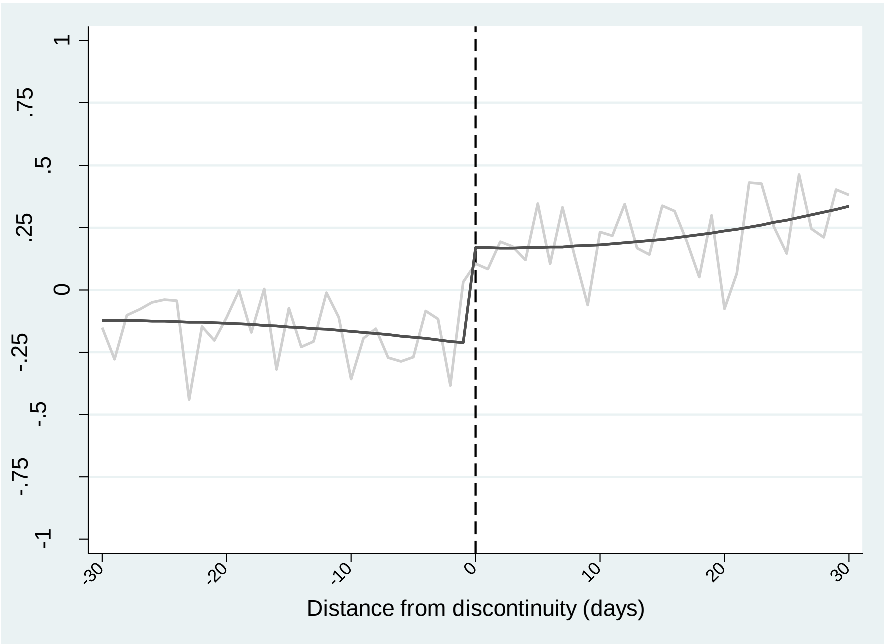
Figure 1 shows the discontinuity in national test scores when September born children are aged 7 years and 8 months and August born children are aged 6 years and 9 months, while Figure 2 shows the discontinuity when September borns are aged 11 years and 8 months and August borns are aged 10 years and 9months, both for those born up to 30 days either side of the 1st September cut off. It is clear that there is a large jump in test scores, particularly at age 7, when those marginally assigned to the treatment group (the oldest in the academic year) score, on average, around 0.7 standard deviations higher than the youngest in the academic year. The gap is somewhat smaller at age 11, at around 0.4 standard deviations, but it remains substantial.

Figure 1 Discontinuity in KS1 scores



Note: A window of 30 days either side of the discontinuity is applied. Both discontinuities (across cohorts) have been pooled to increase sample sizes. The model is as specified in Section 2, omitting background characteristics.

Figure 2 Discontinuity in KS2 scores



See notes to Figure 1.

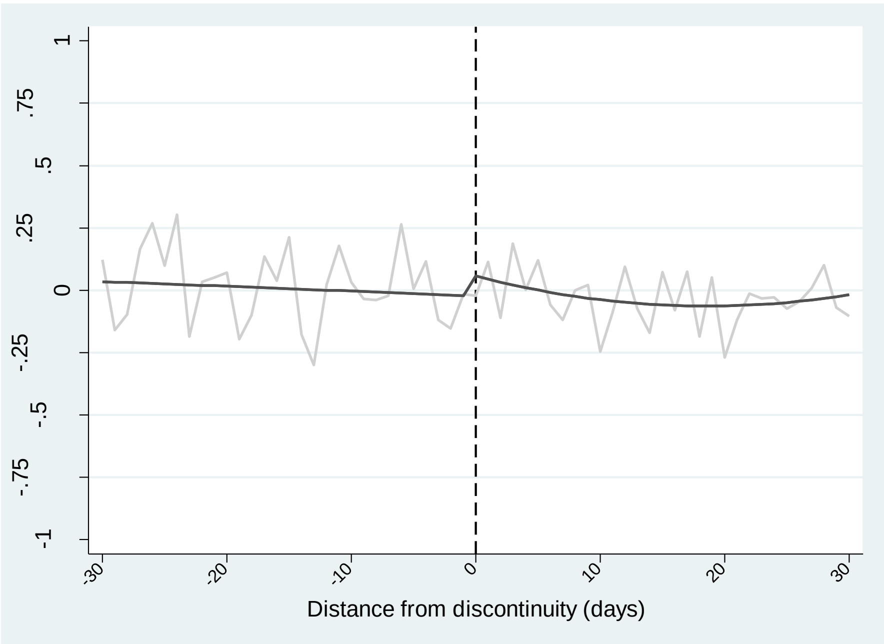
Table 3 presents the corresponding regression discontinuity design (RDD) estimation results for national achievement test scores at ages 7 and 11 in Columns 1 and 2. These models include a quadratic control for distance from the discontinuity and selected background characteristics (see Section 3 above). The estimated impact of being the oldest in the academic cohort (receiving the treatment) is around 0.7 standard deviations when pupils are around age 7, which confirms the graphical representation of the "treatment effect" around the discontinuity in Figure 1. The estimated impact of being the oldest in the academic cohort declines to just over 0.4 standard deviations when pupils are assessed at the end of primary school, at age 10 or 11. The declining importance of a young person's month of birth on their educational attainment is consistent with previous work using linear regression methods (e.g. Crawford et al., 2007, 2011). Indeed, the RDD method produces treatment effects that are not statistically different from those in Crawford et al. (2011) using the same sample of children, though they are slightly larger in magnitude.

Table 3 RD design estimates: cognitive skills

	National achievement test scores		IQ (WISC)
	Age 7	Age 11	Age 8
Treatment effect	0.707***	0.424***	0.061
	[0.160]	[0.105]	[0.090]
Distance	*Yes*	*Yes*	*Yes*
Distance X Treatment	*Yes*	*Yes*	*Yes*
Distance^2	*Yes*	*Yes*	*Yes*
Distance^2 X Treatment	*Yes*	*Yes*	*Yes*
Background characteristics	*Yes*	*Yes*	*Yes*
N	3,026	3,133	1,359
R-squared	0.258	0.232	0.201

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity. The model is as specified above and includes a range of background characteristics.

Figure 3 Discontinuity in IQ (WISC) standardised score



See notes to Figure 1.

Figure 3 and the third column of Table 3 move on to document differences in cognitive skills using an IQ (WISC) test measured around the time of the child's eighth birthday. In comparison to tests that are nationally administered, where children are assessed on the same day rather than at the same age, the "age at test" effect is eliminated. If we find differences in outcomes for children either side of the discontinuity when using this test, then it suggests that other factors, such as relative age, length of schooling and/or age of starting school have an impact. This finding would have important policy implications; simply age normalising nationally set and administered tests may not be sufficient to overcome observed differences in cognitive outcomes.

Figure 3 shows that there is only a very small jump in test scores at the time of the discontinuity, which column 3 of Table 3 suggests is not significantly different from zero. This is in marked contrast to the results from national achievement tests taken at age 7 and age 11, which can be affected by the age at which a child sits the test. This suggests that either:

a) The age at which a child sits a test is the most important driver of the difference in outcomes for children that are the oldest and youngest in their cohort;

b) Drivers of the differences in outcomes change significantly between age 7 and age 8;

c) WISC is not comparable to the Key Stage tests.

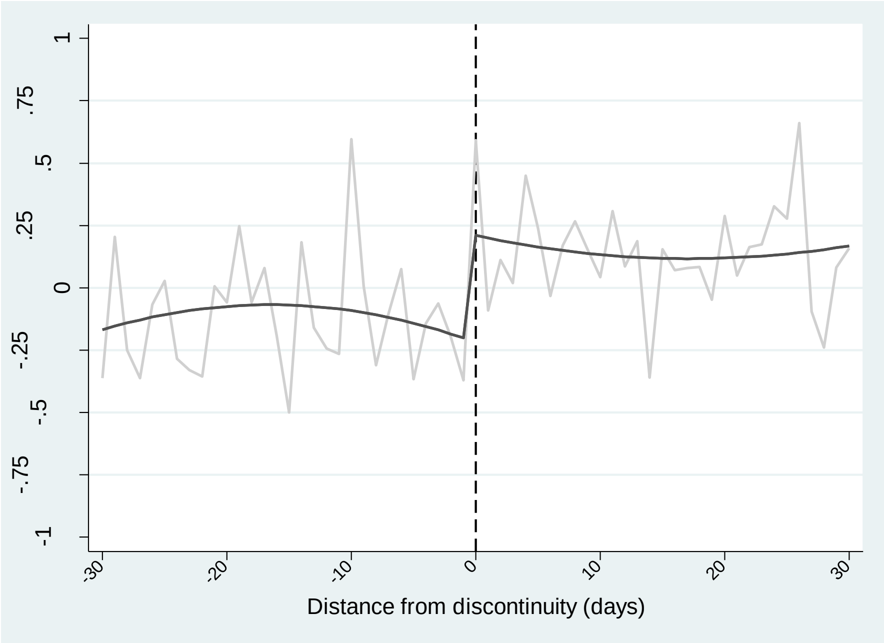
The most likely explanation is the first, namely that the significant driver of differences in educational attainment between those born in different months is the age at which a child sits a test. The second potential explanation is unlikely, as we observe significant differences between those born either side of the discontinuity at older ages. The third potential explanation is feasible, as WISC is designed to measure IQ rather than learned material, but it is unlikely that the differences in content are sufficient to remove the large significant difference observed at age 7. In particular, the evidence presented in section 2 suggests that these measures largely reflect similar measures of children's cognitive development.

If we are confident that the Key Stage and IQ (WISC) tests are relatively comparable, then this analysis suggests that it is the age at test effect that drives the differences in test scores between children born at the start and end of the academic year. This corroborates evidence found elsewhere (e.g. Crawford et al, 2007; Smith, 2010) and suggests that one simple solution to the penalty faced by those born at the end of the academic year when they are forced to sit exams on the same day as the older children in their cohort is to age normalise test scores.

Non-cognitive skills

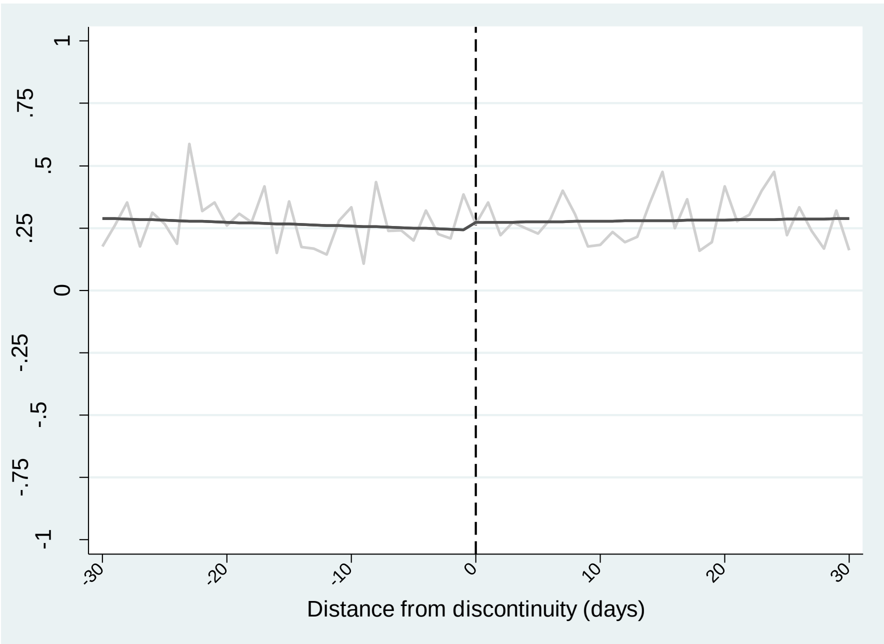
The results above have shown that there is little difference in cognitive test scores between children born at the start and end of the academic year when they are assessed at the same age, suggesting that a policy of age normalising test scores might be sufficient to overcome the disadvantages that summer-born children in England face. This section moves on to consider whether the same is also true for some wider measures of skills, including the child's own assessment of their scholastic competence (Figure 4) and global self-worth (Figure 7), whether they like school (Figure 5) and their locus of control (i.e. how in control they feel of their own destiny) (Figure 6). The RDD estimates and their significance are also show in Table 4.

Figure 4 Discontinuity in scholastic competence (reported by ALSPAC cohort member)



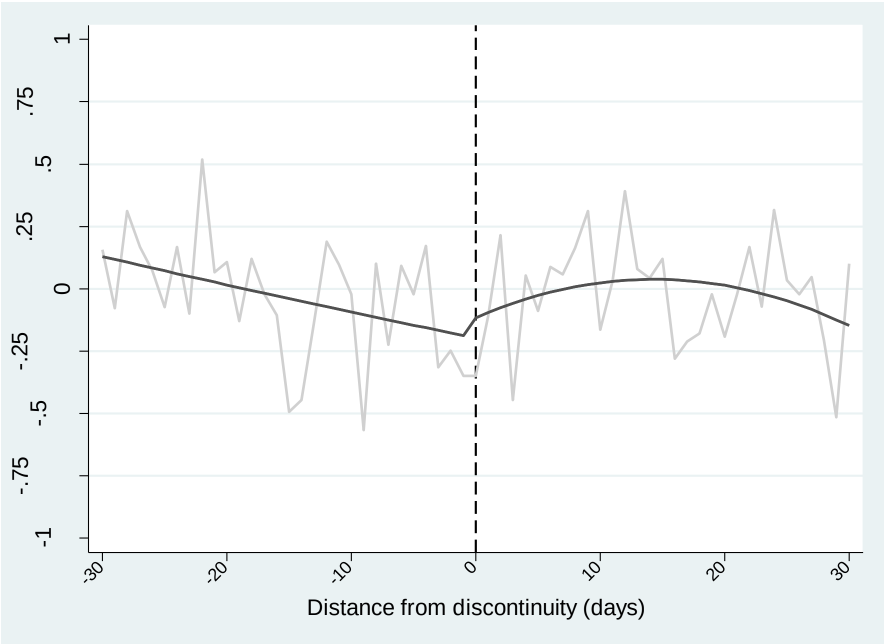
See notes to Figure 1.

Figure 5 Discontinuity in whether the child likes school very much



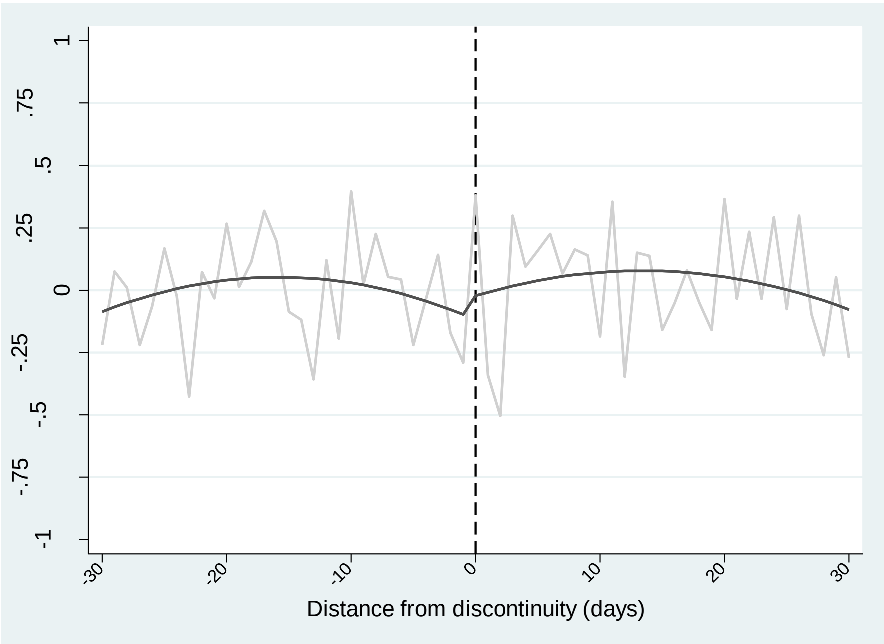
See notes to Figure 1.

Figure 6 Discontinuity in locus of control (reported by ALSPAC cohort member)



See notes to Figure 1.

Figure 7 Discontinuity in global self-worth (reported by ALSPAC cohort member)



See notes to Figure 1.

Figure 4 and column 1 of Table 4 show that there are sizeable and significant differences between August and September born children in terms of their view of their own scholastic competence at age 8, with a magnitude of around 0.4 standard deviations. This difference is the same as the effect size found for national achievement test scores at age 11 and is 40% larger than the difference in scholastic self-competence observed between children born to mothers with high and low levels of education, and twice as large as the difference observed between children born to households with the highest and lowest levels of income. By contrast, Figures 5 to 7 and columns 2 to 4 of Table 4 illustrate that there are no significant differences between children born either side of the discontinuity in terms of their global self-worth, their locus of control or whether they like school.

These results suggest that relative age, length of schooling and/or the age of starting school have a strong negative effect on those born younger in the academic year in terms of their own perceptions of their scholastic self-competence, which is not driven by the age at which they are asked the question. This suggests that a policy response of appropriately age-adjusting tests – as suggested above – may not completely resolve the issues associated with being later in the academic year, although the confidence of these younger children may respond positively to appropriately adjusted test scores.

Table 4 RD design estimates: non-cognitive skills

	Scholastic competence	Likes school very much	Locus of control	Self-esteem
Treatment effect	0.373*	-0.003	0.101	0.033
	[0.165]	[0.065]	[0.164]	[0.186]
Distance	*Yes*	*Yes*	*Yes*	*Yes*
Distance X Treatment	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2 X Treatment	*Yes*	*Yes*	*Yes*	*Yes*
Background characteristics	*Yes*	*Yes*	*Yes*	*Yes*
N	1,308	1,353	1,285	1,308
R-squared	0.084	0.063	0.124	0.064

See notes to Table 3.

5 Conclusions

Using a research design that improves on those previously used in this literature, this paper has confirmed that there are large and significant differences between August- and September-born children in terms of their cognitive skills measured using national achievement tests. In line with other literature (e.g. Crawford, Dearden & Meghir, 2007), the absolute magnitude of these differences decreases as children get older, suggesting that August-borns are 'catching up' with their September-born counterparts in a variety of ways as the difference in relative age becomes smaller over time.

What drives these differences in cognitive outcomes? All nationally set and administered tests are completed on the same day, which means that August born children are almost a year younger than their September born counterparts. This "age at test" effect has the potential to impact children's outcomes, but there are other potential mechanisms: the length of schooling the pupils receive prior to the test, the relative age of the pupil compared to their peers, and the age at which the pupil started school. Outcomes in nationally set and administered tests are potentially affected by all four of these mechanisms, which are difficult to distinguish.

To address this, we compare differences in cognitive outcomes that are affected by all four mechanisms to differences in another, similar, cognitive outcome where the age at test effect is eliminated. This allows us to observe whether the combined effects of relative age, length of schooling, and age of starting school are responsible for the difference observed between August and September born children. We conclude that these combined effects do not have a significant impact on cognitive development, and therefore that age at test is the most important factor driving the difference between the oldest and youngest children in an academic cohort. This finding confirms earlier work using a parametric approach (Crawford et al., 2007). An appropriate policy response to the difference in assessments taken in schools would therefore be to appropriately age-adjust nationally set and administered tests. But would this solution address potential differences in children's wider development?

We find no evidence that children's locus of control, self-assessed global self-worth or enjoyment of school are significantly different when these skills are measured around the child's eighth birthday. However, we find strong evidence that a child's view of their own scholastic competence at age 8 is significantly different between children born in August and September, even when eliminating the age at test effect. This means that other policy responses may be required to correct for differences in outcomes amongst children born in different months, but not necessarily so: it may be that children's view of their scholastic competence would change in response to the introduction of appropriately age-adjusted tests, for example as a result of positive reinforcement.

References

Bedard, K. and E. Dhuey (2006), The persistence of early childhood maturity: international evidence of long-run age effects, *Quarterly Journal of Economics*, **121**, 1437–72.

Bedard, K. and E. Dhuey (2009), School entry policies and skill accumulation across directly and indirectly affected men, University of Toronto, mimeo.

Black, S., Devereux, P. and Salvanes, K. (2008), Too young to leave the nest? The effects of school starting age, NBER Working Paper 13969.

Blundell, R. And M. Costa Dias (2009), Alternative Approaches to Evaluation in Empirical Microeconomics, *Journal of Human Resources*, **44**, 565-640.

Blundell, R., L. Dearden and B. Sianesi (2005), Evaluating the impact of education on earnings: models, methods and results from the NCDS, *Journal of the Royal Statistical Society, Series A*, **168**, 473–512.

Brewer, M. and C. Crawford (2010), Starting school and leaving welfare: the impact of public education on lone parents' welfare receipt, IFS Working Paper W10/19.

Buckles, K. and D. Hungerman (2008), Season of birth and later outcomes: old questions, new answers, NBER Working Paper 14573.

Carneiro, P., C. Crawford and A. Goodman (2007), The impact of early cognitive and non-cognitive skills on later outcomes, CEE Discussion Paper 92.

Cebi, M. (2007), Locus of control and human capital investment revisited, *Journal of Human Resources*, **42**, 919–32.

Chevalier, A. and G. Conlon (2003), Does it pay to attend a prestigious university?, IZA Discussion Paper 848.

Chowdry, H., C. Crawford and A. Goodman (2009), *Drivers and Barriers to Educational Success: Evidence from the Longitudinal Survey of Young People in England*, DCSF Research Report DCSF-RR102.

Crawford, C., L. Dearden and C. Meghir (2007), *When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England*, Centre for the Economics of Education Report to the Department for Children, Schools and Families (http://www.ifs.org.uk/publications/4073).

Crawford, C., Dearden, L. and Greaves, E. (2011), *Does when you are born matter? The impact of month of birth on children's cognitive and non-cognitive skills in England*, Institute for Fiscal Studies (IFS), Briefing Note no. 122 (http://www.ifs.org.uk/publications/5736).

Datar, A. (2006), Does delaying kindergarten entrance give children a head start?, *Economics of Education Review*, **25**, 43–62.

Dearden, L. (1999), The effects of families and ability on men's education and earnings in Britain, *Labour Economics*, **6**, 551–67.

Dearden, L., S. McIntosh, M. Myck and A. Vignoles (2002), The returns to academic and vocational qualifications in Britain, *Bulletin of Economic Research*, **54**, 249–74.

Department for Education (2010), *Month of Birth and Education*, DfE Research Report 17.

Dhuey, E. and S. Lipscomb (2008), What makes a leader? Relative age and high school leadership, *Economics of Education Review*, **27**, 173–83.

Dhuey, E. and S. Lipscomb (2010), Disabled or young? Relative age and special education diagnoses in schools, *Economics of Education Review*, **29**, 857–72.

Dickert-Conlin, S. and T. Elder (2010), Suburban legend: school cutoff dates and the timing of births, *Economics of Education Review*, **29**, 826–41.

Dobkin, C. and F. Ferreira (2010), Do school entry laws affect educational attainment and labor market outcomes?, *Economics of Education Review*, **29**, 40–54.

Duckworth, A., C. Peterson, M. Matthews and D. Kelly (2007), Grit: perseverance and passion for long-term goals, *Journal of Personality & Social Psychology*, **92**, 1087–101.

Duncan, R. and G. Dunifon (1998), Long-run effects of motivation on labor market success, *Social Psychology Quarterly*, **61**, 33–48.

Elder, T. and D. Lubotsky (2009), Kindergarten entrance age and children's achievement: impacts of state policies, family background and peers, *Journal of Human Resources*, **44**, 641–83.

Feinstein, L. (2002a), *Quantitative Estimates of the Social Benefits of Learning, 1: Crime*, Wider Benefits of Learning Research Report 5, Institute of Education, London.

Feinstein, L. (2002b), *Quantitative Estimates of the Social Benefits of Learning, 2: Health (Depression and Obesity)*, Wider Benefits of Learning Research Report 6, Institute of Education, London.

Fredriksson, P. and B. Ockert (2005), Is early learning really more productive? The effect of school starting age on school and labour market performance, IZA Discussion Paper 1659.

Frolich, M. (2007), Regression discontinuity design with covariates, cemmap working paper CWP27/07.

Gans, J. and A. Leigh (2009), Born on the first of July: an (un)natural experiment in birth timing, *Journal of Public Economics*, **93**, 246–63.

Goodman, A. and P. Gregg (2010), *Poorer Children's Educational Attainment: How Important Are Attitudes and Behaviours?*, Report to the Joseph Rowntree Foundation.

Goodman, R., J. Gledhill and T. Ford (2003), Child psychiatric disorder and relative age within school year: cross sectional survey of large population sample, *British Medical Journal*, **327**, 1–4.

Hahn, J., Todd, P., and W. Van der Klaauw (2001), Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design, *Econometrica*, **69**, 201-209.

Hammond, C. and L. Feinstein (2006), *Are Those who Flourished at School Healthier Adults? What Role for Adult Education?*, Wider Benefits of Learning Research Report 17, Institute of Education, London.

Harter, S. (1985), Manual for the self-perception profile for children, University of Denver, mimeo.

HEFCE (2005), "Young participation in higher education", HEFCE Report No. 2005/03.

Kawaguchi, D. (2011), Actual age at school entry, educational outcomes, and earnings, *Journal of the Japanese and International Economies*, **25**(2), 64–80.

Lee, D. and Lemieux, T. (2010) "Regression Discontinuity Designs in Economics", Journal of Economic Literature, 48(2), pp. 281-355.

Mühlenweg, A. (2010), Young and innocent: international evidence on age effects within grades on victimization in elementary school, *Economics Letters*, **109**, 157–60.

Mühlenweg, A., D. Blomeyer and M. Laucht (2011), Effects of age at school entry on the development of non-cognitive skills: evidence from psychometric data, ZEW Discussion Paper 11-017.

Neale, M. (1997), *Neale Analysis of Reading Ability – Revised*, NFER-Nelson, Windsor.

Nowicki, S. and B. Strickland (1973), A locus of control scale for children, *Journal of Consulting and Clinical Psychology*, **40**, 148–54.

Osborne Groves, M. (2005), How important is your personality? Labor market returns to personality for women in the US and UK, *Journal of Economic Psychology*, **26**, 827–41.

Puhani, P. and A. Weber (2007), Does the early bird catch the worm? Instrumental variable estimates of early educational effects of age of school entry in Germany, *Empirical Economics*, **32**, 359–86.

Rust, J. (1996), *The Manual of the Wechsler Objective Language Dimensions*, The Psychological Corporation.

Sharp, C. (1995), What's age got to do with it? A study of patterns of school entry and the impact of season of birth on school attainment, *Education Research*, **37**, 251–65.

Smith, J. (2010), 'How valuable is the gift of time? The factors that drive the birth date effect in education', *Education Finance and Policy*, vol. 5, pp. 247–77.

Solli, I. (2011), Left behind by birth month, University of Stavanger, mimeo.

Wechsler, D., S. Golombok and J. Rust (1992), *Wechsler Intelligence Scale for Children – Third Edition*, UK Manual, The Psychological Corporation, Sidcup, UK.

Appendix A: Continuity of background variables around the discontinuity

Table A1 RD design estimates: background characteristics

	Highest income quintile	Mother married at birth	Mother has a degree	No older siblings	Father in highest social class
Treatment effect	0.018	-0.016	-0.018	0.030	0.004
	[0.032]	[0.044]	[0.027]	[0.047]	[0.020]
Distance	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance X Treatment	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2 X Treatment	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Background characteristics	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
N	3,721	3,721	3,721	3,721	3,721
R-squared	0.002	0.001	0.001	0.002	0.002

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity.

Table A2 RD design estimates: age (in days) at survey

	Focus at 8 clinic
Treatment effect	3.320
	[2.944]
Distance	*Yes*
Distance X Treatment	*Yes*
Distance^2	*Yes*
Distance^2 X Treatment	*Yes*
Background characteristics	*Yes*
N	1,409
R-squared	0.004

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity, measured in days.

Appendix B: Alternative specifications

Table B1: RD design estimates: educational attainment in standardised national tests with different sized windows around discontinuity

	Window of 20 days			Window of 40 days		
	Age 7	Age 11	Age 16	Age 7	Age 11	Age 16
Treatment effect	0.568**	0.289*	0.088	0.664***	0.389***	0.191*
	[0.196]	[0.139]	[0.089]	[0.139]	[0.085]	[0.077]
Distance	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance X Treatment	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2 X Treatment	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Background characteristics	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
N	2,049	2,113	2,044	3,981	4,141	4,049
R-squared	0.260	0.234	0.259	0.253	0.229	0.267

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity.

Table B2: RD design estimates: educational attainment in standardised national tests with different polynomial specifications

	Polynomial of degree 1			Polynomial of degree 3		
	Age 7	Age 11	Age 16	Age 7	Age 11	Age 16
Treatment effect	0.682***	0.386***	0.170**	0.488*	0.258	0.045
	[0.096]	[0.060]	[0.061]	[0.208]	[0.152]	[0.098]
Distance	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance X Treatment	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Distance^2 X Treatment	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Distance^3	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Distance^3 X Treatment	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Background characteristics	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
N	3,026	3,133,	3,048,	3,026	3,133	3,048
R-squared	0.257	0.232	0.263	0.259	0.233	0.264

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity.

Table B3: RD design estimates: SDQ reported by the parent with different sized windows around discontinuity

	Window of 20 days			Window of 40 days		
	WISC Age 8	Scholastic comp. Age 8	Likes school very much Age 8	WISC Age 8	Scholastic comp. Age 8	Likes school very much Age 8
Treatment effect	0.108	0.313	-0.023	0.071	0.296*	0.035
	[0.115]	[0.220]	[0.073]	[0.076]	[0.135]	[0.051]
Distance	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance X Treatment	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2 X Treatment	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Background characteristics	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
N	912	878	912	1,803	1,738	1,800
R-squared	0.249	0.113	0.074	0.187	0.061	0.053

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity.

Table B4: RD design estimates: SDQ reported by the parent with different polynomial specifications

	Polynomial of degree 1			Polynomial of degree 3		
	WISC Age 8	Scholastic comp. Age 8	Likes school very much Age 8	WISC Age 8	Scholastic comp. Age 8	Likes school very much Age 8
Treatment effect	0.042	0.284**	0.033	0.085	0.495*	-0.040
	[0.064]	[0.106]	[0.043]	[0.115]	[0.240]	[0.079]
Distance	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance X Treatment	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance^2	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Distance^2 X Treatment	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Distance^3	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Distance^3 X Treatment	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Background characteristics	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
N	1,359	1,308	1,353	1,359	1,308	1,353
R-squared	0.200	0.083	0.063	0.201	0.085	0.070

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity.

Table B5: RD design estimates: non-cognitive outcomes with different sized windows around discontinuity

	Window of 20 days		Window of 40 days	
	Locus of control	Self-worth	Locus of control	Self-worth
Treatment effect	-0.056	0.017	0.211	0.099
	[0.197]	[0.234]	[0.147]	[0.160]
Distance	*Yes*	*Yes*	*Yes*	*Yes*
Distance X Treatment	*Yes*	*Yes*	*Yes*	*Yes*
$Distance^2$	*Yes*	*Yes*	*Yes*	*Yes*
$Distance^2$ X Treatment	*Yes*	*Yes*	*Yes*	*Yes*
Background characteristics	*Yes*	*Yes*	*Yes*	*Yes*
N	866	879	1,702	1,738
R-squared	0.152	0.078	0.109	0.047

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity.

Table B5: RD design estimates: non-cognitive outcomes with different polynomial specifications

	Polynomial of degree 1		Polynomial of degree 3	
	Locus of control	Self-worth	Locus of control	Self-worth
Treatment effect	0.236*	0.060	0.012	0.128
	[0.116]	[0.121]	[0.198]	[0.258]
Distance	*Yes*	*Yes*	*Yes*	*Yes*
Distance X Treatment	*Yes*	*Yes*	*Yes*	*Yes*
$Distance^2$	*No*	*No*	*Yes*	*Yes*
$Distance^2$ X Treatment	*No*	*No*	*Yes*	*Yes*
$Distance^3$	*No*	*No*	*Yes*	*Yes*
$Distance^3$ X Treatment	*No*	*No*	*Yes*	*Yes*
Background characteristics	*Yes*	*Yes*	*Yes*	*Yes*
N	1,285	1,308	1,285	1,308
R-squared	0.123	0.063	0.124	0.065

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity.

Appendix C: Placebo tests

Table C1: RD design estimates: educational attainment at age 7; Placebo tests with discontinuity at 1st of each month.

	Treatment Effect	Window of 30 days Standard error	Number of obs.
1st September	0.707***	[0.160]	3,026
1st October	-0.052	[0.108]	2,878
1st November	-0.160	[0.082]	2,774
1st December	0.161	[0.104]	2,750
1st January	-0.063	[0.113]	2,211
1st February	-0.028	[0.189]	1,565
1st March	-0.093	[0.152]	1,487
1st April	-0.090	[0.137]	2,307
1st May	-0.106	[0.076]	3,065
1st June	-0.041	[0.098]	3,071
1st July	0.102	[0.114]	3,125
1st August	-0.189	[0.128]	3,119

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity.

Table C2: RD design estimates: WISC at age 8; Placebo tests with discontinuity at 1st of each month.

	Treatment Effect	Window of 30 days Standard error	Number of obs.
1st September	0.061	[0.090]	1359.000
1st October	-0.094	[0.068]	1376.000
1st November	0.068	[0.082]	1335.000
1st December	0.092	[0.088]	1284.000
1st January	-0.107	[0.100]	1047.000
1st February	0.053	[0.085]	744.000
1st March	-0.071	[0.138]	750.000
1st April	-0.042	[0.127]	1065.000
1st May	-0.004	[0.093]	1312.000
1st June	-0.175*	[0.071]	1370.000
1st July	-0.033	[0.076]	1432.000
1st August	0.062	[0.080]	1388.000

Note: Standard errors are robust and clustered by distance to the discontinuity. Statistical significance of the treatment effect is denoted as follows: * p<0.05, ** p<0.01, *** p<0.001. "Distance" refers to the assignment variable, the distance from the discontinuity.

Appendix D: Descriptive Statistics

Table D1 Average outcomes for the sample as a whole and for those born in September

Sample: window of 10 days around discontinuity

Variable	*Left of discontinuity*		*Right of discontinuity*		*Average in window*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Mean KS1 APS	-0.35	1.10	0.15	1.06	-0.09	1.11
Mean KS2 APS	-0.12	0.94	0.19	0.90	0.04	0.93
Mean KS4 capped	-0.03	0.90	0.04	0.97	0.01	0.94
Mean IQ (WISC)	0.00	0.57	-0.01	0.56	0.00	0.56
Mean scholastic competence	-0.1	1.03	0.17	0.92	0.03	0.99
Mean "likes school very much"	0.26	0.44	0.27	0.44	0.26	0.44
Mean locus of control	-0.15	1.01	-0.02	0.99	-0.09	1.00
Mean self-worth	0.02	1.07	0.04	0.96	0.03	1.01

Sample: window of 20 days around discontinuity

Variable	*Left of discontinuity*		*Right of discontinuity*		*Average in window*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Mean KS1 APS	-0.38	1.13	0.19	1.07	-0.09	1.13
Mean KS2 APS	-0.15	0.95	0.19	0.92	0.03	0.95
Mean KS4 capped	-0.08	0.93	0.05	0.95	-0.01	0.94
Mean IQ (WISC)	-0.01	0.56	-0.03	0.55	-0.02	0.56
Mean scholastic competence	-0.09	1.04	0.13	0.95	0.02	1.00
Mean "likes school very much"	0.25	0.44	0.28	0.45	0.26	0.44
Mean locus of control	-0.10	1.02	-0.02	0.98	-0.06	1.00
Mean self-worth	0.02	1.05	0.04	0.98	0.03	1.01

Sample: window of 30 days around discontinuity

Variable	*Left of discontinuity*		*Right of discontinuity*		*Average in window*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Mean KS1 APS	-0.39	1.12	0.2	1.07	-0.10	1.14
Mean KS2 APS	-0.16	0.95	0.22	0.91	0.03	0.95
Mean KS4 capped	-0.09	0.93	0.07	0.95	-0.01	0.94
Mean IQ (WISC)	0.01	0.56	-0.04	0.54	-0.02	0.55
Mean scholastic competence	-0.11	1.04	0.14	0.96	0.02	1.01
Mean "likes school very much"	0.27	0.44	0.28	0.45	0.27	0.45
Mean locus of control	-0.04	1.02	-0.02	0.97	-0.03	0.99
Mean self-worth	0.00	1.06	0.03	0.99	0.02	1.02

Sample: window of 40 days around discontinuity

Variable	*Left of discontinuity*		*Right of discontinuity*		*Average in window*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Mean KS1 APS	-0.35	1.12	0.19	1.10	-0.09	1.14
Mean KS2 APS	-0.13	0.94	0.20	0.93	0.03	0.95
Mean KS4 capped	-0.07	0.93	0.07	0.96	0.00	0.95
Mean IQ (WISC)	0.01	0.55	-0.04	0.54	-0.02	0.55
Mean scholastic competence	-0.10	1.00	0.14	0.95	0.03	0.98
Mean "likes school very much"	0.27	0.44	0.28	0.45	0.28	0.45
Mean locus of control	-0.03	1.01	0.00	0.98	-0.01	0.99
Mean self-worth	-0.02	1.03	0.05	0.99	0.01	1.01

Sample: no window imposed around discontinuity

Variable	*Left of discontinuity*		*Right of discontinuity*		*Average in window*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Mean KS1 APS	-0.23	1.14	0.12	1.11	-0.07	1.14
Mean KS2 APS	-0.06	0.97	0.16	0.95	0.04	0.97
Mean KS4 capped	-0.03	0.93	0.04	0.96	0.00	0.95
Mean IQ (WISC)	0.03	0.56	-0.04	0.55	0.00	0.56
Mean scholastic competence	-0.09	1.00	0.09	0.99	0.00	1.00
Mean "likes school very much"	0.28	0.45	0.28	0.45	0.28	0.45
Mean locus of control	0.00	1.00	0.00	1.00	0.00	1.00
Mean self-worth	-0.01	1.01	0.02	0.99	0.00	1.00