Online Appendix to
"Early School Exposure, Test Scores, and Noncognitive Outcomes"

Thomas Cornelissen and Christian Dustmann *

* Cornelissen: Department of Economics, University of York, Heslington, York YO10 5DD, and Centre for Research and Analysis of Migration (CReAM), 30 Gordon Street, London WC1H 0AX, United Kingdom (thomas.cornelissen@york.ac.uk); Dustmann: Department of Economics, University College London and CReAM, 30 Gordon Street, London WC1H 0AX, United Kingdom (c.dustmann@ucl.ac.uk).

APPENDIX A – Additional Tables and Figures

Table A1: Counts of compliers, always-takers, never-takers

Treatment intensity	Change of EEXP from 4-7 (EEXP7)				Change of EEXP from 7-11 (EEXP11)			
	Compliers	Always-Takers	Never-takers	IV weight	Compliers	Always-Takers	Never-takers	IV weight
4	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
5	0.21	0.78	0.00	0.06	0.01	0.98	0.02	0.00
6	0.22	0.78	0.00	0.06	0.01	0.98	0.02	0.00
7	0.21	0.78	0.01	0.06	0.01	0.97	0.02	0.01
8	0.03	0.84	0.13	0.01	0.37	0.57	0.05	0.20
9	0.03	0.84	0.13	0.01	0.37	0.57	0.05	0.20
10	0.03	0.84	0.14	0.01	0.37	0.57	0.06	0.20
11	0.03	0.83	0.14	0.01	0.37	0.57	0.06	0.20
Sum	**0.76**			**0.20**	**1.53**			**0.80**
Weighted average	**0.04**	**0.16**	**0.005**		**0.29**	**0.46**	**0.046**	

Notes: The fraction of compliers is estimated as follows: For each possible value j of actual exposure we construct a dummy variable d_{ji} indicating whether observation i has actual exposure of greater or equal to j or not. We then regress each d_{ji} on the two dummy instruments EEXP7 and EEXP11 and all exogenous regressors. The coefficient on each instrument then gives us the fraction of compliers with this instrument at treatment intensity j. The fraction of always-takes associated with instrument Z at treatment intensity j is estimated by the average predicted value of that same regression with instrument Z switched off, and the fraction of never-takers is one minus the average prediction of that regression with the instrument Z switched on. The IV weights are calculated as shown in equation A2 of online Appendix C. They sum to one over all treatment intensities and over both instruments. The IV weights are also used to construct the weighted averages reported in the last row of the table.

Table A2: Early exposure effects on the FSP Total score at the end of first grade (age 5) – RDD design

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	First stage (FS)		**Reduced form (RF)**				**RDD Exposure Effect (RF/FS)**
Policy area:	B	C	A		B	C	
Born from March 2001 onwards	-1.720*** (0.374)				-0.137** (0.058)		0.080** (0.038)
Born from January 2001 onwards		-1.503*** (0.366)				-0.141** (0.056)	0.094** (0.044)
Born from May 2001 Onwards		-1.022*** (0.498)				-0.074* (0.044)	0.072 (0.056)
Age (centred around mean)	0.021 (0.025)	0.056 (0.078)	0.067*** (0.002)	0.067*** (0.002)	0.066*** (0.007)	0.063 (0.182)	
Age Squared	-0.002 (0.004)	0.003 (0.007)		-0.001** (0.001)	-0.001 (0.001)	0.000 (0.001)	
R-squared	0.44	0.51	0.20	0.20	0.17	0.17	
N	1,541	1,206	26,716	26,716	7,802	5,908	

Note: Columns (1) and (2) show first-stage (FS) estimates from a fuzzy RDD design regressing actual exposure to reception class onto dummies for whether an individual is born after one of the birth-month cut-offs of the respective policy areas, controlling for a quadratic in age. Columns (3)-(6) show respective reduced-form (RF) regressions in which the dependent variable is the FSP total score at the end of first grade (age 5). Columns (3) and (4) are for policy area A in which there is no birth month cut-off, and thus only an age effect can be identified. Columns (5) and (6) are for policy areas B and C, in which there are cut-offs that allow identification of an exposure effect. Column (7) reports the exposure effect implied by the RDD design by dividing the RF coefficients on the different cut-offs by the FS coefficients for the same cut-offs. Effects in column (7) are effects of a one-month increase in exposure to the first school year on the total test score from the in-school assessment at the end of the first grade at age 5. All regressions include the following control variables: dummies for gender, free school meal eligibility, English first language at home, ethnicity, and local authority. Standard errors clustered at the level of the local authority are in parentheses. Standard errors in column (7) are computed as $Var\left(\frac{\hat{\beta}_{RF}}{\hat{\beta}_{FS}}\right) \approx \sqrt{\left[\hat{\beta}_{RF}^2 Var\left(\hat{\beta}_{RF}\right) + \hat{\beta}_{FS}^2 Var\left(\hat{\beta}_{RF}\right)\right]/\hat{\beta}_{RF}^4}$. Statistically significant at the *** 0.01 level, ** 0.05 level.

Data Source: MCS (Millennium Cohort Study) for first stage, NPD (National Pupil Database) for reduced form.

Table A3: Early exposure effects on cognitive Test Scores at ages 5, 7 and 11 (MCS)

Panel A: Cognitive test scores at age 5		
	1st grade Language Skills	1st grade Numeracy Skills
Exposure	0.088***	0.085***
	(0.024)	(0.026)
N	7,805	7,805
Exposure * male	0.078**	0.082**
	(0.034)	(0.039)
Exposure * female	0.097***	0.082***
	(0.028)	(0.028)
N	7,805	7,805
Panel B: Cognitive test scores at age 7		
	3rd grade Language Skills	3rd grade Numeracy Skills
Exposure	0.016	0.005
	(0.027)	(0.027)
N	5,782	5,780
Exposure * male	0.034	0.033
	(0.034)	(0.035)
Exposure * female	-0.004	-0.018
	(0.037)	(0.034)
N	5,782	5,780
Panel C: Cognitive test scores at age 11		
	7th grade Language Skills	7th grade Numeracy Skills
Exposure	-0.025	0.016
	(0.018)	(0.036)
N	4,715	4,668
Exposure * male	-0.002	0.026
	(0.026)	(0.055)
Exposure * female	-0.040	0.027
	(0.027)	(0.044)
N	4,715	4,668

Note: The table shows IV estimates for the effect of a one-month increase in exposure to the first school year on cognitive assessments at ages 5, 7 and 11. Control variables: dummies for gender, free school meal eligibility, English first language at home, ethnicity, birth month, and local authority. Separate results for males and females are obtained by interacting all regressors with male/female dummies. Standard errors clustered at the level of the local authority are in parentheses. Statistically significant at the *** 0.01 level, ** 0.05 level, * 0.10 level.
Data Source: Millenium Cohort Study.

Table A4: Family characteristics by gender of the child and SES

	Male			Female			
	high SES	low SES	Diff.	high SES	low SES	Diff.	Diff-in-Diff
Formal child care before school entry							
No formal child care	0.13	0.15	0.020*	0.13	0.15	0.0	0.007
Hours of formal child care	16.78	14.77	-2.011***	16.57	15.13	-1.441***	-0.456
Socio-economic characteristics							
OECD poverty indicator	0.26	0.56	0.301***	0.27	0.57	0.299***	0.007
Homeownership	0.71	0.44	-0.270***	0.71	0.42	-0.289***	0.016
Social housing	0.19	0.44	0.257***	0.19	0.44	0.255***	0.001
Mother's education/work							
Mother low education	0.41	0.62	0.208***	0.40	0.61	0.204***	0.008
Lone parent	0.17	0.28	0.104***	0.15	0.29	0.131***	-0.026
Mother works throughout survey waves	0.36	0.18	-0.176***	0.37	0.15	-0.215***	0.032
Parenting style							
Not read to child daily, age 3	0.40	0.54	0.135***	0.38	0.54	0.156***	-0.016
Helped with reading, at most once week, age5	0.11	0.16	0.045***	0.10	0.13	0.026**	0.012
Exercise less than 1 day per week, age 5	0.51	0.69	0.184***	0.43	0.67	0.240***	-0.056**
More than 3 hrs TV per week day, age 5	0.15	0.21	0.060***	0.14	0.20	0.059***	-0.012
More than 1hr comp. game per week day, age 5	0.27	0.35	0.073***	0.18	0.21	0.032**	0.026
No regular bed time on week days, age 5	0.09	0.15	0.055***	0.08	0.15	0.072***	-0.019
Problem with sleeping habit, age 5	0.10	0.11	0.010	0.09	0.14	0.053***	-0.045***
Does usually not eat at regular times, age 5	0.07	0.12	0.055***	0.06	0.12	0.061***	-0.012
How often read, age 5	5.22	5.00	-0.225***	5.24	5.01	-0.230***	0.017
How often tell story, age 5	3.47	3.41	-0.062	3.52	3.50	-0.015	-0.034
How often musical activity, age 5	4.60	4.51	-0.091*	4.84	4.82	-0.011	-0.054
How often paint/draw, age 5	3.72	3.73	0.0165	3.93	3.79	-0.146***	0.136**
How often play games/toys, age 5	4.48	4.40	-0.084**	4.43	4.36	-0.063	-0.008
How often visit library, age 5	2.64	2.45	-0.187***	2.75	2.44	-0.316***	0.104
# obs.	1,200	3,200		1,162	3,077		

Notes: The table shows mean values of family characteristics by gender of the child and high and low socio-economic status (SES). See section III.D in the paper for the SES definition. Columns labelled "Diff" show the difference of the low-SES minus the high-SES outcome. The column labelled "Diff-in-Diff" shows the difference of the low-SES minus the high-SES outcome for boys, minus the same difference for girls. Most outcomes are dummy variables, except the hours of formal child care, and the last six outcomes, which are coded as 6 ("every day"), 5 ("several times a week"), 4 ("once or twice a week"), 3 ("once or twice a month"), 2 ("less often") and 1 ("not at all"). Statistically significant at the *** 0.01 level, ** 0.05 level, * 0.10 level. Data Source: Millennium Cohort Study.

Table A5: Early exposure effects on outcomes at ages 5, 7 and 11, by gender and SES

Panel A: Age 5

	Cognitive outcomes, 1st grade		***Noncognitive outcomes***		
Male	Language	Numeracy	Creative Development	Physical Development	Personal Social and Emotional Development
Exposure * High SES	0.036	0.046	0.013	0.033	0.026
	(0.035)	(0.037)	(0.036)	(0.031)	(0.034)
Exposure * Low SES	0.174***	0.167***	0.090*	0.122**	0.113**
	(0.042)	(0.054)	(0.051)	(0.059)	(0.048)
Female					
Exposure * High SES	0.099***	0.079***	0.085***	0.062**	0.066**
	(0.029)	(0.030)	(0.033)	(0.029)	(0.030)
Exposure * Low SES	0.099***	0.091***	0.062	0.055	0.055
	(0.032)	(0.034)	(0.041)	(0.039)	(0.038)
N	7768	7769	7768	7768	7768

Panel B: Age 7

	Cognitive outcomes, 3rd grade		***Noncognitive outcomes***			
Male	Language	Numeracy	Teacher relationship I	Academic Interest I	Positive self perception I	Disruptive behaviour I
Exposure * High SES	-0.013	-0.013	0.112***	0.091	0.016	-0.022
	(0.031)	(0.034)	(0.041)	(0.064)	(0.067)	(0.046)
Exposure * Low SES	0.102**	0.107**	0.043	0.163**	0.069	0.032
	(0.044)	(0.046)	(0.059)	(0.075)	(0.070)	(0.053)
Female						
Exposure * High SES	-0.011	-0.012	0.044	0.011	0.045	-0.067
	(0.038)	(0.036)	(0.039)	(0.031)	(0.052)	(0.047)
Exposure * Low SES	0.016	-0.040	-0.005	-0.045	0.031	-0.091**
	(0.042)	(0.035)	(0.049)	(0.042)	(0.047)	(0.043)
N	5763	5761	6131	6242	6316	6367

- **Table A5 to be continued -**

Table A5 – continued

Panel C: Age 11						
	Cognitive outcomes, 7th grade		***Noncognitive outcomes***			
	Language	Numeracy	Teacher relationship II	Academic Interest II	Positive self perception II	Disruptive behaviour II
Male						
Exposure * High SES	-0.014	0.005	0.089	0.132***	0.034	-0.092**
	(0.029)	(0.056)	(0.060)	(0.045)	(0.043)	(0.046)
Exposure * Low SES	-0.007	0.033	0.173**	0.068	0.032	-0.093
	(0.026)	(0.056)	(0.081)	(0.066)	(0.052)	(0.062)
Female						
Exposure * High SES	-0.050*	0.020	0.020	-0.045	-0.020	-0.072**
	(0.029)	(0.044)	(0.051)	(0.043)	(0.040)	(0.036)
Exposure * Low SES	-0.001	0.061	-0.033	-0.062	-0.002	-0.036
	(0.028)	(0.049)	(0.061)	(0.056)	(0.049)	(0.030)
N	4700	4653	5298	5835	6223	6218

Note: The table shows IV estimates for the effect of a one-month increase in exposure to the first school year on cognitive and non-cognitive and behavioural outcomes at ages 5, 7 and 11 by gender and socio-economic status (SES). SES is defined depending on parental occupation according to the National Statistics Socio-Economic Classification devised by the UK Office for National Statistics and provided in the MCS dataset (see section III.D of the text for more detailed information). Low SES is a dummy variable indicating that a family belongs to the bottom quartile of this SES measure, high SES a dummy for the first three top quartiles. See Appendix F for a description of the dependent variables used in this table. Control variables: dummies for gender, free school meal eligibility, English first language at home, ethnicity, birth month, and local authority. Separate results for males and females are obtained by interacting all regressors with male/female dummies.
Standard errors clustered at the level of the local authority are in parentheses.
Statistically significant at the *** 0.01 level, ** 0.05 level, * 0.10 level.
Data Source: MCS (Millennium Cohort Study).

APPENDIX B – Regression Discontinuity Design

In order to relax the difference-in-differences common-trend assumption, we exploit the fact that the exposure effect can also be identified based on a regression discontinuity (RD) design within policy areas that have a birth-month cut-off. We start by describing the RD estimator based on data from policy area B alone, in which children born before March 2001 are supposed to enter in the first term, and children born from March 2001 onwards are supposed to enter in the second term of the academic year. The fuzzy RD design has the first stage regression:

$$EXP_{imr} = \theta_0 + \theta_1 DAfter_{imr} + f(AGET_{imr}) + \mathrm{Z}_{imr}'\delta + \varepsilon_{imr}$$

where $DAfter_{imr}$ is a dummy variable equal to 1 for birth months after the cut-off, and equal to zero otherwise, and $f(AGET_{imr})$ is a continuous function (such as a polynomial) in $AGET_{imr}$. Observed characteristics Z_{imr} can be included to increase precision. The coefficient θ_1 picks up by how much exposure to schooling decreases at the cutoff. The corresponding reduced-form equation for an outcome of interest y_{imr} is

$$y_{imr} = \pi_0 + \pi_1 DAfter_{imr} + f(AGET_{imr}) + \mathrm{Z}_{imr}'\tau + \xi_{imr},$$

and the RD estimator of the exposure effect is $\hat{\gamma}_1^{RD} = \frac{\widehat{\pi_1}}{\widehat{\theta_1}}$.

In policy area C, there are two birth-month cutoffs and three possible school-entry dates. Children born before January 2006 are supposed to enter school in the first term, children born from January to March 2006 are supposed to enter in the second term, and children born from April 2006 are supposed to enter in the third term. Using data from policy area C alone allows estimating two RD effects by extending equations (3) and (4) with an additional dummy variable for a second cut-off. The assumptions for the RD estimators to be valid are that the probability of being born just before or just after the cutoff is not endogenous (e.g., parents do not manipulate the birth month of their child), and that the function of the running variable age $f(AGET_{imr})$ is correctly specified. If this precondition is met, by conducting the analysis just within policy region B or C, the RD design controls implicitly for *policy-region specific* age and birth month effects, thus allowing for potential *interactions* between birth month and policy region in the error terms of the outcome and first–stage equations (1) and

(2) in the paper.[1] Thus, the RD design relaxes the parallel trends assumption of the difference-in-differences design described in section II.A of the paper but instead requires the assumption that the continuous function of the running variable age $f(AGET_{imr})$ is correctly specified. In our setting, we need to rely on this assumption quite heavily because we are not able to go very close to the cut-off for the reasons of having a discrete running variable, and because in the survey data the sample size is limited, in particular at later ages. We therefore chose the difference-in-differences IV estimator as the preferred approach in the paper because it exploits all cut-offs in all policy areas to obtain an overall estimate, which is more efficient than separate RD designs by policy area. It is also attractive because it allows controlling completely flexibly (by monthly dummies) for birth month and age at test.

In Table A2 we show the results from the fuzzy RDD design. The first-stage regressions are based on the MCS data, while the reduced form is based on the NPD data. The first two columns of Table A2 show the first stage, in which we regress actual exposure to reception class (in months) on the cut-offs of policy areas B, in column (1), and C, in column (2), controlling for a second-order polynomial in age, and for control variables. In policy area B, the results indicate a reduction of exposure to schooling by about 1.7 months at the cut-off. This number is consistent with the drop observed for policy area B in Figure 3 in the paper. In policy area C, the reduction in exposure at the two cut-offs is slightly smaller, again reflecting the visual impression for this policy area in Figure 3.

Columns (3) to (6) of Table A2 show reduced-form estimates, for the outcome of the standardized total score from the teacher assessment at the end of the first school year. In columns (3) and (4) we first show regressions for policy area A only, in which there exists no cut-off. Policy area A alone thus only identifies the classical age-at entry effect. Whether we specify age linearly (column 3), or quadratic (column 4), we find an age effect of about 6-7% of a standard deviation for a one-month increase in age, with the small quadratic term indicating that the effect is linear in age.[2] In terms of our notation in the paper, this is an estimate of $\gamma_2 = \beta_2 + \beta_3$, the effect of being older at school entry and at the test, in absolute

[1] For most of the outcomes we use, age at test is collinear with birth month, and therefore $f(AGET_{imr})$ controls for both, age at test and birth month effects. A possible reason for policy-region specific birth month effects is variation in maternal characteristics across birth months (Buckles and Hungerman 2013) that could differ across policy regions. The RD approach controls for this, as it estimates a flexible functional form in birth month (age at test) separately by policy region.

[2] We verified that when specifying age non-parametrically as a set of monthly age dummies, the pattern of the coefficients on the age dummies is also linear.

and relative terms, but potentially confounded by the effect of being born in an earlier birth month. The size of this effect is of plausible magnitude. For comparison, Elder and Lubotsky (2009) find a comparable age effect of about 4-5% of a standard deviation.[3] In columns (5) and (6) we repeat this regression for policy areas B and C, adding the respective cut-off dummies. Depending on the cut-off and policy area, test scores fall by between 7% and 14% of a standard deviation, conforming well to the visual impression of Figure 4 in the paper. In column (7) we obtain an RDD estimate of the exposure effect by relating the drop in test scores at the cut-off by the corresponding drop in actual exposure for the three cut-offs across the two regions. Two out of the three exposure effects are statistically significant at the 10%-level and imply that a one-month increase in exposure to reception class raises the overall test score at the end of first grade by between 7% and 9% of a standard deviation. In terms of our notation in the paper, this is an estimate of $\gamma_1 = \beta_1 - \beta_3$, a clean effect of gaining additional exposure to schooling by entering school younger, unconfounded by birth month, relative age, or academic ability. This range of results corresponds exactly to what we find in our baseline IV specification in Table 3 in the paper.

APPENDIX C – Interpretation of the TSLS coefficient with two instruments and a treatment intensity

Expected exposure EEXP is a multivalued instrument taking on the values 4, 7 and 11. We can fully describe it by the two dummy variable instruments $z_1 = \mathrm{I}(EEXP \geq 7)$ and $z_2 = \mathrm{I}(EEXP = 11)$. In the following, we keep conditioning on covariates and conditioning on the respective other instrument implicit.[4]

Let the predicted first stage be $\widehat{EXP_i} = \pi_0 + \pi_1 z_{1i} + \pi_2 z_{2i}$. The TSLS estimator of the effect of the endogenous variable EXP on the outcome Y then identifies (Angrist and Pischke 2009, p. 174):

[3] They report an effect of 53% of a standard deviation on reading test scores for being one year older at the kindergarten entry assessment, corresponding to an effect of 53%/12=4.4% for being one month older.

[4] In Angrist and Pischke (2009, p. 174) the two instruments are assumed mutually exclusive. We do not make that assumption, and we therefore condition each instrument's Wald ratio not only on covariates, but also on the other instrument. To make the notation less burdensome, we do not write the conditioning explicitly.

$$\gamma_1^{TSLS} = \frac{\text{Cov}(Y_i, \widehat{EXP_i})}{\text{Cov}(EXP_i, \widehat{EXP_i})} = \mu_1 \frac{E[Y_i|z_{1i}=1] - E[Y_i|z_{1i}=0]}{E[EXP_i|z_{1i}=1] - E[EXP_i|z_{1i}=0]}$$
$$+\mu_2 \frac{E[Y_i|z_{2i}=1] - E[Y_i|z_{2i}=0]}{E[EXP_i|z_{2i}=1] - E[EXP_i|z_{2i}=0]}$$

This is a weighted average of two Wald ratios. The first Wald ratio is the effect induced by a shift in expected exposure from 4 to 7 months (i.e., exploiting z_1 conditional on z_2), while the second is the effect induced by a shift of expected exposure from 7 to 11 months (i.e., exploiting z_2 conditional on z_1). The weights are:

$$\mu_1 = \frac{\pi_1 \text{Cov}(EXP, z_1)}{\pi_1 \text{Cov}(EXP, z_1) + \pi_2 \text{Cov}(EXP, z_2)}, \mu_2 = 1 - \mu_1.$$

We estimate $\hat{\mu}_1 = .2$ and $\hat{\mu}_2 = .8$, which implies that our TSLS estimate is predominantly driven by instrument changes from 7 to 11 months rather than by changes from 4 to 7 months.

The variable EXP is a treatment intensity varying in the range between 1 and 12.[5] Based on theorem 4.5.3 in Angrist and Pischke (2009, p. 182), the estimate can thus be further decomposed into:

$$\gamma_1^{TSLS} = \mu_1 \sum_{\text{s=1}}^{12} \varpi_{1,\text{s}} E\left[Y_{\text{s,i}} - Y_{\text{s-1,i}} | EXP_{z_1=1,i} \geq \text{s} > EXP_{z_1=0,i}\right]$$
$$+ \mu_2 \sum_{\text{s=1}}^{12} \varpi_{2,\text{s}} E\left[Y_{\text{s,i}} - Y_{\text{s-1,i}} | EXP_{z_2=1,i} \geq \text{s} > EXP_{z_2=0,i}\right] \quad (A1)$$

Here, $E\left[Y_{\text{s,i}} - Y_{\text{s-1,i}} | EXP_{z_k=1,i} \geq \text{s} > EXP_{z_k=0,i}\right]$ is the marginal causal effect ("unit causal response") of an increase in the treatment intensity from *s*-1 to *s* for compliers with the instrument z_k (k=1,2) at the treatment intensity *s*. A complier at treatment intensity *s* is an individual who would choose a treatment intensity smaller than *s* when the instrument is switched off, and a treatment intensity larger or equal than *s* if the instrument is switched on. The weights for each unit causal response are

$$\mu_k \varpi_{\text{k,s}} = \mu_k \frac{P\left[EXP_{z_k=1,i} \geq \text{s} > EXP_{z_k=0,i}\right]}{\sum_{\text{l=m}}^{12} P\left[EXP_{z_k=1,i} \geq \text{m} > EXP_{z_k=0,i}\right]}, k = (1,2) \quad (A2)$$

(Angrist and Pischke 2009, p. 182). We present these weights in Table A1.

[5] Although negligible, there are a few cases in our data in which the school entry date was August 2005, leading to an actual exposure of 12 months (up to July 2006).

Equation A1 illustrates that the TSLS effect we identify with our empirical strategy is a weighted average of unit causal responses, averaged over both, treatment intensities and instruments.

At each value *s* of the treatment intensity, we can define a unit causal response for the "non-treated" as $E\left[Y_{\mathrm{s,i}} - Y_{\mathrm{s\text{-}1,i}} | EXP_i < \mathrm{s}\right]$, which is the effect of increasing the treatment from *s*-1 to *s* on individuals who currently receive a treatment smaller than *s*. We can also define never-takers at treatment intensity *s* as individuals who take a treatment intensity of less than *s* even if the instrument is switched on ($EXP_{z_k=1,i} < \mathrm{s}$), and always-takers as individuals who take a treatment intensity greater than or equal to *s*, even if the instrument is switched off ($EXP_{z_k=0,i} \geq \mathrm{s}$).[6] In this framework, there are only two mutually exclusive ways in which an individual can be non-treated at treatment intensity *s* ($EXP_i < \mathrm{s}$): either by being a never-taker, or by being a complier with the instrument switched off. Thus, conditional on being non-treated, the probabilities of being a complier and a never-taker have to sum to one:

$$P\left[EXP_{z_k=1,i} \geq \mathrm{s} > EXP_{z_k=0,i}, z_k = 0 | EXP_i < \mathrm{s}\right] + P\left[EXP_{z_k=1,i} < \mathrm{s} | EXP_i < \mathrm{s}\right] = 1,$$

and these two probabilities are the weights for decomposing the treatment effect on the non-treated at treatment intensity *s* into a weighted average of the treatment effects for compliers and the treatment effect for never-takers:

$$\begin{aligned} &E\left[Y_{\mathrm{s,i}} - Y_{\mathrm{s\text{-}1,i}} | EXP_i < \mathrm{s}\right] \\ &= E\left[Y_{\mathrm{s,i}} - Y_{\mathrm{s\text{-}1,i}} | EXP_{z_k=1,i} \geq \mathrm{s} > EXP_{z_k=0,i}\right] P\left[EXP_{z_k=1,i} \geq \mathrm{s} > EXP_{z_k=0,i}, z_k = 0 | EXP_i < \mathrm{s}\right] \\ &+ E\left[Y_{\mathrm{s,i}} - Y_{\mathrm{s\text{-}1,i}} | EXP_{z_k=1,i} < \mathrm{s}\right] P\left[EXP_{z_k=1,i} < \mathrm{s} | EXP_i < \mathrm{s}\right] \qquad (A3) \end{aligned}$$

If there are no never-takers at treatment intensity *s*, then the first weight will be equal to 1, and the second weight equal to zero. Therefore, the unit causal response for compliers will then be an effect on the non-treated (i.e., those who receive a treatment smaller than *s*). To get an idea of whether we can interpret our overall TSLS effect in eq. A1 as a treatment effect on the non-treated, we estimate the shares of never-takers at each treatment intensity and for each of the instruments, and compute a weighted average using the weights given in eq. A2 (see Appendix table A1). The resulting numbers, reported in the last row of table A1, show that after weighting the shares of never-takers for each instrument at each treatment intensity

[6] Above we have already defined compliers by the condition $EXP_{z_k=1,i} \geq \mathrm{s} > EXP_{z_k=0,i}$. The no-defiers assumption in this context is $EXP_{z_k=1,i} \geq EXP_{z_k=0,i}$, the treatment intensity cannot decrease as a result of switching on the instrument.

with the weights according to which the instruments and the treatment intensities enter our overall TSLS estimate, the share of never-takers is only around 5%. That suggests that our treatment effect should be close to the treatment effect on the non-treated, hence the effect of extending early schooling on those who currently receive little early schooling.

References:

Angrist, J., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics* (Princeton: Princeton University Press).

APPENDIX D - One-sided noncompliance

In Table A1, we report the estimated fractions of compliers, never-takers, and always-takers at different values of exposure (i.e., different treatment intensities) and for the two possible instrument changes.[7] The last row of the table gives the weighted sum of these fractions over all treatment intensities and over the two instruments.[8] On average, an instrument change from 4 to 7 contributes 0.5% of never-takers, while an instrument change from 7 to 11 contributes 4.6% of never-takers, implying an average group size of never-takers of 5%. In contrast, the average group size of always-takers with respect to both instruments is 62%. The low estimate for the never-takers implies that the identified effect is close to the treatment effect on the nontreated. This is because under the usual LATE framework assumptions (monotonicity, no defiers), there can be only two mutually exclusive ways of being nontreated: being a complier with the instrument switched off or being a never-taker (Angrist and Pischke 2009).[9] If there are no never-takers, however, then all nontreated are compliers, meaning that the treatment effect estimated on the basis of compliers is the average treatment effect on the nontreated. The interpretation of our TSLS estimate, therefore, is that it mainly captures the effect of extending exposure to early schooling for those individuals who currently have low levels of exposure. Moreover, one-sided non-

[7] Never-takers are individuals who choose low treatment intensities even when the instrument is switched on; always-takers are individuals who choose high treatment intensities even when the instrument is switched off (Angrist, Imbens and Rubin 1996). We define this more formally in Appendix C.

[8] The weighting over the two instruments is done by using the corresponding IV weights, see Mogstad and Wiswall (2010) for computational details.

[9] Because we have a multivalued treatment intensity, compliers, never-takers, and the treatment effect on the nontreated can be calculated at each value of the treatment intensity and the unit causal response at a given treatment intensity *s* will be an effect on the nontreated at point *s* when there are no never-takers at point *s* (see eq. A3 in the Appendix C).

compliance rules out the existence of defiers, and monotonicity is automatically satisfied (Imbens 2014).

References:

Angrist, J., and G.W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association*, 90(430): 431–442.

Angrist, J., G.W. Imbens, and D.B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 94(434): 444–455.

Angrist, J., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics* (Princeton: Princeton University Press).

Imbens, G. W. 2014. "Instrumental Variables: An Econometrician's Perspective." *Statistical Science* 29(3): 323–358.

Imbens, G. W., and J. D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2): 467–475.

Mogstad, M. and M. Wiswall. 2010. "Linearity in Instrumental Variables Estimation: Problems and Solutions." IZA Discussion Paper No. 5216 (Institute for the Study of Labor).

APPENDIX E - Admissions Policies

Policy A: All children, regardless of age, start school in the September of the academic year in which they turn 5. This policy covers 63% of children.

Policy B: Children born September 1 to February 29 start school in the September of the academic year in which they turn 5; children born March 1 to August 31 start school in the January of the academic year in which they turn 5. This policy covers 18% of children.

Policy C (the rising 5 policy): Children start school at the beginning of the term in which they turn 5, so children born September 1 to December 31 start school in September, children born January 1 to April 30 start school in January, and children born May 1 to August 31 start school in April. This policy covers close to 15% of children.

Policy D: Children born September 1 to April 30 start school in the September of the academic year in which they turn 5 and children born May 1 to August 31 start school in the January of the academic year in which they turn 5. This policy covers 3% of children.

Policy E: Children born September 1 to December 31 start school in the September of the academic year in which they turn 5 and children born January 1 to August 31 start school in the January of the academic year in which they turn 5. This policy covers 1% of children.

APPENDIX F - Description of the Cognitive and Noncognitive Outcome Measures

The Foundation Stage Profile at age 5 (MCS and NPD data): This assessment at the end of first grade measures cognitive outcomes (communication, language, and literacy skills; mathematical development; knowledge and understanding of the world) and noncognitive outcomes (creative development; physical development; and personal, social, and emotional development), and also provides a total score. Please refer to Appendix G for a list of the items for each of these scores. We use the "communication, language, and literacy" score as our 1st grade Language Skills score, and the common factor of the "mathematical development" and "knowledge and understanding of the world" scores as our 1st grade Numeracy Skills score.

The Key Stage 1 Assessment (MCS and NPD data): This in-school assessment takes place in third grade (officially called "Year 2" of primary school) when children are usually aged 7. Nationwide test materials are developed and provided centrally by the Standards & Testing Agency (STA) of the Department for Education and administered by schools under supervision from local authorities. The teacher guidelines for this assessment are attached as Appendix H below. The test is administered and graded by class teachers based on the official grading schemes. The assessment provides point scores for reading, writing, mathematics, and science, as well as an average point score over these four subjects. We use the common factor of the reading and writing score as our 3rd grade Language Skills score and the common factor of the math and science score as our 3rd grade Numeracy Skills score.

The Key Stage 2 Assessment (NPD data): This in-school assessment is taken at the end of seventh grade (officially called "Year 6" of primary school) when children are usually aged 11. It is based on national tests developed and provided by the Standards & Testing Agency (STA) of the Department for Education, and administered in schools according to a binding national time schedule. The Key Stage 2 Assessments are marked externally. The NPD data contains reading, writing, mathematics, and science scores from that assessment. We use the common factor of the reading and writing score as our 7th grade Language Skills score and the common factor of the math and science score as our 7th grade Numeracy Skills score.

Self-Assessment (child) Questionnaire at age 7 (MCS): We aggregate a range of variables from the child self-assessment questionnaire into the following factors:

Teacher Relationship I: Factor aggregating the variables "How often does your teacher think you are clever?", "What I do if worried – tell a teacher."

Academic Interest I: Factor aggregating the variable "How often do you get fed up at school?" and six further variables "How much do you like [school, reading, numbers, science, PE, answering in class]?"

Positive Self-Perception I: Factor aggregating the variables "How often do you try to do your best at school?" and "How often does your teacher think you are clever?"

Disruptive Behavior I: Factor aggregating the variables "I behave always well in school" and "I behave sometimes well in school", multiplied by minus 1 to make it indicate disruptive behavior.

Parental Questionnaire and Self-Assessment (child) Questionnaire at age 11 (MCS): We aggregate variables from both the parental and the self-assessment questionnaire into the following factors:

Teacher Relationship II: Factor aggregating the variables, "How much do you like your class teacher?", "How often do you think your class teacher is getting at you?", "How often does your child look forward to seeing the school teacher?"

Academic Interest II: Factor aggregating the variables "How often do you find school interesting?", "How often do you get tired at school?", "How often to you try your best at school?", Happiness with school work and the school, "Would your child like to stay on at school [after the end of compulsory schooling]?", "How long does your child spend on homework?" and "How often does your child seem bored at school?"

Positive Self-Perception II: Factor aggregating the variables "How strongly do you agree: I have a number of good qualities" and "How strongly do you agree: I am able to do things as well as most other people."

Disruptive Behavior II: Factor aggregating the variables "Have you ever been noisy or rude in a public place?" and "Have you ever taken something from a shop without paying for it?", "Have you ever written things or spayed paint on a building?", "Have you ever damaged anything on purpose in a public place?", and "How often do you misbehave or cause trouble in class?"

APPENDIX G – First-Grade Assessment Scales (Foundation Stage Profile)

There are 13 assessment scales on each of which a child can gain a maximum of 9 points. The items are ordered in ascending difficulties, and the child gets assigned the highest point it reaches on each of these scales.

PERSONAL, SOCIAL & EMOTIONAL DEVELOPMENT

1. Dispositions & attitudes

D1. Shows an interest in activities through observation or participation.
D2. Dresses, undresses & manages own personal hygiene with adult support.
D3. Displays high levels of involvement in self-chosen activities.
D4. Dresses & undresses independently & manages own personal hygiene.
D5. Selects & uses activities and resources independently.
D6. Continues to be interested, motivated & excited to learn.
D7. ls confidant to try new activities, initiate ideas & speak in a familiar group.
D8. Maintains attention & concentrates.
D9. Sustains involvement & perseveres, particularly when trying to solve a problem or reach a conclusion.

2. Social development

S1. Plays alongside others.
S2. Builds relationships through gesture & talk.
S3. Takes turns and shares with adult support.
S4. Works as part of a group or class, taking turns, sharing fairly.
S5. Forms good relationships with adults & peers.
S6. Understands that there needs to be agreed values & codes of behavior to work together.
S7. Understands that people have different needs, views, cultures & beliefs.
S8. Understands that s/he can expect others to treat her/his needs, views, cultures, & beliefs with respect.
S9.Takes in to account the needs of others.

3. Emotional development

E1. Separates from main carer with support
E2. Communicates freely about home and community.
E3. Expresses needs and feelings in appropriate ways.
E4. Responds to significant experiences showing a range of feelings.
E5. Has a developing awareness of own needs.
E6. Has a developing respect for own cultures and beliefs.
E7. Considers the consequences of words and actions for self and others.
E8.Understands what is right, what is wrong and why.
E9.Displays a strong and positive sense of self-identity & is able to express a range of emotions fluently & appropriately.

COMMUNICATION, LANGUAGE & LITERACY

4. Language for Communication and thinking

C1. Listens & responds.
C2. Initiates communication with others, displaying greater confidence in more informal contexts.
C3. Talks activities through, reflecting on & modifying actions.
C4. Listens with enjoyment to stories, songs, rhymes, poems, sustains attentive listening & responds with relevant comments, questions, actions.
C5. Use language to imagine and recreate roles and experiences.
C6. Interacts with others in a variety of contexts, negotiating plans and activities and taking turns in conversation.
C7. Uses talk to organize, sequence and clarify thinking, ideas, feelings, events, exploring the meanings and sounds of words
C8. Speaks clearly with confidence & control, showing awareness of the listener.
C9. Talks & listens confidently & with control, consistently showing by awareness of the listener by including relevant detail. Uses language to work out & clarify ideas, showing control of a range of appropriate vocabulary.

5. Linking sounds and letters

L1. Joins in with rhyming and rhythmic activities
L2. Shows & interest in rhyme & alliteration
L3. Links some sounds to letters.

L4. Links sounds to letters, naming & sounding letters of the alphabet.
L5. Hears & says initial & final sounds in words. L6.Hears & says short vowel sounds within words.
L7. Uses phonic knowledge to read simple regular words.
L8. Attempts to read more complex words, using phonic knowledge.
L9. Uses knowledge of letters, sounds & words when reading & writing independently.

6. Reading

R1. Is developing an interest in books.
R2. Knows that print conveys meaning.
R3. Recognizes a few familiar words.
R4. Knows that, in English, print is read from left to right, top to bottom.
R5. Shows and understanding of the elements of stories, such as main character, sequence of events, openings.
R6. Reads a range of familiar words & simple sentences independently.
R7. Retells narrative in the correct sequence, drawing on language patterns of stories.
R8. Shows an understanding of how information can be found in non-fiction texts to answer questions about where, who, why, how.
R9. Reads books of own choice with some fluency & accuracy.

7. Writing

W1. Experiments with mark making, sometimes ascribing meaning to marks.
W2. Uses some clearly identifiable letters to communicate meaning.
W3. Represents some sounds correctly in writing.
W4. Writes own name & other words from memory.
W5.Hold a pencil and uses it effectively to form recognizable letters, lots of which are correctly formed.
W6. Attempts writing for a variety of purposes using features of different forms
W7. Uses phonic knowledge to write simple regular words and makes phonetically plausible attempts at more complex words.
W8. Begins to form captions, simple sentences, sometimes using punctuation
W9. Communicates meaning through phrases & simple sentences with some consistency in punctuating sentences.

MATHEMATICAL DEVELOPMENT

8. Numbers as labels

N1. Says some number names in familiar contexts such as nursery rhymes.
N2. Counts reliably up to three everyday objects.
N3. Counts reliably up to six everyday objects.
N4. Says number names in order
N5. Recognizes numerals 1 to 9.
N6. Counts reliably up to 10 everyday objects.
N7.Orders numbers, up to 10.
N8. Uses developing mathematical ideas & methods to solve practical problems.
N9.Recognisies, counts, orders, writes and uses numbers up to 20.

9. Calculating

C1. Responds to the vocabulary involved in addition & subtraction in rhymes 7 games. C2. Recognizes differences in quantity when comparing a set of objects.
C3. Finds one more or one less from a group of up to five objects.
C4. Relates addition to combining 2 groups.
C5.Relates subtraction to taking away.
C6. ln practical activities & discussion, begins to use the vocabulary involved in addition & subtracting.
C7. Finds one more or one less than a number from 1 to 10.
C8. Uses developing mathematical ideas and methods to solve practical problems C9. Uses a range of strategies for addition & subtraction, including some mental recall of number bonds.

10. Shape, space & measures

S1. Experiments with a range of objects & materials showing some mathematical awareness.
S2. Sorts or matches objects & talks about sorting.
S3. Describes shapes in simple models,
S4.Talks about, recognizes & recreates simple patterns.
S5.Uses everyday words to describe position.
S6. Uses language such as ‘circle’ or ‘bigger’ to describe the shape and size of solids and flat shapes.
S7. Uses language such as 'greater', 'smaller', 'heavier', and 'lighter' to compare quantities.
S8. Uses developing mathematical ideas & methods to solve practical problems.
S9. Uses mathematical language to describe solid (3D) objects and flat (2D) shapes.

11. KNOWLEDGE & UNDERSTANDING Of THE WORLD

K1. Shows curiosity and interest by exploring surroundings.

K2. Observes, selects & manipulates objects & materials. Identifies simple features and significant personal events.

K3. Identifies obvious similarities & differences when exploring & observing. Constructs in a purposeful way, using simple tools & techniques.

K4. lnvestigates places, objects, materials & living things by using all the senses as appropriate. Identifies some features & talks about those features s/he likes & dislikes. K5. Asks questions about why things happen and how things work. Looks closely at similarities, differences, patterns & change

K6. Finds out about past & present events in own life, & in those family members & other people s/he knows. Begins to know about own culture & beliefs & those of other people.

K7. Finds out about & identifies the uses of everyday technology & uses information and communicate technology and programmable toys to support her/his learning.

K8. Builds & constructs with a wide range of objects, selecting appropriate
resources, tools & techniques & adapting her/his work where necessary.

K9. Communicates simple planning for investigations & constructions & makes simple records & evaluations of her/his work. Identifies & names key features & properties, sometimes linking different experiences, observations & events. Begins to explore what it means to belong to a variety of groups & communities.

12. PHYSICAL DEVELOPMENT

P1. Moves spontaneously, showing some control & co-ordination.

P2. Moves with confidence in a variety of ways, showing some awareness of space.

P3. Usually shows appropriate control in large-and small-scale movements.

P4. Moves with confidence, imagination & in safety. Travels around, under, over & through balancing & climbing equipment. Shows awareness of space, of self & others.

P5. Demonstrates fine motor control and co-ordination.

P6. Uses small & large equipment, showing a range of basic skills.

P7. Handles tools, objects, construction, & malleable materials safely &with basic control.

P8. Recognizes the importance of keeping healthy & those things which contribute to this. Recognizes the changes that happen to her/his body when s/he is active

P9. Repeats, links & adapts simple movements, sometimes commenting on her/his work. Demonstrates co-ordination & control in large & small movements, & in using a range of tools & equipment.

13. CREATIVE DEVELOPMENT

C1. Explores different media & responds to a variety of sensory experiences.
Engages in representational play.

C2. Creates simple representations of events, people & objects & engages in music making.

C3. Tries a capture experiences, using a variety of different media.

C4. Sings simple songs from memory

C5. Explores color, texture, shape, form & space in two or three dimensions.

C6. Recognizes & explores how sounds can be changed. Recognizes repeated sounds & sound patterns & matches movements to music.

C7. Uses imagination in art & design, music, dance, imaginative & role-play & stories. Responds in a variety of ways to what s/he sees, hears, smells, touches & feels.

C8. Expresses & communicates ideas, thoughts & feelings using a range of materials, suitable tools, imaginative and role-play, movement, & designing & making, & a variety of songs & musical instruments.

C9. Expresses feelings & preferences in response to artwork, drama, & music & makes some comparisons & links between different pieces. Responds to own work and that of others when exploring & communicating ideas, feelings & preferences through art, music, dance, role-play and imaginative play.