Supplementary Material for:

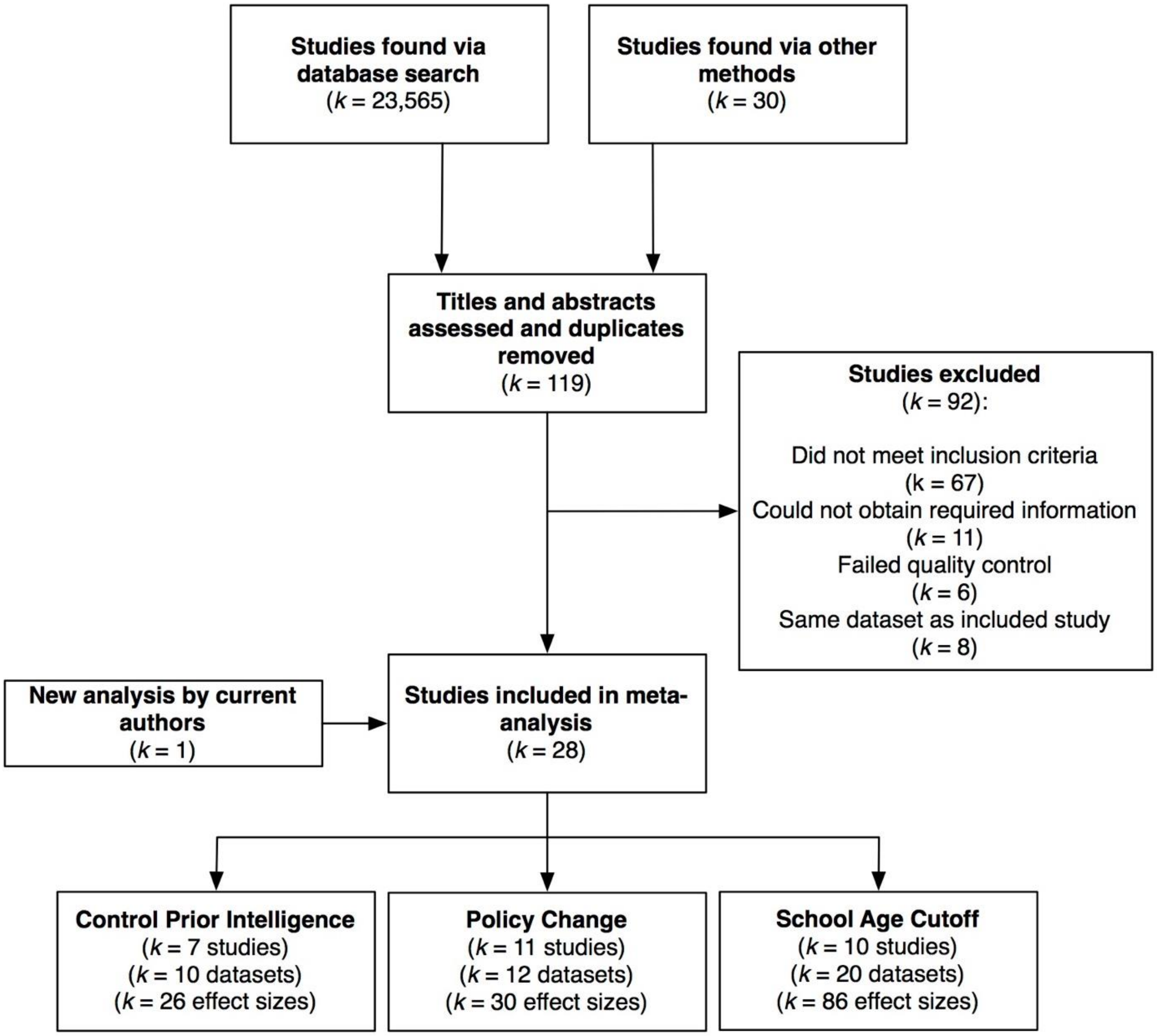
How much does education improve intelligence? A meta-analysis

Stuart J. Ritchie & Elliot M. Tucker-Drob

Contents:

Open Science Framework page URL: https://osf.io/r8a24/

1. *Figure S1.* Flow diagram for literature search and study inclusion



2. *Table S1.* Methodological characteristics of each Policy Change study included in the meta-analysis.

Study	Instrument(s)/intervention(s)	Comparison	Age at which change/difference/intervention occurred
Lager et al. (2016)	Swedish educational reform (1950s): increase lower compulsory education from 8 to 9 years	IQ post-reform versus pre-reform where date of reform varies quasi-randomly across municipalities estimated using multilevel linear regression with municipality fixed effects.	15-16 years
Brinch & Galloway (2012)	Norwegian educational reform (1955-1972): increased compulsory schooling from 7 to 9 years	IQ post-reform versus pre-reform where date of reform varies quasi-randomly across municipalities, estimated using an instrumental variable approach.	14-16 years
Glymour et al. (2008)	Minimum amount of compulsory education calculated as the difference between State/Birth Cohort-specific age of enrolment and State/Birth Cohort-specific compulsory schooling or minimum work age laws	Association of memory and mental status with variation in amount of compulsory education across states and birth cohorts, calculated using a separate-sample instrumental variable approach.	6-14 years
Gorman (2016)	1972 UK schooling reform: raised school leaving age from 15 to 16 years of age	Test scores in subjects born before or after 1 September 1957 (unaffected by reform versus affected) estimated using a regression discontinuity approach.	15-16 years
Banks & Mazzonna (2012)	1947 English schooling reform: raised school leaving age from 14 to 15 years of age	Effect of being born before or after 1 April 1933 (unaffected by reform versus affected) on test scores, estimated with regression discontinuity approach.	14-15 years
Schneeweis et al. (2014)	Years of compulsory schooling (due to both cross-national-differences and cross-time changes in compulsory schooling laws) across 6 European countries	Country- and birth-cohort specific years of compulsory schooling with controls for fixed effects of country and cohort, estimated as an instrumental variable analysis.	11-16 years
Carlsson et al. (2015)	Exogenous random variation in National Service conscription intelligence test dates	More or fewer school days completed at time of intelligence test, estimated in a linear regression with controls for birthdate, parish, and date of expected graduation	~18 years
Kamhöfer & Schmitz (2016)	Post-1964 German reform increasing compulsory schooling, or state-level deviations from national trend in school availability	Affected vs. unaffected by the reform or by greater or lesser school availability (due to cross-state differences), estimated as an instrumental variable analysis, controlling for sex and cohort and state fixed effects	13-14 years

Kamhöfer et al. (2015)	Post-1958 German college availability, or post-1971 German availability of student loans	University graduate vs. non-graduate, estimated using an instrumental variables approach with a large number of controls	18-20 years (?)
Xiao et al. (2017)	2006 Chinese free compulsory education reforms	Subjects in reform provinces versus subjects in non-reform provinces, estimated using a differences-in-differences design with a large number of controls	6-15 years
Davies et al. (2018)	1972 UK schooling reform: raised school leaving age from 15 to 16 years of age	Subjects born before or after 1 September 1957 (unaffected by reform versus affected), estimated using an instrumental variable approach	15-16 years

Note: The following study was excluded for the Policy Change design as we judged it to have a potentially confounded instrument (exposure to the 1959-61 Great Famine in China): Huang, W., & Zhou, Y. (2013). Effects of education on cognition at older ages: evidence from China's Great Famine. *Social Science & Medicine*, *98*, 54-62.

3. Inclusion criteria and quality control for the School Age Cutoff design

This subsection details the strict inclusion criteria we used for studies that used the School Age Cutoff design.

The School Age Cutoff design is an implementation of "fuzzy" regression discontinuity analysis. Regression discontinuity analysis gains its inferential power from the fact that an incentive or eligibility criterion for assignment to an intervention is contingent on scores on an observed assignment variable. In the context of school age cutoff design, initiation of formal schooling at the beginning of the school year is contingent upon the child being a certain age at the school cutoff date: for example, age 5 years by September 1 of the school year. The effect of schooling on the outcome of interest, in this case cognitive test score, is estimated with a regression equation with a term specifying a continuous relation between age and scores on the outcome and a term specifying a discontinuous step function relating criterion-based eligibility (having been at least 5 years old by September 1) to scores on the outcome. The effect of age on scores is estimated as the within-grade age effect on scores, and the effect of a year of schooling on scores is estimated as the displacement of the age regression function for the older grade relative to that for the younger grade.

In reality, school eligibility criteria are not perfectly enforced. Some children who meet the age cutoff for eligibility forgo initiating schooling during what should be their kindergarten year. Other children who do not meet eligibility are sometimes promoted to kindergarten a year before they are eligible. Because grade assignment is probabilistic rather than deterministic, the regression discontinuity analysis is referred to as "fuzzy". This is a particular concern if the decision to promote children late or early relates to their "pre-treatment" IQ scores. This may occur, for example, when parents or teachers decide that children who have not yet mastered literacy or numeracy skills would be best served by waiting an additional year before beginning kindergarten (a phenomenon colloquially known as "redshirting"), or when particularly precocious children are selected to begin kindergarten early. Under these circumstances, an impression of a discontinuity—seemingly indicative of a schooling effect—may occur simply by virtue of more cognitively advanced children being selected into the older grade.

When regression discontinuity data are fuzzy, such that the compliance with the cutoff criterion is imperfect, steps may be taken to mitigate potential bias. The most rigorous approach involves specifying the step function to be a propensity function of the age-criteria *free from information about the grade that the child is actually in* (Lee & Lemieux, 2010). Such an approach can be considered an instrumental variable implementation of an intent-to-treat approach. We did not observe the implementation of this approach in the "School Age Cutoff" studies. However, several other approaches were taken that we deemed acceptable for mitigating bias. One common approach was to remove individuals with birthdates within several months of the cutoff criterion. As the tendency is for individuals who do not comply with the cutoff criterion to have just missed or just exceeded the cutoff criterion by within approximately two months, many of the individuals disposed toward noncompliance are removed with this approach. We therefore deemed this approach adequate for a study to qualify for inclusion in our meta-analysis.

Some studies did not exclude individuals based on their birthdate proximity to the cutoff, but did remove individuals whose grade-level was not commensurate with the grade that they should have been in on the basis of their birthday. We did not find this approach to be adequate to address selection effects, because even if “red-shirted” individuals are excluded from the lower grade, they are absent from the higher grade to which they are assigned on the basis of their birthday. This still produces a selection effect. We therefore excluded studies that solely took this approach, except in two circumstances in which authors explicitly reported that only a very small proportion of participants were non-compliant, such that the selection effect was minimal (Kyriakides & Luyten, 2009; Luyten et al., 2006).

Additionally, we excluded a study (Mayer & Knutson, 1999) that did not estimate a discontinuity or step function. This study appeared to be focused on the effects of starting school earlier versus later. Finally, we did not include studies that simply examined similarly aged children who differed in their grade (e.g. Brod, Bunge, & Shing, 2017). Such studies are causally ambiguous, because children in higher grades may have higher cognitive scores either because of the extra year of schooling, or because they had higher pre-existing scores that led to them being promoted earlier than those with lower pre-existing scores.

4. Analysis software and data description

The main meta-analytic models were run using Mplus v7.2 (Muthén & Muthén, 1998-2014), as described in the Method section of the main document. Publication bias analyses, and their associated illustrations, were produced in R v3.4.0 (R Core Team, 2017) using the *metafor* (Viechtbauer, 2010) and *ggplot2* packages (Wickham, 2009). The flowchart in Figure 1 was created in OmniGraffle 5.4.3. Data and analysis scripts for the meta-analysis are available on the associated Open Science Framework page (https://osf.io/r8a24/).

The following list describes each file and folder on the OSF page:

- The master data spreadsheet ("EduIQ_master.xlsx") contains all data extracted from every included study, along with notes (rightmost column) on where in each paper each estimate was found.
- The master data codebook ("EduIQ_codebook.pdf") provides a brief description of each column in the master data spreadsheet.
- All R scripts and relevant data are available in the folder "R Scripts and Data":
    - A subset of columns from the master dataset are used in the R input data file ("meta_dataset.txt"), a tab-delimited text file;
    - The Mplus input data file ("metadata.dat") was prepared using the R script "Mplus_filemaker";
    - Figure 1 in the manuscript was prepared using the R script "Age_effect_plots.R";
    - Publication bias tests, and their associated plots, were produced using the R script "Publication_bias_tests.R"; the p-curves graphs require the data files "pcurve_graphdata_cpiq.txt", "pcurve_graphdata_poli.txt", "pcurve_graphdata_schl.txt";
    - The overall funnel plots (Figure S3, below) were prepared using the R script "Overall_funnel_plots.R"
    - Forest plots (Figure S2, below) were produced using the R script "Forest_plot.R" and the three forest plot data files ("forest_data_cpiq.txt", "forest_data_poli.txt", and "forest_data_schl.txt").
- All input and output files from Mplus, along with the Mplus input data file, used for all Mplus analyses, are in the folder "Mplus Analyses", which contains two subfolders for conditional and unconditional meta-analyses. The data file "metadata_mplus.dat" is for the main analyses; the data file "metadata_mplus_pluscovs.dat" is for the analysis including maximal covariates.
- All of the figures from the main study and from the Supplementary Materials are available, in .pdf and .jpeg form, in the folder "Figures".
- For the new analysis of the British Cohort Study data*, the analysis R script ("BCS_analysis.R") is available in the "New BCS Analysis" folder.
- For the two studies where we re-analysed the correlation matrices using structural equation modeling (see subsection 6, below), the matrices and Mplus scripts are in the folder "Reanalyses", in the subfolders "Plassman reanalysis" and "Tonkin reanalysis"

- Outputs for all three study designs from the *p*-curve app on http://p-curve.com/ are provided in the folder "p-curves".

*Note that we did not have permission to upload the data for our new analysis of the British Cohort Study (see below). Those data are available by (free) application to the UK Data Service. Access details can be found at the following URL: http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=947&sitesectiontitle=Accessing+the+data.

5. New analysis of British Cohort Study data

Method

Participants

The British Cohort Study (BCS; Elliot & Shepherd, 2006) is an ongoing longitudinal study of all individuals born during one week of April 1970 (initial n = 17,287). The present study uses cognitive data and the report of sex from the wave at approximate age 10 (n = 14,875; 7,162 females [BCS variable *sex10*]; Butler & Bynner, 2008), and numeracy/educational the wave at approximate age 34 (n = 9,665; University of London, Institute of Education, Centre for Longitudinal Studies, 2008). We removed 292 individuals who were twins (reported at age 10; BCS variable *tc10*), thus restricting the analysis to singletons. Sample sizes for each variable are provided in Table S1.

Table S2. Descriptive statistics and correlation matrix for the British Cohort Study variables used in the analysis.

Variable	1.	2.	3.	4.	5.	6.	7.	*n*	M (SD) or %
1. BAS Definitions age 10	-							11,284	10.14 (5.01)
2. BAS Similarities age 10	.66	-						11,227	11.10 (2.54)
3. BAS Digit Recall age 10	.33	.33	-					11,275	22.40 (4.28)
4. BAS Matrices age 10	.44	.44	.26	-				8,533	16.71 (5.00)
5. Cognitive composite age 10	.81	.80	.62	.72	-			8,433	100 (15)
6. Years of Education	.38	.33	.19	.31	.41	-		8,770	13.42 (3.22)
7. Numeracy age 34	.43	.42	.29	.45	.51	.38	-	8,624	17.83 (4.07)
8. Sex (female)	–.11	–.08	.03	.05	–.04	.03	–.17	14,585	48.2% female

Note: BAS = British Ability Scales. For Numeracy age 34, the mean and SD refer to pre-standardization scores. All correlations significant at $p < .006$.

Measures

Childhood intelligence

Participants completed four subscales of the British Ability Scales at the age-10 testing wave. Two subtests tapped verbal abilities: Definitions (37 items; BCS variables *i3504*-*i3540*) and Similarities (21 items; variables *i4201*-*i4221*). Two tapped nonverbal abilities: Recall of Digits (34 items; variables *i3541*-*i3574*) and Matrices (28 items; variables *i3617*-*i3644*). We standardized the sum score from each subtest (z-scoring), then took the average of these z-scores to form a unit-weighted composite childhood cognitive score. This composite variable was converted to an IQ metric (mean = 100, SD = 15).

Years of Education

Since years spent in education were not explicitly reported in the BCS data, we relied on a variable where the participant reported their highest educational qualification (BCS variable *qual34*). We converted this into the years that each qualification would usually take to obtain, as follows. No qualifications: 11 years (11 years' full-time education was compulsory in the United Kingdom at this time; *n* = 800 or 9.1%); CSE, GCSE, or O-Level: 11 years (*n* = 4236 or 48.3%); A-level, SSCE, or AS-level: 13 years (*n* = 832 or 9.5%); Degree, Diploma of Higher Education, other teaching qualification, or nursing qualification: 17 years (*n* = 2358 or 26.9%); Higher degree or PGCE: 21 years (*n* = 544 or 6.2%).

Adult numeracy

At age 34, the cohort members completed an adult numeracy test (BCS variable *numall*). This involved items testing the understanding of mathematical information in different forms and for different purposes. There were 17 multiple-choice items (covering the topics of "basic money", "whole numbers and time", "measures and proportions", "weights and scales", "length and scaling", "charts and data", and "money calculations") and 6 open-response items (involving time calculations, monetary calculations in fictional scenarios, and the extraction of information from a timetable). The total score from this test was negatively skewed (skew = –0.92; kurtosis = 0.44), indicating ceiling effects. We thus square-transformed the variable to reduce the skew (resulting skew = –0.39; kurtosis = –0.79), before converting it to an IQ metric.

Note that a second test, focusing on literacy, was also taken at age 34 (BCS variable *litort*). However, this test showed a strong ceiling effect (substantially stronger than for numeracy; skew = –2.04; kurtosis = 5.54), indicating that it did not have sufficient difficulty to assess the full range of literacy skills. We thus did not use it as an outcome in this analysis.

Statistical Analysis

We used an ordinary least squares linear multiple regression model to predict age-34 Numeracy score from years of education, the age-10 cognitive composite, and sex. Note that this analysis therefore fit into our meta-analytic category of studies that controlled for the intelligence of the participants prior to variation in educational duration ("Control Prior Intelligence").

Results

The results of the regression model, shown in Table S2, indicated that, after control for the age-10 cognitive composite and sex, years of education was still statistically significantly associated with age-34 numeracy. The effect size was estimated to be 0.916 points (on the IQ scale) for an additional year of education (SE = 0.054, $p = 1.14\times10^{-62}$). For comparison, the estimate for years of education without controlling for the cognitive composite variable (in a secondary model without this control) was 1.797 points per year (SE = 0.045, $p = 1.21\times10^{-316}$). This estimate was entered into the meta-analytic database for the main study.

Table S3. Results from the regression of age-34 numeracy score on years of education, the age-10 cognitive composite, and sex (number of observations = 5,296).

Predictor	Unstandardized estimate	SE	*t*-value	*p*-value
(Intercept)	49.737	1.189	41.82	~0.00
Years of Education	0.916	0.054	16.93	$1.14{\times}10^{-62}$
Cognitive composite age 10	0.415	0.012	33.76	$2.19{\times}10^{-226}$
Sex (female)	–4.653	0.326	–14.28	$2.02{\times}10^{-45}$

Note: Numeracy variable standardized such that mean = 100 and SD = 15; estimates are thus in IQ-point units per unit of the predictor (e.g. years for the education variable).

6. *Table S4.* List of all studies included in the final meta-analysis

Study Design	Authors	Year of publication	Journal/Book/etc. of publication	*k* datasets	*k* effect sizes
Control Prior Intelligence	Ritchie et al.	2012	*Psychology and Aging*	1/2*	4
	Ritchie et al.	2015	*Developmental Psychology*	1*	13
	Clouston et al.	2012	*International Journal of Epidemiology*	3	3
	Herrnstein & Murray	1994	*The Bell Curve*	1	1
	Falch & Sandgren Massih	2011	*Economic Inquiry*	1	1
	Tonkin	1999	PhD Thesis	1	2
	Plassman et al.	1995	*Neurology*	1	1
	Ritchie & Tucker-Drob	2017	Present study	1	1
Policy Change	Lager et al.	2016	*International Journal of Epidemiology*	1	1
	Brinch & Galloway	2012	*PNAS*	1	1
	Glymour et al.	2008	*Journal of Epidemiology & Community Health*	1	2
	Gorman	2016	Working paper	1	4
	Banks & Mazzonna	2012	*Economic Journal*	2	4
	Schneeweis et al.	2014	*Demography*	1	5
	Carlsson et al.	2015	*Review of Economics and Statistics*	1	4
	Kämhofer & Schmitz	2016	*Journal of Applied Economics*	3	3
	Kämhofer et al.	2015	Working paper	1	3
	Xiao et al.	2017	Working paper	1	2
	Davies et al.	2018	*Nature Human Behaviour*	1	1
School Age Cutoff	Jabr & Cahan	2015	*School Effectiveness and School Improvement*	3	3
	Kyriakides & Luyten	2009	*School Effectiveness and School Improvement*	1	3
	Cahan & Cohen	1989	*Child Development*	1	15
	Gambrell	2013	PhD Thesis	1	22
	Artman et al.	2006	*Cognitive Development*	1	3
	Wang et al.	2016	*Educational Psychology*	2	10
	Cahan & Noyman	2001	*Education and Psychological Measurement*	1	14
	Luyten	2006	*Oxford Review of Education*	8	16
	Luyten et al.	2008	*American Educational Research Journal*	1	1
	Cliffordson	2010	*Educational Research and Evaluation*	3	1

Note: Full references can be found in the Supplementary References section, below. *The two studies by Ritchie et al. (2012, 2015) overlapped in their analysis of one dataset (the Lothian Birth Cohort 1936), and are thus classed as one study for these purposes.

7. Reanalysis of correlation matrices in Plassman et al. (1995) and Tonkin (1999)

In order to obtain per-year effect size estimates for the studies by Plassman et al. (1995) and Tonkin (1999), we extracted the relevant information from the correlation matrices they provided, along with the relevant descriptive statistics (means and standard deviations), and used these as input to structural equation models in Mplus which were set up to re-estimate the effect sizes in the standardized units required (the outcome variables were rescaled into standard IQ units with a mean of 100 and a standard deviation of 15, to conform with the figures in the rest of the meta-analysis). The relevant data from Plassman et al. (1995) and from Tonkin (1999) are displayed in Tables S3 and S4, respectively.

Table S4. Correlations and descriptive statistics from Plassman et al. (1995).

	1.	2.	Mean (SD)
1. Army General Classification Test (AGCT)	-		106.39 (18.99)
2. Telephone Interview for Cognitive Status (TICS)	.457	-	100 (15)
3. Education (years)	.555	.408	13. 17 (2.98)

Note. Number of observations = 930.

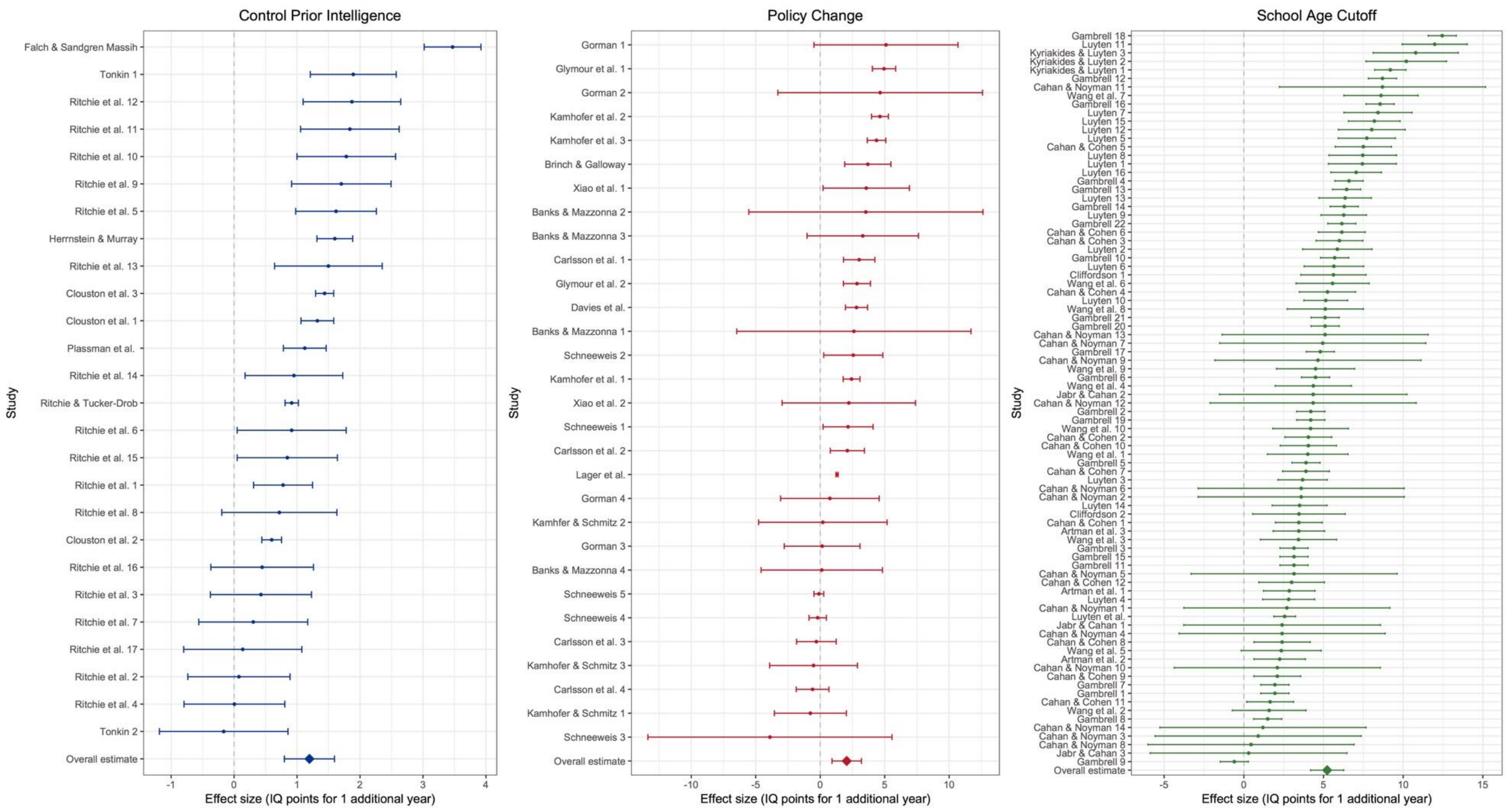
Table S5. Correlations and descriptive statistics from Tonkin et al. (1999).

	1.	2.	3.	Mean (SD)
1. IQ age 18 years	-			122 (11.70)
2. Education (years)	.55	-		16.1 (2.8)
3. Adult Verbal IQ	.77	.67	-	100 (15)
4. Adult Performance IQ	.53	.27	.53	100 (15)

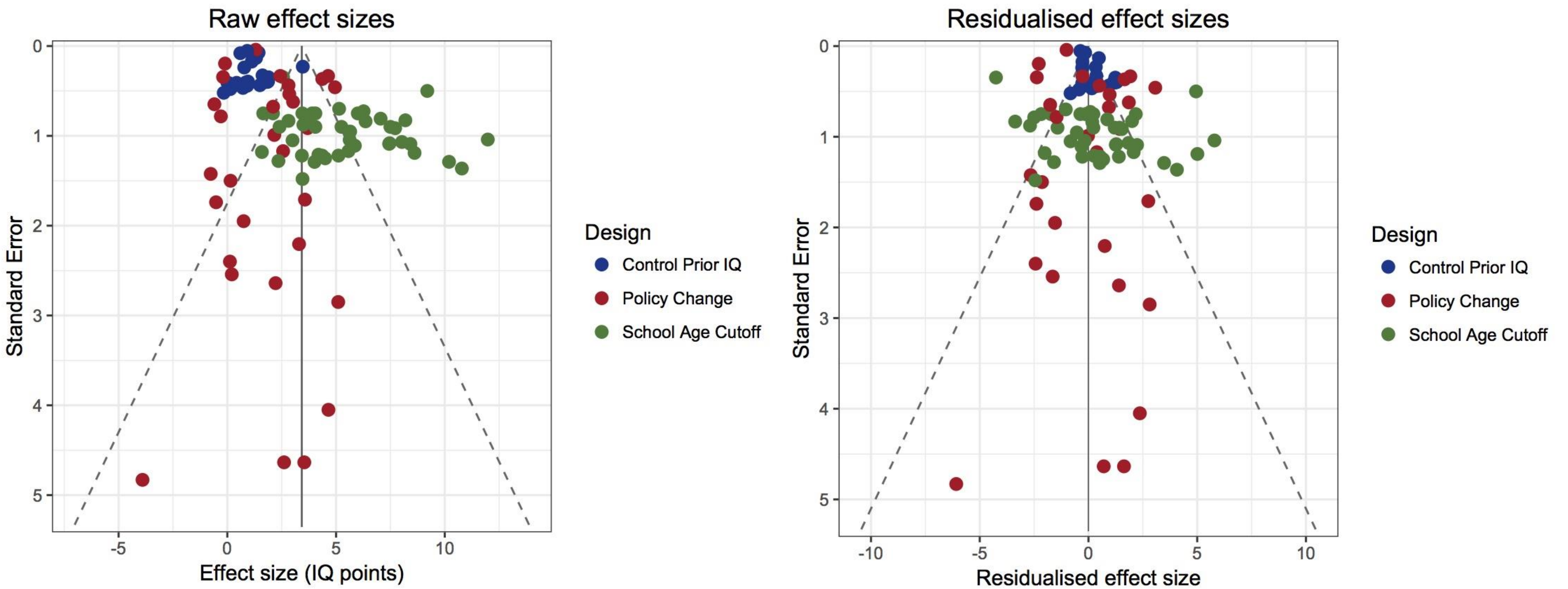
Note. Number of observations = 109.

The Mplus scripts for both these analyses are available alongside the manuscript on the associated Open Science Framework page (https://osf.io/r8a24/). Note that the Tonkin (1999) re-analysis has two scripts: one to estimate the effect on Verbal IQ and one for Performance IQ.

8. *Figure S2.* Forest plots for each of the three study designs

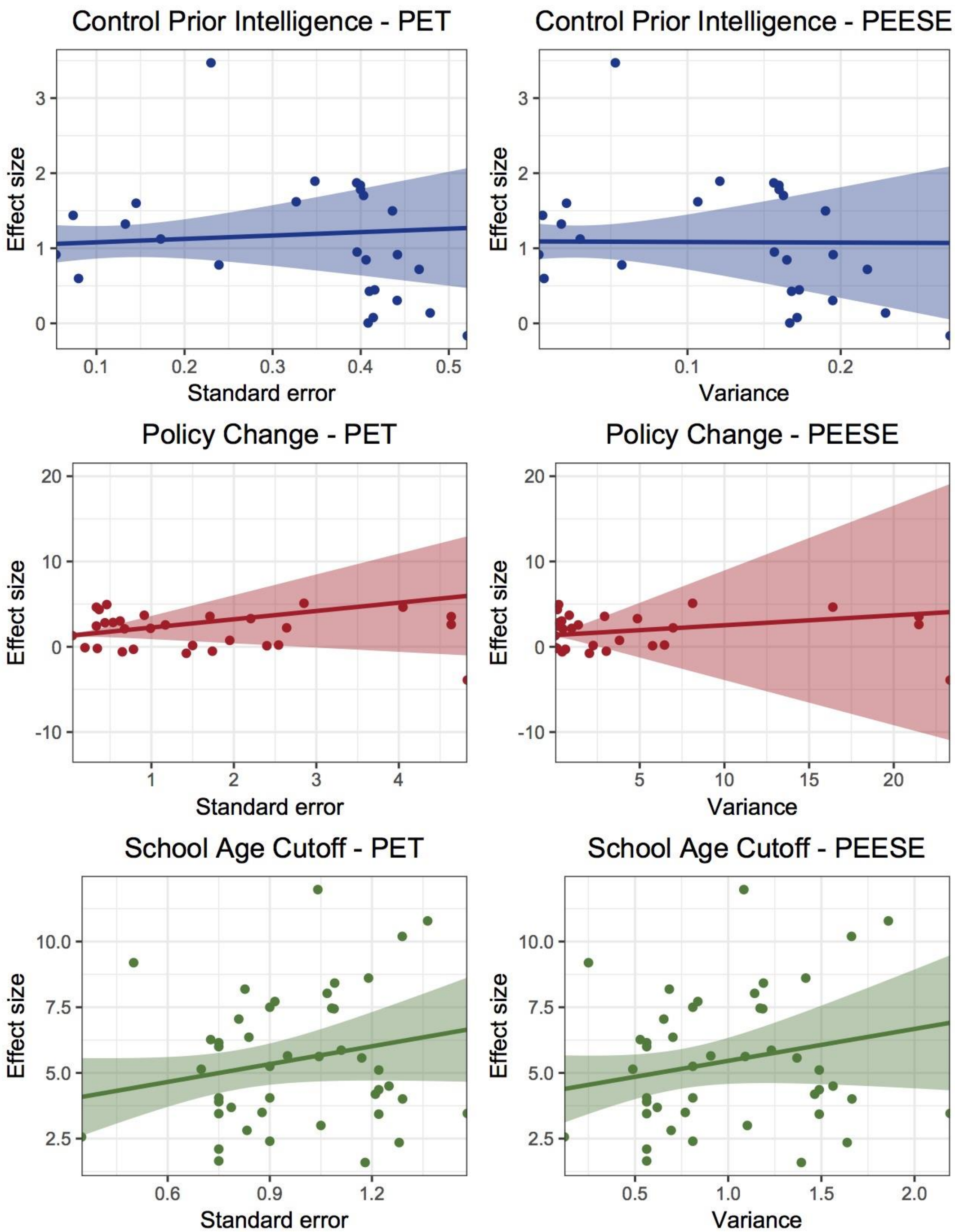


9. *Figure S3.* Funnel plots including estimates from all three study designs (raw and residualised for moderators)



Note: For the residualised effect sizes, estimates were residualised for the moderators described for that design in Table 2 in the main document (that is, there was a different set of moderators for each design).

10. *Figure S4.* PET and PEESE regression graphs for each study design



Note. Effect size is always in IQ points (mean = 100, SD = 15) for one additional year of education. The shaded area around each regression line indicates the 95% confidence interval.

11. Supplementary References

**Studies with an asterisk were included in the meta-analysis*

*Artman, L., Cahan, S., & Avni-Babad, D. (2006). Age, schooling and conditional reasoning. *Cognitive Development*, *21*, 131-145.

*Banks, J., & Mazzonna, F. (2012). The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. *The Economic Journal*, *122*, 418-448.

Butler, N. and Bynner, J.M. (2008). *1970 British Cohort Study: Ten-Year Follow-up, 1980* [computer file]. *3rd Edition.* Colchester, Essex: UK Data Archive [distributor]. SN:3723, doi:10.5255/UKDA-SN-3723-1

*Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences*, *109*, 425-430.

Brod, G., Bunge, S. A., & Shing, Y. L. (2017). Does one year of schooling improve children's cognitive control and alter associated brain activation? *Psychological Science*, advance online publication. doi:10.1177/0956797617699838

*Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, *60*, 1239-1249.

*Cahan, S., & Noyman, A. (2001). The Kaufman Ability Battery for Children mental processing scale: a valid measure of "pure" intelligence? *Educational and Psychological Measurement*, *61*, 827-840.

*Carlsson, M., Dahl, G. B., Öckert, B., & Rooth, D. O. (2015). The effect of schooling on cognitive skills. *Review of Economics and Statistics*, *97*, 533-547.

*Cliffordson, C. (2010). Methodological issues in investigations of the relative effects of schooling and age on school performance: The between-grade regression discontinuity design applied to Swedish TIMSS 1995 data. *Educational Research and Evaluation*, *16*, 39-52.

*Clouston, S. A., Kuh, D., Herd, P., Elliott, J., Richards, M., & Hofer, S. M. (2012). Benefits of educational attainment on adult fluid cognition: International evidence from three birth cohorts. *International Journal of Epidemiology*, *41*, 1729-1736.

*Davies, N. M., Dickson, M., Davey Smith, G., van den Berg, G. J., & Windmeijer, F. (2018). The causal effects of education on health outcomes in the UK Biobank. *Nature Human Behaviour*, *2*, 117-125.†

Elliott, J., & Shepherd, P. (2006). Cohort profile: 1970 British birth cohort (BCS70). *International Journal of Epidemiology*, 35, 836–843.

*Falch, T., & Sandgren Massih, S. (2011). The effect of education on cognitive ability. *Economic Inquiry*, *49*, 838-856.

*Gambrell, J. L. (2013). *Effects of age and schooling on 22 ability and achievement tests*. (Unpublished doctoral dissertation). University of Iowa, Iowa.

*Glymour, M. M., Kawachi, I., Jencks, C. S., & Berkman, L. F. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology & Community Health*, *62*, 532-537.

*Gorman, E. (2017). Does schooling have lasting effects on cognitive function? Evidence from compulsory schooling laws. Working paper.

*Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: Intelligence and class structure in American life* (paperback). New York, NY: Free Press.

*Jabr, D., & Cahan, S. (2015). Between-context variability of the effect of schooling on cognitive development: evidence from the Middle East. *School Effectiveness and School Improvement*, *26*, 441-466.

*Kamhöfer, D. A., & Schmitz, H. (2016). Reanalyzing zero returns to education in Germany. *Journal of Applied Econometrics*, *31*, 912-919.

*Kamhöfer, D. A., Schmitz, H., & Westphal, M. (2016). Heterogeneity in non-monetary returns to higher education. Working paper.

*Kyriakides, L., & Luyten, H. (2009). The contribution of schooling to the cognitive development of secondary education students in Cyprus: An application of regression discontinuity with multiple cut-off points. *School Effectiveness and School Improvement*, *20*, 167-186.

*Lager, A., Seblova, D., Falkstedt, D., & Lövdén, M. (2016). Cognitive and emotional outcomes after prolonged education: a quasi-experiment on 320 182 Swedish boys. *International Journal of Epidemiology*, *46*, 303-311.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*, 281-355.

*Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95. *Oxford Review of Education*, *32*, 397-429.

*Luyten, H., Peschar, J., & Coe, R. (2008). Effects of schooling on reading performance, reading engagement, and reading activities of 15-year-olds in England. *American Educational Research Journal*, *45*, 319-342.

Mayer, S. E., & Knutson, D. (1999). Does the timing of school affect how much children learn? In S. E. Mayer & P. E. Peterson (Eds.), *Earning and Learning: How School Matters* (pp. 79-102). New York, NY: Russell Sage Foundation.

Muthén L. K., & Muthén B. O. (1998–2014). *Mplus user's guide: The omprehensive modeling program for applied researchers.* Los Angeles, CA: Muthén & Muthén.

*Plassman, B. L., Welsh, K. A., Helms, M., Brandt, J., Page, W. F., & Breitner, J. C. S. (1995). Intelligence and education as predictors of cognitive state in late life: A 50-year follow-up. *Neurology*, *45*, 1446-1450.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

*Ritchie, S. J., Bates, T. C., Der, G., Starr, J. M., & Deary, I. J. (2013). Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. *Psychology and Aging*, *28*(2), 515-521.

*Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology*, *51*, 573-582.

*Schneeweis, N., Skirbekk, V., & Winter-Ebmer, R. (2014). Does education improve cognitive performance four decades after school completion? *Demography*, *51*, 619-643.

*Tonkin, M. C. (1999). *The comparative effects of education and the complexity of work on adult intellectual ability*. (Unpublished doctoral dissertation). University of Akron, OH.

University of London, Institute of Education, Centre for Longitudinal Studies (2008). *1970 British Cohort Study: Thirty-Four-Year Follow-up, 2004-2005* [computer file]. *2nd Edition.* Colchester, Essex: UK Data Archive [distributor]. SN: 5585, doi:10.5255/UKDA-SN-5585-1

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1-48.

*Wang, T., Ren, X., Schweizer, K., & Xu, F. (2016). Schooling effects on intelligence development: evidence based on national samples from urban and rural China. *Educational Psychology*, *36*(5), 831-844.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

*Xiao, Y., Li, L., & Zhao, L. (2017). Education on the cheap: The long-run effects of a free compulsory education reform in rural china. Working paper, accessed from: http://conference.iza.org/conference_files/ICID_Renmin_2016/zhao_l9564.pdf

†Note that the preprint version of the study by Davies et al. (2018) was used in our original analysis (preprint version available at: https://www.biorxiv.org/content/early/2016/09/13/074815); this reference was updated when the study was published as we cited it in the Discussion section of the main paper