Research Article

Psychological Science
2018, Vol. 29(8) 1358–1369

DOI: 10.1177/0956797618774253
www.psychologicalscience.org/PS

How Much Does Education Improve Intelligence? A Meta-Analysis

Stuart J. Ritchie[1,2] and Elliot M. Tucker-Drob[3,4]
[1]Department of Psychology, The University of Edinburgh; [2]Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh; [3]Department of Psychology, University of Texas at Austin; and [4]Population Research Center, University of Texas at Austin

Abstract

Intelligence test scores and educational duration are positively correlated. This correlation could be interpreted in two ways: Students with greater propensity for intelligence go on to complete more education, or a longer education increases intelligence. We meta-analyzed three categories of quasiexperimental studies of educational effects on intelligence: those estimating education-intelligence associations after controlling for earlier intelligence, those using compulsory schooling policy changes as instrumental variables, and those using regression-discontinuity designs on school-entry age cutoffs. Across 142 effect sizes from 42 data sets involving over 600,000 participants, we found consistent evidence for beneficial effects of education on cognitive abilities of approximately 1 to 5 IQ points for an additional year of education. Moderator analyses indicated that the effects persisted across the life span and were present on all broad categories of cognitive ability studied. Education appears to be the most consistent, robust, and durable method yet to be identified for raising intelligence.

Keywords

intelligence, education, meta-analysis, quasiexperimental, open data

Received 11/9/17; Revision accepted 3/8/18

There is considerable interest in environmental factors that might improve the cognitive skills measured by intelligence tests, and for good reason: These skills are linked not just to higher educational attainment but to superior performance at work (Kuncel & Hezlett, 2010; Schmidt, Oh, & Shaffer, 2016), better physical and mental health (Gale et al., 2012; Wrulich et al., 2014), and greater longevity (Calvin et al., 2017). The current meta-analysis focused on a potential intelligence-boosting factor that is routinely experienced by children and young adults throughout the world: education. We addressed the question of whether increases in normal-range educational duration after early childhood have positive effects on a student's later intelligence.

On its face, the positive correlation between intelligence test scores and years of completed education (Strenze, 2007) might suggest that the experience of prolonged education has a beneficial effect on intelligence. However, the association could also result from a selection process, whereby more intelligent children progress further in education (Deary & Johnson, 2010). Indeed, there is ample evidence that pervasive selection processes operate in the intelligence-education association: Longitudinal studies demonstrate the predictive power of early intelligence test scores for later educational attainment (Deary, Strand, Smith, & Fernandes, 2007; Roth et al., 2015). The existence of selection processes does not necessarily gainsay any causal effects of education, but it does create an endogeneity problem that renders causal hypotheses difficult to test in observational data. In recent years, however, researchers have increasingly capitalized on a number of sophisticated study designs that circumvent the endogeneity problem, testing the causal hypothesis that more education leads to higher intelligence. This unique class of studies serves as the basis for the current meta-analysis.

Corresponding Author:
Stuart J. Ritchie, Department of Psychology, The University of Edinburgh, Edinburgh, EH8 9JZ, United Kingdom
E-mail: stuart.ritchie@ed.ac.uk

In a seminal review of the effects of educational duration on intelligence, Ceci (1991) adduced evidence from a wide variety of research designs, including studies of intermittent school attendance, studies of the "summer slide" (the drop in children's cognitive performance during summer vacation), and studies using *regression-discontinuity* methods to separate schooling effects from age effects. Ceci's conclusion was that "schooling emerges as an extremely important source of variance" in intelligence test scores (p. 719). However, this and several newer reviews (Deary & Johnson, 2010; Gustaffson, 2001; Snow, 1996; Winship & Korenman, 1997) are all exclusively narrative. In recent years, several high-quality studies investigating educational effects on intelligence have been published, but there continues to be no overall quantitative synthesis of this work. We report the first such synthesis.

We analyzed results from the three most prominent quasiexperimental methods for testing the effects of education on intelligence. We defined intelligence as the score on a cognitive test; see below for consideration of how the test scores might relate to the underlying psychological processes. Each method implements a different approach to minimize effects stemming from selection processes. Full meta-analytic inclusion criteria are reported below, but first we describe each of the three designs, providing a canonical example of each.

The first research design, which we label *control prior intelligence*, is a longitudinal study in which cognitive testing data are collected before and after variation in the duration of education. This allows the relation between education and the second test to be adjusted for by each participant's earlier ability level. An example of this design is the study by Clouston et al. (2012), who analyzed data from three large U.S. and UK cohort studies, all of which had both an adolescent and a midlife cognitive test. Results indicated that completing a university education was linked to higher midlife cognitive ability, above and beyond adolescent intelligence.

The second design, *policy change*, relies on changes in educational duration that are, by all accounts, exogenous to the characteristics of the individuals. An example of this design is the study by Brinch and Galloway (2012), who used large-scale data from a 1960s educational reform in Norway. This reform increased compulsory education by 2 years; critically, it was staggered across municipalities in the country. This allowed the researchers to estimate the effect of an additional year of school on a later intelligence test, taken by males at entry to military service as part of Norway's universal military draft. Under the assumption that the policy change affected intelligence only via increasing years of schooling, the authors used an instrumental-variables analysis to estimate the effect of 1 year of schooling on intelligence at approximately 3.7 points on a standard IQ scale (M = 100, SD = 15 in the population).

The third design takes advantage of a *school-age cutoff*. These studies use regression-discontinuity analysis to leverage the fact that school districts implement a date-of-birth cutoff for school entry. The first study to use this method was by Baltes and Reinert (1969), but the most highly cited example is by Cahan and Cohen (1989), who, in a sample of over 12,000 children across three grades of the Israeli school system (between the ages of approximately 10–12 years), found that schooling exerted positive effects on all of 12 tests covering a variety of cognitive domains. These educational effects were around twice the effect of a year of age. The strict assumptions of this method are sometimes not fully met (Cliffordson, 2010); methodological issues are discussed in more detail below.

After synthesizing the evidence within and across these three research designs, we addressed two further questions. First, which factors moderate the effect of education on intelligence? Perhaps most important, we examined the moderator of age at the outcome test, thus asking whether any educational effects are subject to decline or "fadeout" with increasing age. Second, to what extent is there publication bias in this literature, such that the meta-analytic effects might be biased by a disproportionate number of positive results?

Method

Inclusion criteria, literature search, and quality control

We included data from published articles as well as books, preprint articles, working papers, dissertations, and theses, as long as they met the meta-analytic inclusion criteria. The criteria were as follows. First, the outcome cognitive measures had to be objective (not, e.g., subjective teacher ratings) and continuous (not, e.g., categorical indicators such as the presence of mild cognitive impairment). Second, variation in education had to be after age 6 (i.e., the meta-analysis was not focused on interventions such as preschool but instead on variation later in the educational process). Third, the population under study had to be generally healthy and neurotypical. We thus did not include studies that focused specifically on samples of patients with dementia, individuals with neurodevelopmental disorders, or other such selected groups.

Fourth, studies had to fit into one of the three study design types described above. That is, they had to (a) use earlier cognitive test scores as a control variable in a model predicting cognitive test scores after some

variation in educational duration (control prior intelligence), (b) use data from a natural experiment that specifically affected educational duration prior to the outcome cognitive test or tests (policy change), or (c) use a regression-discontinuity design to analyze cognitive test scores from individuals born on either side of a cutoff date for school entry (school-age cutoff).

We began by searching Google Scholar for articles that had cited Ceci's (1991) review on the effects of education on intelligence, and then searching through the references within each of those studies. Next, we ran searches of APA PsycINFO, Google Scholar, and the ProQuest Dissertations and Theses online database, using search terms related to the three study design types in our inclusion criteria. These terms included combinations of general terms related to the broad topic—"intelligence," "cognitive ability," "cognition," "mental ability," "IQ," "achievement," "ability," "reasoning," "fluid intelligence," "general intelligence," "education," "educational," "school," "schooling"—with specific terms related to the study designs, such as "return to education/school," "influence/effect of education/school," "regression discontinuity," "instrumental variables," "two-stage least squares," "difference-in-difference," "natural experiment," and "quasi-experiment." Having selected the relevant studies from these searches and removed duplicates, we then searched the references within each report to find any additional studies of interest. Finally, we e-mailed authors of multiple studies to request any unpublished preprint articles or working papers that we had not already found. A flow diagram of the overall literature search process is shown in Figure S1 in the Supplemental Material available online.

After arriving at a set of studies that fit the inclusion criteria, we closely examined each report and removed any studies that we deemed did not fit our quality criterion. No studies were excluded for the control-prior-intelligence design. One study was excluded for the policy-change design as we judged it to have a potentially confounded instrument. See Table S1 in the Supplemental Material for a brief description of the design of each included policy-change study, along with other relevant details. For the school-age-cutoff design, we excluded five studies because they did not explicitly report dealing with threats to the validity of the regression-discontinuity analysis related to selection or noncompliance with the cutoff age. We detail the inclusion criteria and quality control for the school-age-cutoff design in the Supplemental Material.

We also produced one new analysis for inclusion in the meta-analysis, using data from a large longitudinal study to which we had access (the British Cohort Study; Elliot & Shepherd, 2006), where the critical control-prior-intelligence analysis had not—to our knowledge—previously been performed. Full details of this analysis are available in the Supplemental Material.

When multiple results were available for a single data set, we coded all relevant cognitive outcomes. However, where multiple estimates from different analyses of the same cognitive outcomes within a data set were available, we used the following criteria to select the estimate for meta-analysis. First, for articles using an instrumental-variables approach in which an alternative ordinary least squares regression analysis was also available, we always took the estimates from the instrumental-variables analysis (although we also recorded the ordinary least squares regression estimates in our data spreadsheet). Second, to reduce heterogeneity due to between-study differences in what covariates were included, we took the analysis that adjusted for the fewest number of covariates. Of the effect sizes that remained after fulfilling the first two criteria, we took the estimate that involved the largest sample size. The precise sources (table, section, or paragraph) for each estimate are described in notes in the master data spreadsheet, available on the Open Science Framework page for this study (https://osf.io/r8a24/). Note that for two of the studies (Ritchie et al., 2013; Ritchie, Bates, & Deary, 2015), we had the data from the cohorts available and recalculated the estimates to remove one of the covariates (see the data spreadsheet). For comparison, we also provide an estimate where maximal covariates were included. A full list of all studies included in the final meta-analysis is shown in Table S4 in the Supplemental Material.

Statistical analysis

Calculating effect sizes. We rescaled each effect size into the number of IQ point units, on the standard IQ scale (M = 100, SD = 15), associated with 1 additional year of education. We also made the corresponding correction to the standard error associated with each rescaled effect size. For example, we multiplied z-scored per-year effect sizes by 15, and we divided unstandardized per-year effect sizes by the associated standard deviation of the cognitive test before multiplying them by 15 (effect-size calculations are described in the master data spreadsheet). For two studies, we recalculated the effect size using structural equation modeling of the correlation matrix provided in the report (see the Supplemental Material). Where effect-size recalculation was not possible from the data provided in the original reports—for example, because of missing standard errors or the effect size being in units other than years of education—we contacted the authors to request further information.

Meta-analytic structural equation models. To produce the main estimates, we used random-effects meta-analytic structural equation modeling, as described by Cheung (2008). This approach is mathematically equivalent to conventional random-effects meta-analytic approaches but has the added advantage of being able to capitalize on special features of structural equation modeling software, such as the correction of standard errors for nonindependence of observations.

Many studies reported effect sizes for more than one cognitive test outcome. For instance, they might have reported effects on a test of memory and a test of executive function. Instead of producing a per-study average of these estimates, we included them all individually, weighting each estimate by the reciprocal of the number of effect sizes provided from each study. In addition, using the TYPE = COMPLEX and the CLUSTER functions in Mplus (Version 7.3; Muthén & Muthén, 2014), we employed a sandwich estimator to correct standard errors for dependencies associated with the clustering of effect sizes within studies.

Moderators. We used the tau (τ) statistic, an estimate of the standard deviation of the true meta-analytic effect, as an index of heterogeneity. To attempt to explain any heterogeneity, we tested a somewhat different set of moderators for each of the three study designs, as appropriate given their methodological differences. For all three designs, we tested the moderators of the age at the outcome test and the outcome test category (classified in two different ways, as described below). For both the control-prior-intelligence and the policy-change designs, we tested the moderators of participant age at the early (control) test or at the policy change (for the control-prior-intelligence design, we also tested the moderator of the gap between the two tests, though this was heavily related to the age at outcome test) and of whether the study was male only or mixed sex (several studies in these designs relied on military draft data and were thus restricted to male participants; note that this variable is confounded with the representativeness of the study because, aside from their single-sex nature, military draft studies will tend to include a more representative sample of the population than others). Where we combined all three study designs, we tested whether design was a moderator.

We classified outcome tests in two ways. The first was into the broad intelligence subtype: fluid tests (tests that assessed skills such as reasoning, memory, processing speed, and other tasks that could be completed without outside knowledge from the world), crystallized tests (tests that assessed skills such as vocabulary and general knowledge), and composite tests (tests that assessed a mixture of fluid and crystallized skills; in one instance, this composite was formally estimated as a latent factor with fluid and crystallized indicators). The second classification method was to highlight tests that might be considered achievement measures. To do this, we classified every test that would likely have involved content that was directly taught at school (including reading, arithmetic, and science tests) as "achievement," and the remaining tests, which generally involved IQ-type measures (ranging from processing speed to reasoning to vocabulary), as "other" tests.

Publication-bias tests. We used four separate methods to assess the degree of publication bias in the data set. First, we tested whether the effect sizes were larger in peer-reviewed, published studies versus unpublished studies (for example, PhD dissertations or non-peer-reviewed books). If unpublished studies have significantly smaller effect sizes, this may indicate publication bias.

Second, we produced funnel plots, visually inspecting them and testing their symmetry using Egger's test. Significant funnel plot asymmetry (where, for example, low-precision studies with small effects were systematically missing) was taken as a potential indication of publication bias in the data.

Third, we used *p*-curve (Simonsohn, Simmons, & Nelson, 2015) to assess the evidential value of the data set using just the significant *p* values. A left-skewed *p*-curve (with more *p* values near the alpha level, in this case .05) indicates possible publication bias or so-called *p*-hacking (use of questionable research practices, such as the ad hoc exclusion of participants or inclusion of covariates, in order to turn a nonsignificant result into a significant one) in the data set. Conversely, a right-skewed *p*-curve indicates evidential value. The shape of the curve is tested using both a binomial test (for the proportion of values where $p < .025$) and a continuous test, which produces "*pp* values" (the probability of finding a *p* value as extreme as or more extreme than the observed *p* value under the null hypothesis), and combines them to produce a *z* score using Stouffer's method. We used the online *p*-curve app (http://www.p-curve.com/) to compute the analyses.

Fourth, we used the Precision Effect Test–Precision Effect Estimate with Standard Errors technique (PET-PEESE; Stanley & Doucouliagos, 2014). The method first uses a weighted metaregression of the effect sizes on the standard errors, using the intercept of this regression—which estimates a hypothetical "perfect" study with full precision, and thus a standard error of zero—as the corrected "true" meta-analytic estimate (called the PET estimate). However, Stanley and Doucouliagos (2014) advised that, where the PET estimate was significantly different from zero, a less biased estimate can be produced by using the variance instead of the standard

errors. The intercept of this regression is the PEESE estimate. We followed this conditional logic in our PET-PEESE analysis. References for all of the analysis software are provided in the Supplemental Material.

Results

The selection process resulted in a final meta-analytic data set including 142 effect sizes from 42 data sets, analyzed in 28 studies. The total sample size across all three designs was 615,812. See Table 1 for a breakdown of study characteristics by study design. Figure S2 shows forest plots for each design.

Overall meta-analytic estimates

In three separate unconditional random-effects meta-analytic models (one for each study design), we estimated the effect of 1 additional year of education on cognitive outcomes. For all three study designs, there was a significant effect of 1 additional year of education. For control prior intelligence, the effect was 1.197 IQ points ($SE = 0.203$, $p = 3.84\times10^{-09}$); for policy change, it was 2.056 IQ points ($SE = 0.583$, $p = 4.23\times10^{-04}$); and for school-age cutoff, it was 5.229 IQ points ($SE = 0.530$, $p = 6.33\times10^{-23}$). An overall model including all estimates from all three designs found an average effect size of 3.394 IQ points for 1 year of education ($SE = 0.503$, $p = 1.55\times10^{-11}$).

The overall model, considering all study designs simultaneously and including study design as a nominal moderator variable, found that the estimate for school-age cutoff was significantly larger than that for control prior intelligence ($SE = 0.564$, $p = 1.98\times10^{-13}$) and for policy change ($SE = 0.790$, $p = 5.34\times10^{-05}$). There was no significant difference between the estimates for control prior intelligence and policy change ($SE = 0.608$, $p = .116$).

The estimates above had minimal covariates included; for 27 of the 142 effect sizes, it was possible to extract an estimate that included a larger number of covariates (see data spreadsheet). This maximal-covariate analysis yielded reduced, though similar and still significant, effect-size estimates for the control-prior-intelligence design (0.903 IQ points, $SE = 0.372$, $p = .015$), and for the quasiexperimental design (1.852 IQ points, $SE = 0.508$, $p = 2.71\times10^{-04}$). There were no additional covariates to include for the school-age-cutoff design.

Heterogeneity and moderator analyses

There was significant heterogeneity in the unconditional meta-analyses from all three designs (control prior intelligence: $\tau = 0.721$, $SE = 0.250$, $p = .004$; policy change: $\tau = 1.552$, $SE = 0.144$, $p = 3.40\times10^{-27}$; school-age cutoff: $\tau = 1.896$, $SE = 0.226$, $p = 5.38\times10^{-17}$). This was also the case for the overall model including all the data, which included study design as a nominal moderator ($\tau = 2.353$, $SE = 0.272$, $p = 5.72\times10^{-18}$). We explored which moderators might explain the heterogeneity within each of the three study designs. Descriptive statistics for each moderator are shown in Table 1.

Age at early test and time lag between tests. For the control-prior-intelligence design, we tested whether the age at which the participants had taken the initial (control) cognitive test, or the gap between this early test and the outcome test, moderated the effect size. The age at the early test, which did not vary substantially (see Table

Table 1. Descriptive Statistics for Each Study Design

Design	Control prior intelligence	Policy change	School age cutoff
k studies	7	11	10
k data sets	10	12	20
k effect sizes	26	30	86
N participants	51,645	456,963	107,204
Mean age at early test in years (*SD*)	12.35 (2.90)	—	—
Mean time lag between tests in years (*SD*)	53.17 (15.47)	—	—
Mean age at policy change in years (*SD*)	—	14.80 (2.59)	—
Mean age at outcome test in years (*SD*)	63.48 (18.80)	47.92 (19.39)	10.36 (1.60)
n outcome test category (composite:fluid:crystallized)	5:20:1	2:23:5	3:67:16
n achievement tests (achievement:other)	1:25	7:23	38:48
Male-only estimates (male only:mixed sex)	2:24	8:22	0:86
Publication status (published:unpublished)	22:4	21:9	64:22

Note: To estimate *N* from studies with multiple effect sizes with different *n*s, we averaged sample sizes across effect sizes within each data set and rounded to the nearest integer. "Unpublished" refers to any study not published in a peer-reviewed journal.

1), was not significantly related to the effect size (−0.024 IQ points per year, $SE = 0.064$, $p = .706$). For the youngest early-test age (10 years), the metaregression model indicated that the effect size of 1 additional year of education was 1.243 IQ points; for the oldest (16 years), the effect size was 1.099 IQ points. Conversely, the time lag between the tests was a significant moderator of the effect size (−0.031 IQ points per year, $SE = 0.015$, $p = .033$). This metaregression indicated that at the smallest age gap (5 years), the effect size was 2.398 IQ points, whereas for the largest age gap (72.44 years), the effect size was a substantially smaller 0.317 IQ points. Note that this age gap is almost fully confounded with the age at the outcome test ($r = .988$), assessed as a moderator below.

Age at intervention. For the policy-change design, we tested whether the age at which the educational policy change produced an increment in compulsory schooling moderated the intervention effect. This was not the case: The effect size increased by a nonsignificant 0.038 IQ points per year of age at the intervention ($SE = 0.228$, $p = .867$). The metaregression model implied that at the youngest intervention age (7.5 years), the effect size was 1.765 IQ points, and at the oldest (19 years) it was 2.204 IQ points.

Age at outcome test. Figure 1 shows the effect sizes in the first two study designs as a function of the participants' mean age at the outcome test. For the control-prior-intelligence studies, outcome age was a significant moderator: The effect size of education declined by −0.026 IQ points per year of age ($SE = 0.012$, $p = .029$). At the youngest age (18 years) the effect size of having had an additional year of education was 2.154 IQ points, whereas at the oldest age (83 years) the effect size was 0.485 IQ points. This effect was smaller but still significant

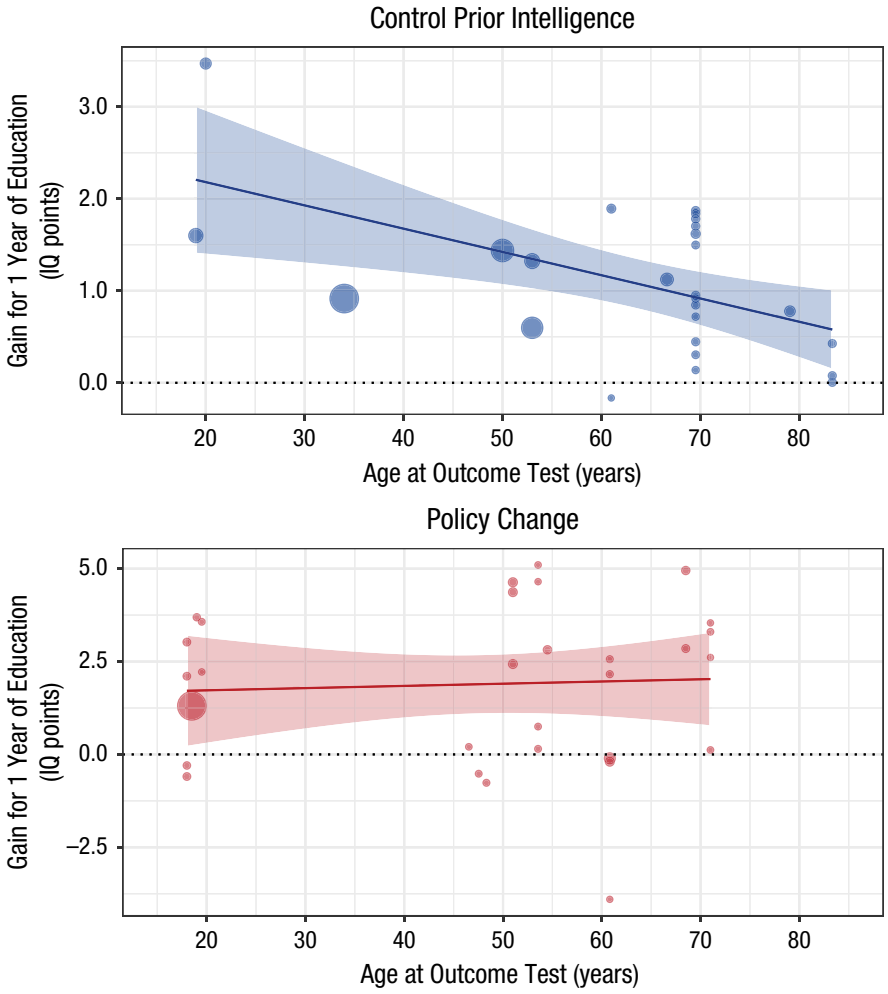

Fig. 1. Effect of 1 additional year of education as a function of age at the outcome test, separately for control-prior-intelligence and policy-change study designs. Bubble size is proportional to the inverse variance for each estimate (larger bubbles = more precise studies). Estimates in these illustrations differ slightly from the final metaregression estimate, which accounted for clustering. The shaded area around the regression line represents the 95% confidence interval.

if the largest effect size (> 3 IQ points for an additional year) was excluded (–0.011 IQ points per year of age, $SE = 0.005$, $p = .018$). There was no significant moderating effect of age at outcome for the policy-change studies (0.014 IQ points per year, $SE = 0.022$, $p = .543$): The effect at the youngest age (18 years; 1.690 IQ points) was not significantly different from the effect at the oldest age (71 years; 2.413 IQ points). There was comparatively little variation in the age at the outcome test for the school-age-cutoff design ($SD = 1.6$ years; see Table 1); there was no significant age moderation effect (–0.027 points per year, $SE = 0.399$, $p = .947$).

Outcome test category. Splitting the outcome cognitive tests into three broad categories—composite, fluid, and crystallized tests—we tested whether the category moderated effect sizes. For the control-prior-intelligence design, there were stronger educational impacts on composite tests (1.876 IQ points, $SE = 0.467$, $p = 5.94\times10^{-05}$) than on fluid tests (0.836 points, $SE = 0.097$, $p = .152$), and the difference between composite and fluid was significant (1.039 points, $SE = 0.496$, $p = .036$). There was only one crystallized outcome test for the control-prior-intelligence design (1.893 points, $SE = 0.348$, $p = 5.34\times10^{-08}$), so we did not include it in the moderator comparison here. For the policy-change design, there were significant effects for both composite (2.314 IQ points, $SE = 0.869$, $p = .008$) and fluid (2.272 points, $SE = 0.765$, $p = .003$) but not crystallized (1.012 points, $SE = 1.125$, $p = .368$) tests; however, the effects on the three different categories were not significantly different from one another (all difference p values > .35). Finally, for the school-age-cutoff design, there were significant effects of a year of education on composite (6.534 points, $SE = 2.433$, $p = .007$), fluid (5.104 points, $SE = 0.621$, $p = 2.05\times10^{-16}$), and crystallized (5.428 points, $SE = 0.170$, $p = 1.04\times10^{-223}$) tests; there were, however, no significant differences between effect sizes across outcome types (difference p values > .5).

We then split the outcome tests into "achievement" tests versus "other" tests. There was only one achievement test in the control-prior-intelligence design, so we did not run this analysis. For policy change, there was no significant difference in the educational effect on the 7 achievement tests (2.760 IQ points, $SE = 0.968$, $p = .004$) versus the 23 other tests (1.784 points, $SE = 0.553$, $p = .001$; difference $SE = 1.011$, $p = .334$). However, for school-age cutoff, which had the largest proportion of achievement tests (38 of the 86 tests were classed as achievement tests), achievement tests showed a substantially and significantly larger educational effect (6.231 points, $SE = 0.339$, $p = 2.85\times10^{-75}$) than other tests (3.839 points, $SE = 0.412$, $p = 1.11\times10^{-20}$; difference $SE = 0.371$, $p = 1.19\times10^{-10}$).

Male-only studies. We tested whether studies that included only male participants showed a differential educational effect. This was not the case for control prior intelligence (effect for the 2 male-only estimates: 2.261 IQ points, $SE = 0.897$, $p = .012$; effect for 24 mixed-sex estimates: 1.027 IQ points, $SE = 0.110$, $p = 6.79\times10^{-21}$; difference $SE = 0.905$, $p = .173$), or for policy change (effect for the 8 male-only estimates: 1.683 IQ points, $SE = 0.507$, $p = .001$; effect for 22 mixed-sex estimates: 2.215 IQ points, $SE = 0.788$, $p = .005$; difference $SE = 0.941$, $p = .572$). There were no male-only school-age-cutoff studies.

Multiple moderators. Table 2 shows the results from each study design after the inclusion of multiple moderators. We included as many moderators as possible for each design, though we chose to include the "achievement" categorization of the outcome test for the school-age-cutoff design (instead of the alternative categorization involving composite, fluid, and crystallized tests, which we used

Table 2. Simultaneous Multiple-Moderator Analyses for Each Study Design

Design	Control prior intelligence	Policy change	School age cutoff
Intercept	2.079 (0.665)**	–1.068 (4.429)	2.603 (4.426)
Early test age	–0.011 (0.026)	—	—
Age at intervention	—	0.137 (0.220)	—
Outcome test age	–0.020 (0.005)***	0.023 (0.048)	0.111 (0.416)
Outcome test category (composite vs. fluid/crystallized)	–0.689 (0.234)***	–1.479 (1.045)	—
Outcome test category (achievement vs. other)	—	—	2.471 (0.524)***
Male only	0.874 (0.258)**	–0.644 (1.766)	—
τ	0.305 (0.056)***	1.454 (0.206)***	1.740 (0.310)***
Change in τ	0.416	0.098	0.156

Note: Values are estimates (in IQ point units); standard errors are in parentheses. The change in the τ statistic refers to that from the unconditional models, as reported in the row above.
**$p < .01$. ***$p < .001$.

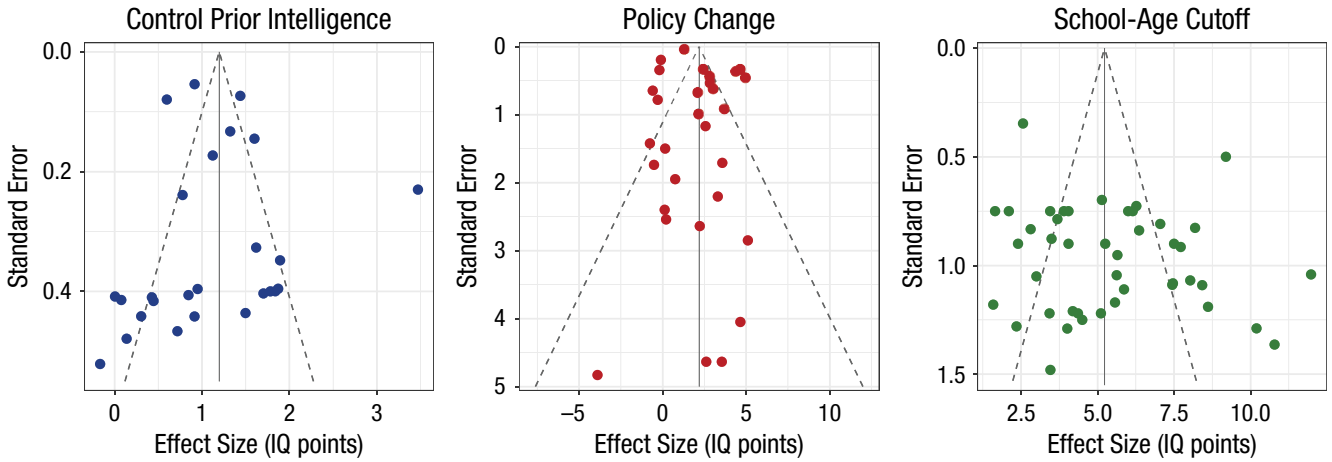

Fig. 2. Funnel plots showing standard error as a function of effect size, separately for each of the three study designs. The dotted lines form a triangular region (with a central vertical line showing the mean effect size) where 95% of estimates should lie in the case of zero within-group heterogeneity in population effect sizes. Note that 42 of the total 86 standard errors reported as approximate or as averages in the original studies were not included for the school-age-cutoff design.

for the other designs) because it had such a substantial effect in the single-moderator model. Including multiple moderators reduced the τ statistic—to a larger degree for the control-prior-intelligence design than for the others—though significant heterogeneity remained in all cases. The moderators that were individually significant (e.g., age in the control-prior-intelligence design or achievement tests in the school-age-cutoff design) were also significant in the multiple-moderator model, indicating that their effects were incremental of the other moderators that we included.

Publication-bias tests

Publication status. As an initial test of publication bias, we tested whether the effect sizes were larger in studies published in peer-reviewed journals compared with those that were either unpublished or published elsewhere. For control prior intelligence, there were 4 published versus 22 unpublished estimates; there was no significant difference in their effect sizes (the effect was 0.036 points larger in published studies, $SE = 0.343$, $p = .915$). For policy-change studies, for which there were 21 published and 9 unpublished estimates, the effect size was significantly larger in unpublished studies, though there were still significant effects within each set of studies (published effect = 1.635 points, $SE = 0.575$, $p = .004$; unpublished effect = 3.469 points, $SE = 0.388$, $p = 4.11\times10^{-19}$; difference $SE = 0.710$, $p = .010$). For school-age-cutoff studies, there was no significant difference between published and unpublished studies (for which there were 64 and 22 estimates, respectively; difference = 0.509 points higher in published studies, $SE = 0.689$, $p = .460$).

Funnel plots. Funnel plots for the three study designs are shown in Figure 2. Note that, for the school-age-cutoff design, 42 of the 86 standard errors were reported as approximate or as averages; because they were inexact, we used them in the estimates of the meta-analytic effects above but did not use them to estimate the funnel plots (or for the PET-PEESE analysis below). Egger's test found no evidence of funnel plot asymmetry for any of the designs (control prior intelligence: $z = -1.378$, $p = .168$; policy change: $z = -0.486$, $p = .627$; school-age cutoff: $z = 0.941$, $p = .347$). However, only the funnel for control prior intelligence displayed an approximately funnel-like shape. See Figure S3 in the Supplemental Material for funnel plots including studies from all three designs, one using the raw effect sizes and using effect sizes residualized for the moderators shown in Table 2.

p-curve. Next, we used *p*-curve to examine the distribution of study *p* values. The *p*-curves are shown in Figure 3. For the control-prior-intelligence design, the binomial test for a right-skewed *p*-curve was significant, indicating evidential value ($p = .0007$); this was also the case for the continuous test (full *p*-curve: $z = -18.50$, $p = 2.06\times10^{-76}$; half *p*-curve: $z = -19.19$, $p = 4.49\times10^{-82}$). For the quasiexperimental design, the binomial test was significant ($p = .0461$), as was the continuous test (full *p*-curve: $z = -15.59$, $p = 8.51\times10^{-55}$; half *p*-curve: $z = -17.64$, $p = 1.21\times10^{-69}$). For school-age-cutoff studies, all three tests were significant (binomial test $p < .0001$; continuous test full *p*-curve: $z = -43.72$, $p \approx .00$; half *p*-curve: $z = -42.61$, $p \approx .00$). For all three designs, *p*-curve estimated that the statistical power of the tests included was 99%. Thus, overall, *p*-curve indicated that all three designs provided evidential value, and there was no evidence for publication

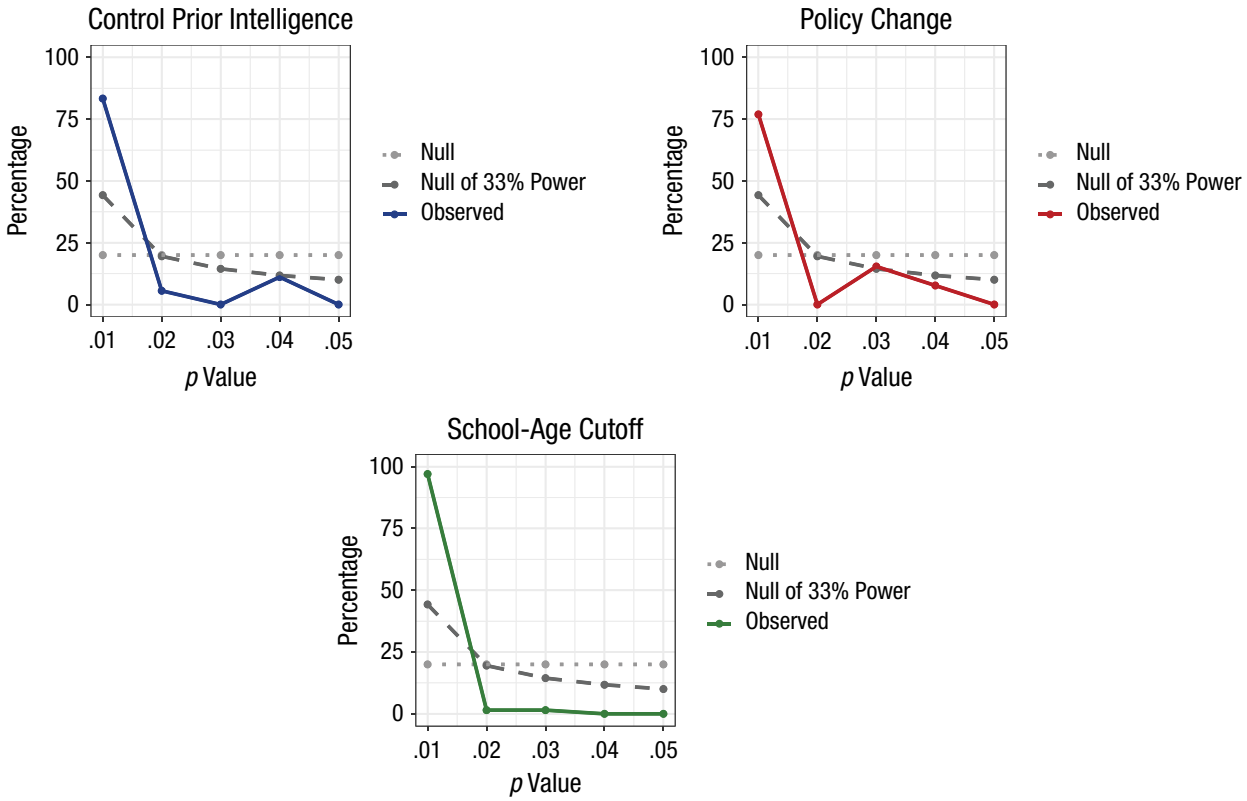

Fig. 3. *p*-curves illustrating the distribution of significant *p* values for each of the three study designs.

bias or *p*-hacking in the studies with statistically significant results (full output from the *p*-curve app is available on the Open Science Framework page for the present study).

PET-PEESE. Finally, we used PET-PEESE to obtain an estimate of the effect size for each study design in a hypothetical study with perfect precision. For all three designs, the PET estimate (intercept) was significant, so we went on to use the PEESE estimate of the intercept, which represents the predicted effect sizes under a counterfactual of no publication bias (control prior intelligence: PET estimate = 1.034 IQ points per year, $SE = 0.153$, $p = 5.45\times10^{-07}$; PEESE estimate = 1.091 points, $SE = 0.117$, $p = 1.76\times10^{-09}$; policy change: PET estimate = 1.286 points, $SE = 0.153$, $p = 3.69\times10^{-09}$; PEESE estimate = 1.371 IQ points, $SE = 0.142$, $p = 2.08\times10^{-10}$; school-age cutoff: PET estimate = 3.299 IQ points, $SE = 1.166$, $p = .007$; PEESE estimate = 4.244 IQ points, $SE = 0.718$, $p = 5.26\times10^{-07}$). Note that only the exact standard errors (i.e., not those reported as approximate or averages, as noted for the funnel plots above) were used for the PET-PEESE analysis. For all three designs, the PEESE test indicated effect sizes that were slightly smaller than in the original estimate but still statistically significant. Graphs of the PET-PEESE estimates for each design are shown in Figure S4 in the Supplemental Material.

Overall, four different publication-bias tests broadly indicated minimal systematic bias in the results: Where there was an unexpected result—unpublished studies producing larger estimates for the policy-change design—this was in the opposite direction to what would be expected under publication bias.

Discussion

In a meta-analysis of three quasiexperimental research designs, we found highly consistent evidence that longer educational duration is associated with increased intelligence test scores. Each of the designs implemented a different approach for limiting endogeneity confounds resulting from selection processes, where individuals with a propensity toward higher intelligence tend to complete more years of education. Thus, the results support the hypothesis that education has a causal effect on intelligence test scores. The effect of 1 additional year of education—contingent on study design, inclusion of moderators, and publication-bias correction—was estimated at approximately 1 to 5 standardized IQ points.

Each research design had its own strengths and weaknesses. The control-prior-intelligence design produced precise, long-range estimates of the educational effect, taking into account the full range of educational

variation. However, this approach did not employ a specific instrument for introducing differences in educational duration, instead capitalizing on naturally occurring variation, which is itself multidetermined. Moreover, because the early and outcome tests were rarely identical (and because the early ability tests likely contained measurement error), the control for preexisting ability levels was likely only partial.

The policy-change design produced causal estimates across large, population-based data sets. However, estimates from this approach were relatively imprecise, as is typical of instrumental-variable analyses. Furthermore, because the policies used as instruments typically increased educational duration only for the subset of individuals who would otherwise have attended school at the preexisting minimum compulsory level, this design should be interpreted as producing a "local average treatment effect" that might not generalize across the full educational range (Morgan & Winship, 2015, p. 305).

The school-age-cutoff design produced the largest number of estimates across a wide range of cognitive abilities, but it was restricted to comparisons across adjacent school years. In this design, the critical causal estimate is based on comparing test scores in a given grade with a counterfactual formed by extrapolating within-grade age trends beyond the cutoff dates. This approach is powerful, but the key assumption—that the age trend extrapolates—is difficult to test. Moreover, although this approach produced large effect-size estimates, we did not identify any studies that tested whether these effects persisted into adulthood. These estimates should thus be regarded with caution.

The finding of educational effects on intelligence raises a number of important questions that we could not fully address with our data. First, are the effects on intelligence additive across multiple years of education? We might expect the marginal cognitive benefits of education to diminish with increasing educational duration, such that the education-intelligence function eventually reaches a plateau. Unfortunately, we are not aware of any studies that have directly addressed this question using a rigorous quasiexperimental method.

Second, are there individual differences in the magnitude of the educational effect? One possibility is the *Matthew effect* (Stanovich, 1986), whereby children at greater initial cognitive (or socioeconomic) advantage benefit more from additional education than those at lower advantage. Another possibility is that education acts as an equalizer, such that children at lower levels of initial advantage benefit most (Downey, von Hippel, & Broh, 2004). Indeed, some evidence of an equalizing effect was reported in a single study by Hansen, Heckman, and Mullen (2004).

Third, why were the effects obtained from the control-prior-intelligence and policy-change designs—which generally came from increases in educational duration that were not explicitly targeted cognitive interventions—still apparent in later life, when effects from targeted educational interventions, such as preschool, have tended to show fade-out into early adulthood (Bailey, Duncan, Odgers, & Yu, 2017; Protzko, 2015)? Even in the control-prior-intelligence design, where the effects showed a decline across time (Fig. 1), estimates remained statistically significant into the eighth and ninth decades of life. One intriguing possibility is that, unlike targeted interventions, increases in educational attainment have lasting influences on a range of downstream social processes, for instance occupational complexity (Kohn & Schooler, 1973) that help to maintain the initial cognitive benefits.

Fourth, which cognitive abilities were impacted? It is important to consider whether specific skills—those described as "malleable but peripheral" by Bailey et al. (2017, p. 15)—or general abilities—such as the general *g* factor of intelligence—have been improved (Jensen, 1989; Protzko, 2016). The vast majority of the studies in our meta-analysis considered specific tests and not a latent *g* factor, so we could not reliably address this question. However, it is of important theoretical and practical interest whether the more superficial test scores or the true underlying cognitive mechanisms are subject to the education effect. In our analyses with test category as a moderator, we generally found educational effects on all broad categories measured (we did observe some differences between the test categories, but it should be noted that differential reliability of the tests might have driven some of these differences). However, further studies are needed to assess educational effects on both specific and general cognitive variables, directly comparing between the two (e.g., Ritchie, Bates, & Deary, 2015).

Fifth, how important are these effects? There is strong evidence from industrial and organizational psychology and cognitive epidemiology studies that IQ is associated with occupational, health, and other outcomes (e.g., Calvin et al., 2017), but to our knowledge, no studies have explicitly tested whether the additional IQ points gained as a result of education themselves go on to improve these outcomes (see Ackerman, 2017, for discussion of this criterion problem in intelligence research). A quasiexperimental study by Davies, Dickson, Davey Smith, van den Berg, and Windmeijer (2018) found that raising the school-leaving age improved not only IQ but also a variety of indicators of health and well-being. It is possible that the educational benefits to the upstream variables were partly mediated via the IQ increases (or vice versa), but this would need explicitly to be investigated.

Finally, what are the underlying psychological mechanisms of the educational effect on intelligence? Ceci (1991) outlined a number of promising pathways,

including the teaching of material directly relevant to the tests, the training of thinking styles such as abstract reasoning, and the instilling of concentration and self-control. Studies that attempt to pin down the proximal educational processes that might, in part, drive the effect (such as reading; Ritchie, Bates, & Plomin, 2015; Stanovich, 1993; though see Watkins & Styck, 2017); those that focus on the differences between the educational effect on specific subtests (e.g., Ritchie et al., 2013), and those that address effects of variation in the quality, not just the quantity, of education (e.g., Allensworth, Moore, Sartain, & de la Torre, 2017; Becker, Lüdtke, Trautwein, Köller, & Baumert, 2012; Gustaffson, 2001) are all promising ways to progress toward clarifying a mechanism.

The results reported here indicate strong, consistent evidence for effects of education on intelligence. Although the effects—on the order of a few IQ points for a year of education—might be considered small, at the societal level they are potentially of great consequence. A crucial next step will be to uncover the mechanisms of these educational effects on intelligence in order to inform educational policy and practice.

Action Editor

Brent W. Roberts served as action editor for this article.

Author Contributions

Both authors developed the study concept. S. J. Ritchie performed the literature search and the initial data coding, then both authors performed quality control on the studies to be included in the meta-analysis and agreed on the final data set. E. M. Tucker-Drob developed the framework for the Mplus syntax and wrote the baseline Mplus scripts (for the meta-analytic models), which were adapted for these analyses by S. J. Ritchie. S. J. Ritchie wrote the R analysis scripts (for the publication-bias tests and figures), which were adapted for the multimoderator figure by E. M. Tucker-Drob. Both authors interpreted the analyses, drafted the manuscript, and approved the final manuscript for submission.

Acknowledgments

We are grateful to Sorel Cahan, Sean Clouston, Neil Davies, James Gambrell, Emma Gorman, Dua Jabr, Daniel Kämhofer, Ben Southwood, and Tengfei Wang for their assistance in finding manuscripts, locating effect-size estimates, and providing additional information on relevant data sets. We thank the Centre for Longitudinal Studies, Institute of Education, and the UK Data Service for the British Cohort Study data. These bodies are not responsible for our analysis or interpretation.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

E. M. Tucker-Drob's contribution to this study was supported by National Institutes of Health (NIH) Research Grant R01HD083613. The Population Research Center at the University of Texas at Austin is supported by NIH Grant R24HD042849.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797618774253

Open Practices

All meta-analytic data and all codebooks and analysis scripts (for Mplus and R) are publicly available at the study's associated page on the Open Science Framework (https://osf.io/r8a24/). These data and scripts are described in the Supplemental Material. The study was not formally preregistered. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797618774253. This article has received the badge for Open Data. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

References

Ackerman, P. L. (2017). Adult intelligence: The construct and the criterion problem. *Perspectives on Psychological Science*, *12*, 987–998.

Allensworth, E. M., Moore, P. T., Sartain, L., & de la Torre, M. (2017). The educational benefits of attending higher performing schools: Evidence from Chicago high schools. *Educational Evaluation and Policy Analysis*, *39*, 175–197.

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*, 7–39.

Baltes, P. B., & Reinert, G. (1969). Cohort effects in cognitive development of children as revealed by cross-sectional sequences. *Developmental Psychology*, *1*, 169–177.

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology*, *104*, 682–699.

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

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

Calvin, C. M., Batty, G. D., Der, G., Brett, C. E., Taylor, A., Pattie, A., . . . Deary, I. J. (2017). Childhood intelligence in relation to major causes of death in 68 year follow-up: Prospective population study. *British Medical Journal, 357*, Article j2708. doi:10.1136/bmj.j2708

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*, 703–722.

Cheung, M. W.-L. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological Methods*, *13*, 182–202.

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.

Deary, I. J., & Johnson, W. (2010). Intelligence and education: Causal perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology*, *39*, 1362–1369.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21.

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? School and non-school sources of inequality in cognitive skills. *American Sociological Review*, *69*, 613–635.

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

Gale, C. R., Batty, G. D., Osborn, D. P., Tynelius, P., Whitley, E., & Rasmussen, F. (2012). Association of mental disorders in early adulthood and later psychiatric hospital admissions and mortality in a cohort study of more than 1 million men. *Archives of General Psychiatry*, *69*, 823–831.

Gustaffson, J.-E. (2001). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. *International Education Journal*, *2*, 166–186.

Hansen, K. T., Heckman, J. J., & Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, *121*, 39–98.

Jensen, A. R. (1989). Raising IQ without increasing *g*? A review of The Milwaukee Project: Preventing mental retardation in children at risk. *Developmental Review*, *9*, 234–258.

Kohn, M. L., & Schooler, C. (1973). Occupational experience and psychological functioning: An assessment of reciprocal effects. *American Sociological Review*, *38*, 97–118.

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, *19*, 339–345.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference* (2nd ed.). Cambridge, England: Cambridge University Press.

Muthén, L. K., & Muthén, B. O. (2014). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence*, *53*, 202–210.

Protzko, J. (2016). Does the raising IQ-raising *g* distinction explain the fadeout effect? *Intelligence*, *56*, 65–71.

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.

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*, 515–521.

Ritchie, S. J., Bates, T. C., & Plomin, R. (2015). Does learning to read improve intelligence? A longitudinal multivariate analysis in identical twins from age 7 to 16. *Child Development*, *86*, 23–36.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137.

Schmidt, F. L., Oh, I.-S., & Shaffer, J. A. (2016). *The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings* (Fox School of Business Research Paper). Retrieved from https://ssrn.com/abstract=2853669

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better *P*-curves: Making *P*-curve analysis more robust to errors, fraud, and ambitious *P*-hacking, a reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, *144*, 1146–1152.

Snow, R. E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law*, *2*, 536–560.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*, 60–78.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407.

Stanovich, K. E. (1993). Does reading make you smarter? Literacy and the development of verbal intelligence. *Advances in Child Development and Behavior*, *24*, 133–180.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*, 401–426.

Watkins, M. W., & Styck, K. M. (2017). A cross-lagged panel analysis of psychometric intelligence and achievement in reading and math. *Journal of Intelligence*, *5*(3), Article 31. doi:10.3390/jintelligence5030031

Winship, C., & Korenman, S. (1997). Does staying in school make you smarter? The effect of education on IQ in the bell curve. In B. Devlin, S. E. Fienberg, D. P. Resnick, & K. Roeder (Eds.), *Intelligence, genes, and success* (pp. 215–234). New York, NY: Springer.

Wrulich, M., Brunner, M., Stadler, G., Schalke, D., Keller, U., & Martin, R. (2014). Forty years on: Childhood intelligence predicts health in middle adulthood. *Health Psychology*, *33*, 292–296.