Does Schooling Improve Cognitive Abilities at Older Ages: Causal Evidence from Nonparametric Bounds

Vikesh Amin[1]; Jere R. Behrman[2]; Jason M. Fletcher[3]; Carlos A. Flores[4]; Alfonso Flores-Lagunes[5]; Hans-Peter Kohler[6]

[1] Department of Economics, Central Michigan University
[2] William R. Kenan, Jr. Professor of Economics and Sociology, University of Pennsylvania
[3] Department of Sociology and La Follette School of Public Affairs, University of Wisconsin-Madison, IZA and NBER
[4] Department of Economics, California Polytechnic State University
[5] Department of Economics, Syracuse University, IZA, and GLO
[6] Fredrick J. Warren Professor of Demography, University of Pennsylvania

Abstract
We revisit the much-investigated relationship between schooling and health, focusing on cognitive abilities at older ages using the Harmonized Cognition Assessment Protocol in the Health & Retirement Study. To address endogeneity concerns, we employ a nonparametric partial identification approach that provides bounds on the population average treatment effect using a monotone instrumental variable together with relatively weak monotonicity assumptions on treatment selection and response. The bounds indicate potentially large effects of increasing schooling from primary to secondary but are also consistent with small and null effects. We find evidence for a causal effect of increasing schooling from secondary to tertiary on cognition. We also replicate findings from the Health & Retirement Study using another sample of older adults from the Midlife in United States Development Study Cognition Project.

JEL Codes: I10, I26, J14.

Key Words: Schooling, Cognition, Bounds, Aging, Partial Identification

1. Introduction

Does schooling have a causal effect on cognition at older ages? This question is important for at least two reasons. *First*, cognitive decline and dementia are major societal issues. An estimated 11% of Americans aged 65 years or older have dementia (Hudomiet et al. 2012), while 19% live with cognitive impairment without dementia (Langa et al. 2017). In the US, dementia cases are projected to increase from 5.2 million in 2019 to 10.2 million by 2050 (GBD 2019 Dementia Forecasting Collaborators 2022). Dementia is a major risk factor for mortality, being associated with an increase in the risk of death by a factor of two (Agüero-Torres et al. 1999). The estimated economic cost of dementia including unpaid care provided by families is $159-$215 billion per year (Hurd et al. 2013). In 2017 the Lancet Commission on Dementia Prevention, Intervention and Care identified schooling attainment as the second most modifiable risk factor for dementia. They estimated that 8% of dementia cases could be avoided by increased levels of basic schooling. An additional grade of schooling has also been associated with a reduction in the risk of dementia by 7% (Xu et al. 2016). *Second*, it is important to understand whether schooling has a causal effect on cognition of aging people for assessing future trends in dementia, as cohorts reaching old age will have higher levels of schooling as a result of schooling expansions in the 20th century, and positive schooling trends are predicted to reduce dementia prevalence.

One mechanism through which schooling can reduce the risk of dementia is by improving cognitive abilities at older ages, since better cognitive performance entails less likelihood that an individual is diagnosed given existing thresholds for classifying individuals with dementia. Schooling might improve cognition at older ages because it is related to mid-and-later life conditions such as health, lifestyle, access to healthcare care, social interactions and networks, income and occupations that are correlated with cognition. Schooling attainment is also a marker for cognitive reserve, which refers to the mind's ability to deal with brain aging. Schooling is hypothesized to transmit skills and knowledge that allow individuals to process tasks that render them less vulnerable to the adverse effects of brain aging. There are also well-established associations between schooling and cognition at older ages. Opdebeek et al. (2015) meta-analyzed 109 studies reporting associations between schooling and cognition in older adults (age 60 years or older) and found that an extra grade of schooling is associated with a 0.04-0.08 standard deviation increase in cognitive performance. This means that an individual with four additional grades of schooling (e.g., a college degree in comparison with high school graduation) would be expected to have an advantage of 0.16-0.32 standard deviations in cognition.

The existing evidence thus suggests that increasing schooling attainment may substantially help reduce the risk and burden of dementia by improving cognitive performance. However, this evidence hinges on whether there is an underlying causal relationship between schooling and cognition. Credibly identifying the causal relationship is difficult due to reverse causality from high cognition in childhood to schooling attainment and confounding from other factors such as genetics, early-life cognition, and family background that are correlated with both schooling and cognition. The literature on causal effects of schooling on cognition at older ages is surprisingly sparse. Existing evidence is largely based on studies that utilize changes in compulsory schooling laws to identify exogenous variation in schooling attainment (e.g., Glymour et al. 2008; Banks and Mazzonna 2012; Schneeweis et al. 2014; Gorman 2017). These studies find protective effects of schooling on immediate and delayed memory, with estimates showing that an extra grade of schooling improves immediate and delayed memory scores by 0.14-0.53 of a standard deviation. However, findings for other cognitive domains are mixed. These estimates also capture the effects only for individuals whose education is causally increased by the compulsory schooling laws—the so-called "compliers" in Angrist et al. (1996)—rather than the effects for the general population. These estimates represent effects of increasing schooling from primary to secondary and are not directly informative about effects of schooling at other parts of the educational distribution (e.g., college graduation). Causal inference based on this strategy relies on the relatively strong assumption that school reforms only affect cognition through their effect on schooling (the exclusion restriction), which may not always hold (Avendano et al. 2020).

We provide new evidence on the causal effect of schooling on cognition at older ages using the Harmonized Cognition Assessment Protocol in the Health & Retirement Study (HRS). Our contribution is to employ a nonparametric partial identification approach (Manski and Pepper 2000), which provides bounds on the causal effect. This approach has several attractive features. *First*, it provides bounds on the population average treatment effect (ATE) as opposed to the average effect for a specific subpopulation. *Second*, it allows for arbitrary correlations between schooling and unobserved factors that can affect cognition. *Third*, it relies on three relatively weak and arguably credible monotonicity assumptions: (1) monotone treatment selection (MTS) that posits that on average individuals with higher schooling attainment have higher latent cognition and (2) monotone treatment response (MTR) that imposes the restriction that more schooling does not worsen cognition. (3) We employ mother's schooling attainment as a monotone instrumental variable (MIV)—a variable that is assumed to have a weakly increasing mean relationship with potential outcomes—to help tighten the bounds under the MTS and MTR assumptions. In our context, the MIV assumption states that individuals with higher-schooled mothers have no lower average latent cognition than those with less-schooled mothers. *Fourth*, this approach allows us to look at dose-response relations between schooling and cognition by providing bounds on the effect of increasing schooling at different parts of the educational distribution (e.g., going from being a high school dropout to a high school graduate, or from being a high school graduate to a college graduate). There may be important effects of obtaining credentials (high school diploma; college degree) on cognition because credentials likely have large effects on mid-life conditions such as income and occupation. Nonlinear effects of schooling have been observed for other health outcomes, including mortality (Montez et al. 2012).

We find that there are potentially large effects of *completing* secondary schooling, with estimated bounds indicating that going from being a high school dropout to a high school graduate *at most* increases immediate and delayed memory by 46% of a standard deviation. However, these bounds are also consistent with small and null effects. We obtain tighter bounds that indicate a statistically significant causal effect of increasing schooling from secondary to tertiary. In particular, transitioning from being a high school to college graduate increases immediate and delayed memory by 9-36% of a standard deviation. These results are robust to accounting for attrition through inverse probability weighting.

To our knowledge, a partial identification approach has not previously been used to conduct inference on the causal effect of schooling on cognition at older ages.[1] Though we do not point-identify the causal effect, we view the nonparametric bounds as providing important new and complementary evidence about the plausible magnitude of the causal effect under relatively weak assumptions. This is in line with the views of Mullahy et al. (2021), who argue that partial identification should be more prevalent in public health and clinical research: rather than focusing on point estimates, base public health recommendations and policies on *ranges* of plausible effects. The approach that we employ is particularly well-suited to nationally representative datasets like the HRS that span several cohorts of individuals, and where natural experiments such as compulsory schooling laws provide inferences only for the small subset of cohorts that are affected by these laws. Finally, we replicate the findings from the HRS in a sample of older adults from the Midlife in United States Development Study Cognition Project. This provides

[1] The Manski and Pepper (2000) framework that we employ has been used, among others, to bound the causal effect of (1) parents' schooling on children's schooling (De Haan 2011), (2) unemployment on mental health (Cyagn-Rehm et al. 2017), (3) English proficiency on labor-market outcomes (Gonzalez 2005), (4) education on social support (Huang et al. 2012), and (5) social activities on cognition (Christelis and Dobrescu 2020).

additional confidence in our results and highlights the value of using partial identification in different datasets to assess external validity.

2. Previous Studies

Two studies have estimated effects of schooling on cognition at older ages using variation in schooling attainment arising from compulsory schooling laws. Glymour et al. (2008) estimated effects of schooling on memory and mental status in the HRS employing a two-sample instrumental variable (IV) approach. In the first stage, they predicted grades of schooling in the 1980 Census 5% sample using compulsory schooling laws between 1907-1961. Predicted grades of schooling were then employed as an independent predictor of cognition in the HRS. Banks and Mazzonna (2012) estimated effects of schooling on memory and executive function in the English Longitudinal Study of Ageing (ELSA) using the 1947 increase in the minimum school leaving age from 14 to 15 within a fuzzy regression discontinuity design (RDD). Both studies found large effects of schooling on memory. An extra grade of schooling was associated with a 0.34 standard deviation increase in memory in the HRS and a 0.50 standard deviation increase in the ELSA. Glymour et al. (2008) found no effect of schooling on mental status. Banks and Mazzonna (2012) found that schooling improved executive functioning for men but not for women.

Other similar studies have used compulsory school leaving laws to estimate effects of schooling on cognition for individuals in their 50s and early 60s. Schneeweis et al. (2014) exploited compulsory schooling laws in the 1950s-1960s across six European countries in the Survey of Health, Ageing and Retirement in Europe (SHARE). Their cognition outcomes were immediate and delayed word recall, verbal fluency, numeracy, and orientation to-date test scores. They found an extra grade of schooling increased immediate memory by 0.14 standard deviations and delayed memory by 0.16 standard deviations. They found no causal effects of schooling on fluency, numeracy and orientation. IV estimates were small in magnitude and frequently were negative. Gorman (2017) estimated the effect of schooling on memory, verbal fluency and numeracy in Understanding Society (a panel study of households in the UK) using the 1972 increase in the minimum school-leaving age from 15 to 16. She found that an additional grade of schooling increased memory scores by 0.53 of a standard deviation, similar in magnitude to the IV estimates based on the 1947 reform in Banks and Mazzonna (2012). She found positive effects of schooling on verbal fluency and numeracy but the IV estimates were imprecise. Lastly, Davies et al. (2018) estimated effects of schooling on health (including fluid intelligence) using the 1972 reform in the UK Biobank (UKB). They found that people who left school at age 16 had a fluid intelligence score of 0.33 standard deviations higher than those who left school at age 15.

IV methods—and the closely related fuzzy RDD—based on compulsory schooling laws provide a strong research design because the policy changes are not caused by individual differences (e.g., family resources, genetics) that are correlated with schooling and cognition. However, causal inferences based on these methods rely on, among others, the exclusion restriction assumption—that school reforms affect cognition only through their effect on schooling—which could be violated in certain contexts. For example, using the 1972 schooling reform in the UK, Avendano et al. (2020) found that education did not improve mental health and that compulsory schooling laws may affect later life mental health through channels other than increased schooling.[2] They argue and provide descriptive evidence that the reform forced young

[2] They estimated effects of education on mental health for individuals in their mid-50s using the Annual Population Survey, Understanding Society, UK Biobank datasets. They found that an extra grade of schooling increased the probability of having a mental health condition, and the probability of having depression/anxiety by about 30% in the Annual Population Survey. Results in the other datasets were imprecise, and signs were not consistent—schooling worsened mental health based on the General Health Questionnaire and SF-12 mental health component in Understanding Society but improved mental health based on the patient health questionnaire and general anxiety disorder measures in the UK Biobank.

people who did not want to stay in school, but rather go to the labor market, to continue their education. These young people may have been negatively affected by being forced to stay in school in a stressful academic environment in which they were less likely to succeed compared to their peers. Courtin et al. (2019) also found that the Berthoin reform—which increased the minimum school leaving age from 14 to 16 in France in 1959—increased depressive symptoms for women in their 60s. These findings imply that results in Gorman (2017) and Davies et al. (2018) who also use the 1972 reform, and more generally those based on compulsory schooling laws as instruments or in a fuzzy RDD, may be invalid because compulsory schooling laws possibly directly lead to worse mental health, which in turn may affect cognition (e.g., Donovan et al. 2017; Nafilyan et al. 2021). Hence, mental health is another channel through which compulsory schooling laws could affect cognition later in life, which would violate the exclusion restriction. The exclusion restriction could also be violated if compulsory schooling laws are correlated with school quality, which can affect cognition independently of level of schooling. Stephens and Yang (2014) show that estimates of the effect of schooling on many outcomes (wages, unemployment, divorce, occupation) in the US using compulsory schooling laws as instruments become insignificant, and in many instances wrong-signed, when controlling for school quality.

Furthermore, in the presence of heterogenous treatment effects, IV methods identify a local average treatment effect (LATE) for those individuals ("compliers") whose treatment is affected by the instrument. In the present context, IV methods estimate the average effect of increasing schooling on cognition only for individuals who increased their education because of the compulsory schooling laws (i.e., those whose schooling would have been lower in the absence of such laws). An analogous interpretation is given to the estimated effect in a fuzzy RDD (e.g., Lee and Lemieux 2010). Previous IV and fuzzy RDD studies are thus not directly informative about the population ATE, or about schooling effects at upper parts of the educational distribution (e.g., college education), as in general compliers are individuals from the lower part of the educational distribution (Clark and Royer 2013). We know of only one IV study that has estimated effects of college education on cognition. Kamhöfer et al. (2019) used student-loan regulations and the availability of colleges in individuals' areas of residence at time of secondary schooling to construct an IV for college education in Germany. They found that college graduates perform 0.73, 1.39 and 1.31 standard deviations better on, respectively, reading speed, reading comprehension and mathematical literacy compared to non-college graduates.

Finally, two studies have employed sibling fixed-effect models with genetic controls. Fletcher et al. (2021) estimated effects of schooling on fluid intelligence in the UKB while also controlling for polygenic scores (summary measures of genetic predisposition) for education, cognition and Alzheimer's disease. They employed a non-linear specification using a series of dummy variables for highest qualification attained rather than grades of schooling. Their sibling-fixed-effect estimates indicate that the fluid intelligence score is 0.35 (0.71) standard deviations higher for high-school (college) graduates compared to high-school dropouts. Using the Wisconsin Longitudinal Study with controls for educational polygenic scores and high-school IQ scores, Herd and Sicinski (2022) found that an extra grade of schooling was associated with a 0.031 standard deviation increase in delayed memory and a 0.065 standard deviation increase in immediate memory though it is important to note that the WLS sample is selected based on a high school "graduate respondent" and thus does not fully represent the distribution of educational outcomes. They also found an extra grade of schooling was associated with a 0.072 standard deviation increase in verbal fluency. Sibling fixed-effects models, however, have the limitations of not providing estimates for the population average effect (given their focus on siblings) and may be prone to unobserved confounding from factors that may vary within siblings.

In sum, the current evidence suggests that there likely is a causal relationship between schooling and memory, but findings for other cognition domains are mixed. The point identification of causal effects, though, rests on strong assumptions, and the effects identified pertain to specific subpopulations (e.g., compliers) and parts of the educational distribution (usually the lower part).

As we describe in the next section, our approach employs relatively weak monotonicity assumptions to provide bounds on the population ATE of increasing schooling at different parts of the educational distribution.

3. Econometric Framework

Let every individual i have a response function $Y(\cdot): T \rightarrow \mathcal{Y}$ which maps treatments $t \epsilon T$ into potential outcomes $Y_i(t) \epsilon \mathcal{Y}$. In our context, the treatment t is schooling attainment consisting of four levels: high-school dropouts (t_1), high-school graduates (t_2), some college education (t_3) and college graduates (t_4). Let S_i denote the realized treatment received by individual i, so that $Y_i \equiv \sum_{t \in T} 1\{S_i = t\} \cdot Y_i(t)$ is the associated observed outcome, where $1\{A\}$ is the indicator function which equals one if the statement A is true and equals zero otherwise. We are interested in the population ATE of, for example, increasing educational attainment from t_1 to t_2 on cognition test scores, defined as:

$$(1)\quad \Delta(t_1, t_2) = E[Y(t_2)] - E[Y(t_1)]$$

Estimation of the ATE is complicated because the potential outcome $Y(t_2)$ is unobserved for individuals with treatment level different from t_2, and the potential outcome $Y(t_1)$ is unobserved for individuals with treatment level different from t_1. This identification problem can be seen by using the law of iterated expectations to write the expected potential outcome $E[Y(t_2)]$ as:

$$(2)\ E[Y(t_2)] = E[Y(t_2)|S < t_2] * P(S < t_2) + E[Y(t_2)|S = t_2] * P(S = t_2) + E[Y(t_2)|S > t_2] * P(S > t_2)$$

The data identify the sample analogues of all the right-side quantities except of the counterfactuals $E[Y(t_2)|S < t_2]$ and $E[Y(t_2)|S > t_2]$. A similar equation applies to $E[Y(t_1)]$. We thus need to impose identifying assumptions about the missing counterfactuals. Manski (1989) suggested a bounded-support assumption, whereby one uses the minimum (Y_{min}) and maximum (Y_{max}) of the outcome variable in place of the counterfactuals (illustrated in Figure 1 panel A). This gives Manski's (1989) "no-assumption" bounds:

$$\begin{aligned}(3)\ & Y_{min} * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + Y_{min} * P(S > t_2) \\ & \leq E[Y(t_2)] \leq \\ & Y_{max} * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + Y_{max} * P(S > t_2)\end{aligned}$$

The no-assumption lower (upper) bound on the ATE $\Delta(t_1, t_2)$ is calculated by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$. Bounds for other treatment effects such as $\Delta(t_2, t_3)$, $\Delta(t_1, t_3)$ or $\Delta(t_2, t_4)$ are computed analogously. In practice, the no-assumption bounds are typically wide and uninformative, and contain zero by construction. To tighten the bounds, we employ three monotonicity assumptions introduced in Manski (1997) and Manski and Pepper (2000): (1) monotone treatment selection; (2) monotone treatment response; and (3) monotone instrumental variable.

3.1 Monotone Treatment Selection (MTS)

We employ the non-negative MTS assumption which captures the notion that, on average, individuals who "selected" into higher education have higher latent cognitive abilities. Formally, for each $t \epsilon T$ and two treatment levels μ_1 and μ_2

$$(4)\ \mu_2 \geq \mu_1 \Rightarrow E[Y(t) \mid S = \mu_2] \geq E[Y(t) \mid S = \mu_1]$$

In our context, the MTS assumption requires that individuals with higher schooling attainment on average have weakly higher potential outcomes at every schooling level t. For example, when comparing college graduates (e.g., $S = \mu_2$) to high school graduates ($S = \mu_1$), the MTS assumption requires that the average potential cognition at older ages at any schooling level t (e.g., as a high-school dropout) of colleges graduates is higher than that of high-school graduates. While the MTS assumption is untestable (since potential outcomes are unobserved), it is plausible in our application. Economic models of human capital posit that individuals with higher innate ability have more schooling (Ben-Porath 1967). Studies have shown that the polygenic scores for education and cognition—which can be interpreted as measures of innate ability—predict cognition at older ages (Ding et al. 2019; Fletcher et al. 2021; Herd and Sicinski 2022), indicating that higher innate ability is also likely related to better cognition at older ages. Thus, given that individuals with higher innate ability are more likely to have more schooling and also better cognition at older ages, it is plausible that, on average, individuals with higher schooling attainment have higher potential cognition at all schooling levels.[3] More generally, the MTS assumption captures the notion that, relative to individuals who self-select into lower schooling levels, individuals who self-select into higher schooling levels are more likely to have pre-treatment characteristics that also make them more likely to have better average potential cognition at older ages at any given schooling level, such as (on average) higher innate ability, better health inputs and better family background, to name a few.

Figure 1 panel B illustrates how the MTS tightens bounds on $E[Y(t_2)]$ relative to the no-assumption bounds. Consider the conditional mean potential outcomes for individuals with $S < t_2$ and $S > t_2$. Under the MTS assumption, $E[\, Y(t_2) \mid S < t_2 \,]$ cannot be more than $E[\, Y(t_2) \mid S = t_2 \,]$, which is identified by the observed mean outcome for those receiving t_2. The observed mean outcome for those receiving t_2 can therefore be used as an upper bound for the mean of $Y(t_2)$ for those with $S < t_2$. Similarly, for the conditional mean potential outcome $E[\, Y(t_2) \mid S > t_2 \,]$, the MTS implies that the unidentified quantity cannot be smaller than $E[\, Y(t_2) \mid S = t_2 \,]$, or the observed mean outcome for $S = t_2,\ E[Y|S = t_2]$. This implies that the observed mean outcome for those receiving t_2 can be used as a lower bound for the mean of $Y(t_2)$ for those with $S > t_2$. Then, the MTS bounds on $E[Y(t_2)]$ are given by (Manski and Pepper, 2000):

$$(5)\ Y_{min} * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + E[Y|S = t_2] * P(S > t_2) \\ \leq E[Y(t_2)] \leq \\ E[Y|S = t_2] * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + Y_{max} * P(S > t_2)$$

As before, the lower (upper) bound on the ATE $\Delta\,(t_1, t_2)$ is calculated by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$, and likewise for other comparisons of interest.

3.2 Monotone Treatment Response (MTR)

We employ the non-negative MTR assumption, imposing the restriction that higher schooling attainment does not decrease cognitive ability at older ages. Formally, for each individual and any treatment levels t_k and t_j:

$$(6)\ t_j \geq t_k \Longrightarrow Y(t_j) \geq Y(t_k)$$

[3] Correlations between other unobserved factors such as innate ability with schooling and cognition are consistent with the MTS. The MTS rules out the possibility that third factors affect cognition in such a way that, on average, at a given level of schooling, more-schooled individuals have worse latent cognitive performance than less-schooled individuals.

There are theoretical models that suggest that the MTR assumption holds. In the Grossman (1972) model of health production individuals choose to invest in health to alter the stock of health, which provides utility itself and because health influences income. Schooling may directly increase health production by increasing the marginal productivity of health inputs or behaviors (productive efficiency) and by enhancing individuals' ability to acquire and process health information (allocative efficiency). Cognition at older ages can be viewed as a component of overall health, which is increased through schooling affecting productive and allocative efficiency. Mazzonna and Perachi (2017) extended the Grossman (1972) model to analyze how retirement affects cognition. Their model shows an increase in the rate of decline of cognitive abilities after retirement, which reflects reduced incentive to invest in cognitive-repair activities (e.g., reading newspapers, engagement in social and cultural activities, clubs). In their model, heterogeneity in preferences for cognitively stimulating activities in the utility function affects the level and age-related decline of cognition. Schooling can thus lead to higher levels of cognition and slower age-related decline if it is associated with undertaking cognitive-repair activities. Similarly, schooling is associated with working in more cognitively demanding occupations, which protects against cognitive decline. The MTR assumption is also consistent with theories of cognitive reserve. The active cognitive-reserve hypothesis (Stern 2002) posits that individuals with more education make more efficient use of brain networks and process tasks more efficiently. This means individuals with more education experience less cognitive decline from brain aging compared to less-educated individuals. The common-cause hypothesis (Stern 2002) argues that if cognition declines in age come from a common cause (e.g., declines in speed of processing), then the cognition of higher-educated individuals will decline at a similar rate to the population rate. However, more-educated individuals will continue to perform at a higher level at a given age because of a greater baseline brain reserve. The compensation hypothesis (Stern 2002) states that education allows more cognitive domains to fully develop, and once brain aging affects cognition, the domains not affected compensate for declines in the other cognitive domains.[4]

The available empirical evidence is also consistent with the MTR. For instance, the Opdebeek et al. (2015) meta-analysis found that an extra grade of schooling is associated with a 0.04-0.08 standard deviation increase in cognitive performance in older adults. Note, however, that the available empirical evidence does not directly inform the MTR assumption, as this assumption is imposed at the individual level and entails a causal effect (rather than associations).

The MTR assumption is stronger than the MTS assumption as it is required to hold for each individual, rather than on average. The MTR assumption would be violated if more schooling leads to worse cognitive performance for some individuals. One could argue, for example, that more schooling may worsen mental health for individuals who work in stressful jobs, or for individuals forced to stay in school by compulsory schooling laws. The deterioration in mental health could lead to worse cognition. Alternatively, the negative impacts of poor mental health on cognition could be outweighed by the positive impacts of schooling on cognitive reserve and mid-life conditions.

A key implication from MTR is that, for example, $E[Y(t_2)|S = t_\ell] \geq E[Y(t_1)|S = t_\ell]$ for any ℓ, given that $t_2 > t_1$. Panel C in Figure 1 illustrates how the MTR assumption narrows bounds for $E[Y(t_2)]$. For any treatment levels $t < t_2$, MTR implies that the conditional mean $E[Y(t_2)|S = t]$ is no less than $E[Y(t)|S = t]$, or the observed mean of Y at t, $E[Y|S = t]$. This increases the lower bound on $E[Y(t_2)]$, relative to that obtained from the bounded support assumption alone. Further, for treatment levels $t' > t_2$, MTR implies that the conditional mean $E[Y(t_2)|S = t']$ cannot be more than $E[Y(t')|S = t']$, which is identified by the observed mean of Y at t', $E[Y|S = t']$. This reduces the upper bound on the unconditional mean $E[Y(t_2)]$ when compared to the no-assumption upper bound.

[4] For example, evidence suggests that more highly educated older persons may capitalize on their higher crystallized abilities to supplement declining fluid abilities (Christensen et al. 1997; Compton et al. 2000).

The MTR bounds on $E[Y(t_2)]$ are given by (Manski 1997):

$$(7)\ E[Y|S<t_2]*P(S<t_2)+ E[Y|S=t_2]*P(S=t_2)+ Y_{min}*P(S>t_2) \\ \leq E[Y(t_2)] \leq \\ Y_{max}*P(S<t_2) + E[Y|S=t_2]*P(S=t_2)+E[Y|S>t_2]*P(S>t_2)$$

As usual, the MTR lower (respectively, upper) bound on the ATE $\Delta(t_1,t_2)$ is calculated by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$. Under the non-negative MTR assumption the lower bound on $\Delta(t_1,t_2)$ is never below zero, because the MTR rules out the possibility that more education worsens cognitive abilities.

The MTR and MTS assumptions can be combined to provide tighter bounds on $E[Y(t_2)]$, given by (Manski and Pepper, 2000):

$$(8)\ E[Y|S<t_2]*P(S<t_2)+ E[Y|S=t_2]*P(S=t_2)+ E[Y|S=t_2]*P(S>t_2) \\ \leq E[Y(t_2)] \leq \\ E[Y|S=t_2]*P(S<t_2) + E[Y|S=t_2]*P(S=t_2)+E[Y|S>t_2]*P(S>t_2)$$

As before, the MTR+MTS lower (upper) bound on the ATE $\Delta(t_1,t_2)$ is calculated by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$. The MTS and MTR assumptions imposed together yield a testable implication that observed mean cognition scores are weakly increasing in schooling attainment. That is, for any two treatments t_k and t_j, $t_j > t_k$ implies that $E\left[Y\middle|S=t_j\right] \geq E[Y|S=t_k]$. This is the case because $t_j > t_k$ implies:

$$(9)\ E\left[Y\middle|S=t_j\right]=E\left[Y(t_j)\middle|S=t_j\right] \geq E\left[Y\left(t_j\right)\middle|S=t_k\right] \geq E[Y(t_k)|S=t_k]=E[Y|S=t_k],$$

where the first inequality follows from the MTS assumption and the second from the MTR. We verify this condition in the next section. Lastly, note that these inequalities help to highlight a key distinction between the MTS and MTR assumptions: while the MTS compares the mean of the same potential outcome for two different subpopulations defined by their observed levels of t, the MTR compares different potential outcomes for the same individual(s). For example, the MTS would compare the average potential cognition at a given schooling level (e.g., as a high school dropout) of college versus high school graduates, while the MTR would compare the potential cognition under a college degree versus the potential cognition under a high school degree for the same individual(s).

3.3 Monotone Instrumental Variable (MIV)

The MTR-MTS bounds can be further narrowed by using a monotone instrumental variable (MIV), which is a variable that has a monotone (weakly increasing or weakly decreasing) mean relationship with the potential outcomes $Y(t)$. The MIV assumption is weaker than the exclusion restriction in IV models, which requires the instrument to be mean independent of the outcome. The MIV assumption also does not require that the variable has a causal effect on the outcome. Specifically, a weakly increasing MIV Z satisfies:

$$(10)\ \ m_1 \leq m \leq m_2 \ \Rightarrow\ \ E[Y(t)|Z=m_1] \leq E[Y(t)|Z=m] \leq E[Y(t)|Z=m_2]$$

for all treatment levels $t \in T$.

We use mother's schooling attainment as the MIV. This is a natural MIV as several studies have shown that higher parental schooling, and more generally childhood socioeconomic status, are associated with better cognition at older ages (see Greenfield and Moorman 2019 for a review).

Theoretically, it has been suggested that higher socioeconomic status in childhood directly affects old-age cognition because it provides resources such as better nutrition or a cognitively stimulating home environment that affect neurocognitive development (Hackman and Farah 2009). There may also be an indirect association through adult outcomes such as schooling attainment.

With a variable Z satisfying the MIV assumption, we can divide the sample into bins defined by the values of Z and compute the MTR-MTS bounds within each bin. In our case of a non-negative MIV, equation (11) implies that the lower bound on $E[Y(t_2)|Z = m]$ is no lower than the lower bound on $E[Y(t_2)|Z = m_1]$, and its upper bound is no higher than the upper bound on $E[Y(t_2)|Z = m_2]$. For the bin where Z has a value of m, we can thus obtain a new lower bound by taking the largest lower bound over all bins where $Z \leq m$. Likewise, we can obtain a new upper bound by taking the smallest upper bound over all bins where $Z \geq m$. The MIV-MTR-MTS bounds are then obtained by taking the weighted average over all the conditional-on-Z bounds (which follows from the law of iterated expectations):

$$(11)\ \sum_{m \in M} P(Z = m) * \left[max_{m_1 \leq m} LB_{E[Y(t_2)|Z=m_1]}\right] \\ \leq E[Y(t_2)] \leq \\ \sum_{m \in M} P(Z = m) * \left[min_{m_2 \geq m} UB_{E[Y(t_2)|Z=m_2]}\right]$$

where LB denotes the MTR+MTS lower bound from equation (8) on $E[Y(t_2)]$ at values $Z = m_1$ of the MIV. Likewise, UB represents the MTR+MTS upper bound on $E[Y(t_2)]$ conditional on values $Z = m_2$ of the MIV.

The MTR+MTS+MIV lower (upper) bound on the ATE $\Delta\ (t_1, t_2)$ is calculated once again by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$.

3.4 Estimation and Inference

The no-assumption, MTS, MTR, and MTR-MTS bounds are all estimated straightforwardly by plugging in sample analogs for the expectations and probabilities in the corresponding bounds' expressions. Inference is undertaken by constructing Imbens and Manski (2004) confidence intervals, which cover the true value of the population average treatment effect of interest with a specified probability (e.g., 95%). Estimation and inference under the MTR+MTS+MIV bounds require that we deal with two issues that have been noted since Manski and Pepper (2000)—see, e.g., Tamer (2010) and references therein. The first is that the plug-in estimators of Equation (12)—an example of so-called intersection bounds—suffer from bias in finite samples that makes them narrower relative to the corresponding true identified set. The bias then carries over to estimated bounds on the average treatment effects of interest. The second, related issue is that the corresponding confidence intervals do not have the expected coverage at the desired level. Both of these issues arise because of the non-concavity and non-convexity, respectively, of the min and max operators in equation (12).

We address both issues in the bounds involving the MIV assumption by employing the estimation and valid-inference procedure in Chernozhukov et al. (2013; hereafter, CLR) for intersection bounds.[5] The CLR procedure allows us to obtain lower- and upper-bound estimators that satisfy a half-median unbiasedness property, that is, the estimated lower (upper) bound will fall below (above) the true lower (upper) bound with a probability of at least one-half asymptotically. This property is important because Hirano and Porter (2012) showed that there exist no locally asymptotically unbiased estimators of parameters that contain min and max

[5] See also Flores and Flores-Lagunes (2013) for additional discussion on the CLR procedure and an application estimating bounds on local-average-treatment effects without the exclusion restriction under a different set of monotonicity assumptions.

operators, implying that methods aimed at reducing bias (such as those based on the bootstrap) cannot completely eliminate it and reducing bias too much eventually leads the variance of such methods to increase significantly. The details on our implementation of the CLR procedure can be found in appendix A.

4. Data

The HRS is a nationally representative longitudinal survey of more than 37,000 individuals over age 50 in 23,000 households in the US. The HRS started in 1992 and data are collected every two years on income and wealth, health, cognition, use of healthcare services; work and retirement, and family connections. The initial HRS cohort consisted of persons born in 1931-41 (then aged 51-61) and their spouses of any age. A second study, Asset and Health Dynamics Among the Oldest (AHEAD), was fielded the next year to capture an older birth cohort, those born in 1890-1923. In 1998, the two studies were merged, and, in order to make the sample fully representative of the older US population, two new cohorts were enrolled, the Children of the Depression (CODA), born in 1924-1930, and the War Babies, born in 1942-1947. The HRS now employs a steady-state design, replenishing the sample every six years with younger cohorts to continue making it fully representative of the population over age 50.

4.1 Schooling Attainment

We obtain schooling attainment of HRS participants and their mothers' schooling attainment from the RAND HRS dataset (version V1), which is a cleaned and streamlined version of the HRS. In both cases, schooling attainment is measured with grades of schooling ranging from 0-17. We discretize participants' grades of schooling into high-school dropout (<12), high-school graduation (12), some college education (13-15) and college graduation (≥16). Mother's schooling, employed as a MIV, is discretized into high-school dropout, high-school graduate and more than high school.

4.2 Cognition Measures

Our cognition measures come from the Harmonized Cognition Protocol Assessment (HCAP), which was initiated in 2016. Participants were selected to be part of HCAP if they were 65 years or older (born 1952 or earlier) and had completed the 2016 interview. HCAP consisted of two parts, a respondent and an informant interview. In the respondent interview, participants completed a comprehensive, in-person neuropsychological assessment (see appendix B for a description of the tests) that took about one hour. Immediately afterwards an individual nominated by the HRS respondent completed an informant interview in another room answering questions on the respondent's functioning and change in ability over the last 10 years. Of the eligible 4,425 participants, 3,496 completed the HCAP interview. There were 149 cases where the HRS respondent was not able to conduct an interview and only the informant interview was conducted.

Table 1 (adapted from Langa et al. 2020) shows the order in which the cognitive tests were undertaken, cognitive domains assessed, number of observations and reasons for missing observations. Generally, missing observations are due to participants refusing to complete the test. Other reasons also include partial interviews, not understanding instructions, and in some instances the test being completed but the data not being recorded. The HRS number series test has the most missing observations (572) due to it being skipped by design (453), participants not understanding the instructions (86), participants refusing to do the test (17) and partial interviews (16).

For comparisons with the literature, we primarily focus on a composite score for immediate and delayed memory constructed by summing up the scores on the HRS TICS, CERAD word list-immediate, CERAD word list-delayed, CERAD constructional praxis-delayed, story-recall immediate, and story-recall delayed tests. The other cognition outcomes we examine are: mini-mental state examination which is seen as a measure of global cognition, recognition memory (sum of scores on CERAD word list-recognition and story recall recognition), verbal fluency (animal naming test), executive function (Raven's test), attention/speed (sum of scores on backward counting and letter cancellation tests) and visuospatial (CERAD constructional praxis-immediate).

5. Results

5.1 Descriptive Statistics

Summary statistics are shown in Table 2. We have an analytical sample of 3,203 individuals with non-missing data on schooling, mother's schooling and race. From the full HCAP sample of 3,496 observations, the majority (287 observations; 8%) are lost due to missing data on mother's schooling. The average age is 77 years (range of 65-103 years), 60% are female, and 73% are non-Hispanic white. Mother's schooling—our MIV—is concentrated at the lower end of the educational distribution. Over half (55%) of individuals had mothers who never graduated from high school and 31% of individuals had mothers who graduated from high school. Only 14% of individuals had mothers who had more than high-school education. Average grade of schooling for HRS participants is 12.86, with 18% of individuals being high-school dropouts and 26% being college graduates. Table 3 shows that average cognition scores are increasing in schooling attainment, which is consistent with the testable implication of the MTS-MTR assumptions discussed in section 3.2.

5.2 Results for Immediate and Delayed Memory

For all results presented henceforth, prior to the analysis all outcomes were adjusted for age, year of birth, gender and race by calculating the residuals from an OLS regression of the outcomes on age, age squared, year-of-birth indicators, and dummy variables for female, black, and Hispanic. The residuals were then scaled back to the outcome by adding the global mean of the outcome to each value. The methods are implemented using the STATA command `mpclr` described and available in Germinario et al. (2021).[6]

Panel A in Table 4 gives results for increasing schooling from primary (high-school dropout) to secondary (high-school graduate). The OLS estimate in column 2 indicates that the memory score of high-school graduates is on average 4.13 points higher than that of high-school dropouts. This represents an average effect of 32% relative to the standard deviation for high-school dropouts (12.87). Columns 3-7 show the estimated bounds and 95% confidence intervals under different set of assumptions. The no-assumption bounds are wide, which is typical for this kind of bounds. They indicate that the true average causal effect of completing secondary schooling could at worst lower memory scores by 59.80 points and at most improve scores by 60.65 points. Adding the MTS assumption—that individuals with higher schooling attainment have on average higher potential cognition—substantially reduces the estimated upper bound. Completing secondary schooling under the MTS assumption in column 4 at most increases memory scores by 25.15 points. The MTS bounds are still wide and null effects cannot be ruled out. Adding the MTR assumption by itself in column 5 reduces the lower bound mechanically to zero, because the MTR assumption rules out the possibility that schooling worsens cognition. The estimated upper bounds under the MTR assumption alone are larger than those under the MTS assumption. The combination of the MTS and MTR assumptions in column 6 provides considerably tighter bounds compared to the previous bounds. Completing secondary schooling at worst has no effect and at most increases memory score by 6.62 points (51% of a standard deviation). To tighten the MTS+MTR bounds we use mother's schooling as a MIV with three bins (high-school dropout, high-school graduate, more than high school). Adding the MIV to the

[6] Programs are available at: https://github.com/grpgerminario/mpclr

MTS+MTR assumptions slightly reduces the estimated upper bound to 5.86 points (46% of a standard deviation) in column 7.

How does the range of causal effects for completing secondary schooling under the MTS+MTR+MIV assumptions compare with previous results in the literature that exploit variation in schooling coming from compulsory schooling laws within an IV or fuzzy RDD context? We first note that the bounds identify the population ATE of completing secondary schooling and are thus not directly comparable to prior IV and fuzzy RDD estimates. Even though these latter estimates also capture the effect at the lower part of the educational distribution, they identify a LATE for individuals who continued their education because of the compulsory schooling laws (the compliers), which may differ for example from the effect for individuals who would have remained in school regardless of the compulsory schooling laws (Clark and Royer, 2013), and thus from the population average effect. Nevertheless, the bounds indicate that there is potentially a very substantial effect of completing secondary schooling—46% of a standard deviation. This implies that an additional grade of schooling increases memory scores by at most 12% of a standard deviation, given a difference of four grades of schooling between high-school graduates and dropouts.[7] Our estimated upper bound is thus substantially smaller than previous IV and fuzzy RDD estimates of an extra grade of schooling for the US (33% of a standard deviation; Glymour et al. 2008) and UK (50-53% of a standard deviation; Banks and Mazzonna 2012; Gorman 2017), but in line with IV estimates identified for Europe (14-16% of a standard deviation; Schneeweis et al. 2014). Under the assumptions imposed by IV methods (including the exclusion restriction) and those on which our bounds are based (MTS+MTR+MIV), the fact that the IV estimate for the US in Glymour et al. (2008) is above our estimated upper bound may be interpreted as reflecting treatment effect heterogeneity. Since compulsory schooling laws are most likely to affect the schooling level of individuals who would otherwise have relatively low schooling (Card, 1999), the average effect for the compliers being larger than for the population would be consistent with these individuals having higher marginal returns to an additional year of secondary schooling in terms of cognition at older ages relative to the overall population. A similar reasoning has been used before in the context of estimating the effect of schooling on earnings, where Card (1999) points to possible differences in the returns to education as a potentially important reason why IV estimates of this effect based on compulsory schooling laws tend to exceed the corresponding OLS estimates. Finally, note that our bounds include the OLS estimates, whereas prior IV estimates are larger than the OLS estimates.[8]

Panel B shows results for increasing schooling from high-school graduation to some college education. The pattern of results from sequentially adding the MTS and MTR assumptions to the no-assumptions bounds is the same as in panel A. Compared to the results in panel A, adding the MIV helps to tighten the MTS+MTR bounds much more. Under the MTS+MTR assumptions, going from being a high-school graduate to having some college education at worst has no effect on memory and at best increases it by 4.64 points. Adding the MIV almost halves the estimated upper bound to 2.36 (17% of a standard deviation). The MTS+MTR+MIV bounds and 95% confidence interval include zero but exclude the OLS point estimate (2.89). Panel C gives results for going from having some college education to graduating from college. The tightest bounds are obtained under the MTS+MTR+MIV assumptions in column 7. They indicate that, on average, graduating from college increases memory scores at most by 4.31 points (32% of a standard deviation) relative to having some college education. The bounds and 95% confidence interval both include zero and the OLS estimate (3.91). The estimated upper bounds

[7] High school dropouts have 8.23 grades of schooling on average in our estimation sample.

[8] In the presence of heterogeneous treatment effects, OLS and IV methods estimate different parameters. Specifically, while OLS estimates the population ATE, IV estimates a LATE. Therefore, differences between OLS and IV estimates may come from possible bias in OLS estimates or from the fact that OLS and IV methods estimate effects for different populations.

under the MTS+MTR+MIV assumptions off increasing schooling from (a) high-school graduation to some college and (b) some college to college graduation are much smaller compared to those from completing high school.

Panel D provides results for increasing schooling from secondary (high-school graduation) to tertiary (college graduation). Here, we obtain fairly tight informative bounds under the MTS+MTR+MIV assumptions. Increasing schooling from secondary to tertiary increases average memory score by 1.28-5.04 points (9-36% of a standard deviation). The estimated bounds exclude the null and OLS estimate (6.80), as does the 95% confidence interval, which implies that the true effect is between 0.60 and 5.70 points (4-41% of a standard deviation). Given a difference of four grades of schooling between college and high-school graduates, the estimated bounds (95% confidence interval) imply that an additional grade of schooling increases memory scores by 2-9% (1-10%) of a standard deviation.

We also obtain somewhat informative bounds under the MTS+MTR+MIV assumptions for increasing schooling from primary to tertiary in panel E. The bounds indicate that the average causal effect is between 2.04-9.88 points (16-77% of a standard deviation). While the 95% confidence interval includes the OLS estimate (10.94), it excludes zero, implying a statistically significant average memory effect of at least 11% of a standard deviation of increasing schooling from primary to tertiary.

5.3 Results for Other Cognition Domains

In this section, we focus on results from increasing schooling from primary to secondary (for comparisons with studies using compulsory schooling laws) and from secondary to tertiary (to see if the informative bounds for memory replicate). Full results for increasing schooling at the different levels on each cognitive domain are given in appendix C.[9] Panel A of Figure 2 presents the effects of completing secondary schooling, indicating that there could be no causal effect, small effects, or potentially large effects on the cognition domains. The literature has found mixed results for cognition domains other than memory, and given the width of the bounds, we cannot make strong comparisons. For example, the bounds (the upper end of the 95% confidence interval) indicate that completing secondary schooling increases verbal fluency scores at most by 46% (52%) of a standard deviation. The estimated upper bound implies that an additional grade of schooling increases verbal fluency by 12% of a standard deviation. This is smaller than the IV estimate in Gorman (2017), who found that an extra grade of schooling increased verbal fluency by 15% of a standard deviation but was imprecisely estimated. The bounds do exclude IV estimates in Schneeweis et al. (2014), where the IV estimates for verbal fluency were all negative and statistically insignificant.

Panel B of Figure 2 shows results for increasing schooling from secondary to tertiary. All of the bounds, except for recognition memory and visuospatial, statistically exclude zero, implying statistically significant average effects of increasing schooling from secondary to tertiary on these cognitive measures. The tightest bounds are obtained for the mini-mental state examination (MMSE), which show that transitioning from being a high-school to college graduate increases MMSE scores by 9-24% of a standard deviation (95% confidence interval of 2-30% of a standard deviation). The width of the bounds is similar for three of the domains (verbal fluency, executive function, attention/speed), indicating average causal effects of about 10-35% of a standard

[9] One small difference in the pattern of results is that adding the MIV to the MTS+MTR assumptions leads to a small increase in the estimated bounds for the effect of going from primary to secondary schooling for MMSE, visuospatial, and recognition memory. This oddity—wider estimated bounds resulting from adding an assumption—can happen in finite samples due to the CLR bias-correction that is applied to the intersection bounds under MTS+MTR+MIV. The estimated bounds under the MTS+MTR (MTS+MTR+MIV) assumptions for these outcomes are as follows: MMSE [0, 1.90] (0, 2.06); visuospatial; [0, 1.01] (0, 1.13); recognition memory [0, 1.45] (0, 1.56).

deviation. All of the bounds and 95% confidence intervals (except for the MMSE and recognition memory) exclude the OLS estimates. These results highlight the potential role that increasing schooling from secondary to tertiary can have in improving cognitive abilities at older ages.

5.4 Replication in the Midlife in United States Development Study (MIDUS)

To assess the external validity of the results from the HRS, we examined the effects of schooling on cognition for a sample of older adults in the MIDUS. The MIDUS is a national sample of 7,108 adults aged 25-74 who were selected via random digit dialing and first interviewed in 1995-96. Nine years later (2004-06) the second wave (MIDUS 2) included data from about 75% (N=4,963) of the original respondents. We use the MIDUS 2 Cognitive Project where 4,512 participants undertook the Brief Test of Adult Cognition by Telephone (BTACT). The BTACT included measures of memory (immediate and delayed recall of 15 words), inductive reasoning (number series; completing a pattern in a series of 5 numbers), verbal fluency (the number of words produced from the category of animals in 60 seconds, as in the HRS), processing speed (backward counting, as in the HRS) and working memory (backward digit span; the highest span achieved in repeating strings of digits in reverse order). Despite its brief length, the BTACT is a reliable and valid measure of cognition (Lachman et al. 2014).

We restricted our analysis to 1,016 individuals aged 65 years or older with data on schooling and mothers' schooling. Summary statistics for the MIDUS analytical sample are given in Table 5. The average age and proportion of women in the MIDUS is similar to the HRS. The distribution of mothers' schooling in the MIDUS and HRS is also similar. Average grades of schooling is higher in the MIDUS (13.79) compared to the HRS (12.86). This reflects that in the MIDUS, 95% of respondents are white whereas in the HRS 73% are white. Average grades of schooling of white individuals in the HRS analytical sample is 13.46, similar to the MIDUS.

Panel A of Figure 3 shows the MTS+MTR+MIV bounds for increasing schooling from primary to secondary on the common cognition domains in the MIDUS and HRS (full results are given in appendix table C7). As found in the HRS, the MIDUS bounds show that the average causal effect of completing secondary schooling could be zero, small, equal to OLS estimates, or potentially larger (but no more than about 50 to 55% of a standard deviation, depending on the specific measure). The OLS estimates and the width of the bounds are quite similar, especially for immediate and delayed memory. Panel B compares the MTS+MTR+MIV bounds for increasing schooling from secondary to tertiary. In the MIDUS, we can statistically exclude null effects for immediate and delayed memory, similar to the HRS. The estimated bounds are tight. Increasing schooling from secondary to tertiary increases average memory scores by 11-21% of a standard deviation. The 95% confidence interval is much wider indicating the causal effect is between 1-34% of a standard deviation. In the MIDUS, the bounds for verbal fluency and attention/speed do not statistically exclude zero, whereas they do in the HRS. In both datasets, all the estimated upper bounds are smaller than the OLS estimates.

5.5 Robustness to Attrition

We examined the robustness of the estimated bounds to attrition in the HRS through inverse probability weighting. We first performed a logit regression on the probability of being in the HRS 2016 survey as a function of year of birth, gender, schooling, mother's schooling (missing values imputed with the sample mean and controlled for with a missingness dummy), self-reported health, mental health (measured by the eight-item center for epidemiological studies depression scale; CES-D), cognition scores, and body mass index (BMI).[10] Self-reported health,

[10] We used the summary measure of cognition in the RAND dataset, which sums up scores from (1) a 10-word immediate and delayed recall tests of memory; (2) a serial 7s subtraction test of working memory; (3) counting backwards to assess attention and processing speed; (4) an object-naming test to assess language; and (5) recall of the date and president and vice-president to assess orientation.

BMI, CES-D, and cognition scores were averaged across the first observed wave through the last observed wave (not including 2016). The inverse of the predicted probabilities was then used to weight the observations when computing the OLS estimates and estimated bounds. Figure 4 shows the inverse-probability-weighted OLS estimates and estimated bounds under the MTS+MTR+MIV assumptions for high-school graduates vs high-school dropouts and college vs high-school graduates. Inverse-probability-weighted OLS estimates and estimated bounds for the other schooling comparisons are given in appendix Table C8. Overall, Figure 4 shows that our estimated bounds and general conclusions are robust to using inverse probability weighting to account for non-random attrition.

6. Summary

Does schooling have a causal effect on cognition at older ages? The evidence for this important question is surprisingly limited, given growing dementia cases, the recognition of schooling as the largest non-biological life-cycle intervention, and the many associations without attempts to provide casual estimates between schooling and various dimensions of aging. We contribute to the literature by employing a partial-identification approach to determine a range of plausible values for the population-average-causal effect of schooling on cognition in the HRS, under weak monotonicity assumptions. We find that the average causal effect of increasing schooling from primary to secondary levels on immediate and delayed memory could be zero, small, or potentially large, but no more than 46% of a standard deviation. The estimated upper bound implies that an additional grade of schooling increases memory scores by at most 12% of a standard deviation. This is substantially smaller than estimates from studies using compulsory schooling laws for identification, where estimates represent a local-average effect only for those who increase their schooling due to these laws. Since compulsory schooling laws are most likely to affect the schooling of people who would otherwise choose low levels of schooling, this result would be consistent with these individuals having higher marginal returns to an additional year of secondary schooling in terms of cognition at older ages relative to the overall population. We also reach similar conclusions for global cognition, verbal fluency, executive function, recognition memory and visuospatial. We further provide new evidence that there are important effects of schooling on cognition at older ages at other parts of the educational distribution. This is critical because the previous literature using compulsory-schooling laws as instruments or in a fuzzy RDD design focuses on obtaining LATE estimates for secondary-school completion, and the effects may differ at other points of the educational distribution. We obtain a fairly narrow range of estimated average causal effects of increasing schooling from secondary to tertiary on all cognition domains. Moreover, these estimated bounds statistically rule out a zero effect for all the cognition domains considered (except for recognition memory). For example, transitioning from being a high-school to a college graduate increases average immediate and delayed memory by between 9% and 36% of a standard deviation. Thus, our analyses lead to a more nuanced and extended understanding of the impacts of different levels of schooling on cognition at older ages in the U.S.

References

Agüero-Torres, H., Fratiglioni, L., Guo, Z., Viitanen, M., & Winblad, B. (1999). Mortality from dementia in advanced age: a 5-year follow-up study of incident dementia cases. *Journal of Clinical Epidemiology*, *52*(8), 737-743.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444-455.

Avendano, M., De Coulon, A., & Nafilyan, V. (2020). Does longer compulsory schooling affect mental health? Evidence from a British reform. *Journal of Public Economics*, *183*, 104137.

Banks, J., & Mazzonna, F. (2012). The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. *The Economic Journal*, *122*(560), 418-448.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, *75*(4, Part 1), 352-365.

Card, D. (1999), "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics* (Vol. 3A), eds. O. Ashenfelter and D. Card, Amsterdam: North Holland, pp. 1801–1863

Chernozhukov, V., Lee, S., & Rosen, A. M. (2013). Intersection bounds: estimation and inference. *Econometrica*, *81*(2), 667-737.

Christelis, D., & Dobrescu, L. (2020). The causal effect of social activities on cognition: Evidence from 20 European countries. *Social Science & Medicine, 247,* 1-9

Christensen, H., Korten, A. E., Jorm, A. F., Henderson, A. S., Jacomb, P. A., Rodgers, B., & Mackinnon, A. J. (1997). Education and decline in cognitive performance: compensatory but not protective. *International Journal of Geriatric Psychiatry*, *12*(3), 323-330.

Clark, Damon, and Heather Royer (2013). The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review*, 103 (6): 2087-2120.

Compton, D. M., Bachman, L. D., Brand, D., & Avet, T. L. (2000). Age-associated changes in cognitive function in highly educated adults: emerging myths and realities. *International Journal of Geriatric Psychiatry*, *15*(1), 75-85.

Courtin, E., Nafilyan, V., Glymour, M., Goldberg, M., Berr, C., Berkman, L. F., ... & Avendano, M. (2019). Long-term effects of compulsory schooling on physical, mental and cognitive ageing: a natural experiment. *Journal of Epidemiology Community Health*, *73*(4), 370-376.

Cygan-Rehm, K., Kuehnle, D., & Oberfichtner, M. (2017). Bounding the causal effect of unemployment on mental health: Nonparametric evidence from four countries. *Health Economics*, *26*(12), 1844-1861.

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

De Haan, M. (2011). The effect of parents' schooling on child's schooling: a nonparametric bounds analysis. *Journal of Labor Economics*, *29*(4), 859-892

Ding, X., Barban, N., Tropf, F. C., & Mills, M. C. (2019). The relationship between cognitive decline and a genetic predictor of educational attainment. *Social Science & Medicine*, *239*, 112549.

Donovan, N. J., Wu, Q., Rentz, D. M., Sperling, R. A., Marshall, G. A., & Glymour, M. M. (2017). Loneliness, depression and cognitive function in older US adults. *International Journal of Geriatric Psychiatry*, *32*(5), 564-573.

Fletcher, J., Topping, M., Zheng, F., & Lu, Q. (2021). The effects of education on cognition in older age: Evidence from genotyped Siblings. *Social Science & Medicine*, *280*, 114044.

Flores, C. A., & Flores-Lagunes, A. (2013). Partial identification of local average treatment effects with an invalid instrument. *Journal of Business & Economic Statistics*, *31*(4), 534-545.

GBD 2019 Dementia Forecasting Collaborators. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*.

Germinario, G., Amin, V., Flores, C.A., & Flores-Lagunes, A. (2022). What Can We Learn About the Effect of Mental Health on Labor Market Outcomes Under Weak Assumptions? Evidence from the NLSY79. Mimeographed: Department of Economics, Syracuse University.

Germinario, G., Flores, C.A., & Flores-Lagunes, A. (2021). mpclr: A Stata command for bias-corrected estimation and valid inference for nonparametric bounds on treatment effects. Mimeographed: Department of Economics, Syracuse University.

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*(6), 532-537.

Gonzalez, L. (2005). Nonparametric bounds on the returns to language skills. *Journal of Applied Econometrics*, *20*(6), 771-795.

Gorman, E. (2017). Schooling, occupation and cognitive function: evidence from compulsory schooling laws. *SocArXiv. October*, *23*.

Greenfield, E. A., & Moorman, S. M. (2019). Childhood socioeconomic status and later life cognition: Evidence from the Wisconsin Longitudinal Study. *Journal of Aging and Health*, *31*(9), 1589-1615.

Grossman, M. (1972): On the concept of health capital and the demand for health. *Journal of Political Economy,* 80(2), 223–255.

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13*(2), 65-73.

Herd, P., & Sicinski, K. (2022). Using sibling models to unpack the relationship between education and cognitive functioning in later life. *Social Science & Medicine-Population Health*, *17*, 100960.

Hirano, K., & Porter, J. R. (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, *80*(4), 1769-1790.

Huang, J., van den Brink, H. M., & Groot, W. (2012). Does education promote social capital? Evidence from IV analysis and nonparametric-bound analysis. *Empirical Economics*, *42*(3), 1011-1034.

Hudomiet, P., Hurd, M. D., & Rohwedder, S. (2018). Dementia prevalence in the United States in 2000 and 2012: estimates based on a nationally representative study. *The Journals of Gerontology: Series B*, *73*(suppl_1), S10-S19.

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., & Langa, K. M. (2013). Monetary costs of dementia in the United States. *New England Journal of Medicine*, *368*(14), 1326-1334.

Imbens, G. W., & Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, *72*(6), 1845-1857.

Kamhöfer, D. A., Schmitz, H., & Westphal, M. (2019). Heterogeneity in marginal non-monetary returns to higher education. *Journal of the European Economic Association*, *17*(1), 205-244.

Lachman, M. E., Agrigoroaei, S., Tun, P. A., & Weaver, S. L. (2014). Monitoring cognitive functioning: Psychometric properties of the brief test of adult cognition by telephone. *Assessment*, *21*(4), 404-417.

Langa, Kenneth M., Eric B. Larson, Eileen M. Crimmins, Jessica D. Faul, Deborah A. Levine, Mohammed U. Kabeto, and David R. Weir. (2017). "A comparison of the prevalence of dementia in the United States in 2000 and 2012." *JAMA Internal Medicine* 177(1): 51-58.

Langa, K. M., Ryan, L. H., McCammon, R. J., Jones, R. N., Manly, J. J., Levine, D. A., ... & Weir, D. R. (2020). The health and retirement study harmonized cognitive assessment protocol project: study design and methods. *Neuroepidemiology*, *54*(1), 64-74.

Lee, David, and Thomas Lemieux (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48 (2): 281-355.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human Resources*, 343-360.

Manski, C. F. (1997). Monotone treatment response. *Econometrica*, 65, 1311-1334.

Manski, C. F., & Pepper, J. V. (2000). Monotone instrumental variables with an application to the returns to schooling. *Econometrica,* 68(July), 997-1010

Mazzonna, F., & Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, *56*(4), 691-710.

Montez, J. K., Hummer, R. A., & Hayward, M. D. (2012). Educational attainment and adult mortality in the United States: A systematic analysis of functional form. *Demography*, *49*(1), 315-336.

Mullahy, J., Venkataramani, A., Millimet, D. & Manski, C. (2021). Embracing uncertainty: The value of partial identification in public health and clinical research. *American Journal of Preventive Medicine*, Forthcoming.

Nafilyan, V., Avendano, M., & de Coulon, A. (2021). The Causal Impact of Depression on Cognitive Functioning: Evidence from Europe. IZA DP No. 14049

Opdebeeck, C., Martyr, A., & Clare, L. (2016). Cognitive reserve and cognitive function in healthy older people: a meta-analysis. *Aging, Neuropsychology, and Cognition*, *23*(1), 40-60.

Schneeweis, N., Skirbekk, V., & Winter-Ebmer, R. (2014). Does education improve cognitive performance four decades after school completion?. *Demography*, *51*(2), 619-643.

Stephens Jr, M., & Yang, D. Y. (2014). Compulsory education and the benefits of schooling. *American Economic Review*, *104*(6), 1777-92.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, *8*(3), 448-460.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics*, *2*, 167-195.

Xu, W., Tan, L., Wang, H. F., Tan, M. S., Tan, L., Li, J. Q., ... & Yu, J. T. (2016). Education and risk of dementia: dose-response meta-analysis of prospective cohort studies. *Molecular Neurobiology*, *53*(5), 3113-3123.

Figure 1: No Assumption, MTS and MTR Bounds for $E[Y(t_2)]$

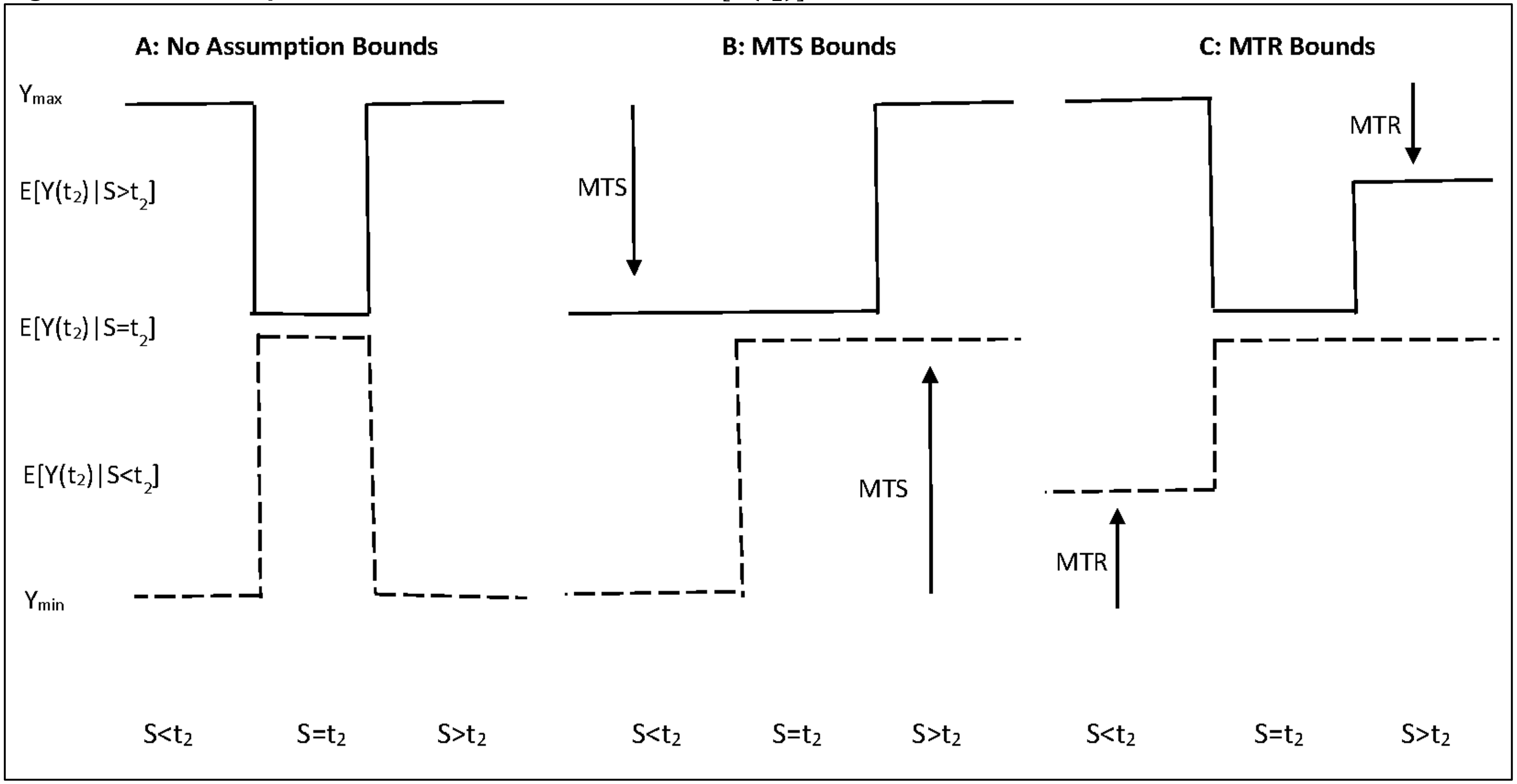


Table 1: Cognition Tests in HCAP

Test	Domains Assessed	Observations	Reasons for Missing Observations (Number)
Mini Mental State Examination	O; MemI; MemD; VF; AS; Viz	3,347	
HRS TICS	MemI	3,344	Refused (2) Not attempted/partial interview (1)
CERAD Word List-Immediate	MemI	3,343	Refused (3) Not attempted/partial interview (1)
Animal Naming	VF	3,345	Not attempted/partial interview (2)
Letter Cancellation	AS	3,197	Refused (64) Not attempted/partial interview (34) Didn't understand instructions (18) Not attempted, interviewer decision (42) Test completed, form lost/not recorded (22)
Backward Counting	AS	3,282	Refused (34) Not attempted/partial interview (5) Didn't understand instructions (25) Not attempted, interviewer decision (1)
Community Screening Instrument for Dementia (CSI-D)	O; EF; VF	3,341	Not attempted/partial interview (6)
CERAD Word List-Delayed	MemD	3,324	Refused (17) Not attempted/partial interview (6)
Story Recall Immediate	MemI	3,306	Refused (26) Not attempted/partial interview (8) Skipped by design (7)
CERAD Word List Recognition	MemR	3,330	Refused (8) Not attempted/partial interview (9)
CERAD Constructional Praxis-Immediate	Viz	3,337	Refused (1) Not attempted/partial interview (9)
Symbol Digit Modalities Test (SDMT)	EF; AS	3,168	Refused (65) Not attempted/partial interview (13) Didn't understand instructions (68) Not attempted, interviewer decision (33)
CERAD Constructional Praxis-Delayed	MemD	3,307	Refused (24) Not attempted/partial interview (13) Completed but can't draw and don't have final score (7)
Story Recall	MemD	3,253	Refused (72)

-Delayed			Not attempted/partial interview (15)
Story Recall -Recognition	MemR	3,243	Refused (80) Not attempted/partial interview (15) Didn't understand instructions (2) Skipped by design (7)
HRS Number Series	EF	2,775	Refused (17) Not attempted/partial interview (16) Didn't understand instructions (86) Skipped by design (453)
Raven's matrices	EF	3,303	Refused (27) Not attempted/partial interview (17)
Trail making Part A	AS	3,243	Refused (33) Not attempted/partial interview (20) Didn't understand instructions (39) Not attempted, interviewer decision (4) Test completed, record lost (9)
Trail making Part B	AS	3,132	Refused (74) Not attempted/partial interview (20) Didn't understand instructions (33) Not attempted, interviewer decision (7) Test completed, record lost (15) Skipped by design (66)

Notes: MemD: Delayed Memory MemI: Immediate Memory MemR: Recognition Memory O: Orientation VF: Verbal Fluency AS: Attention/Speed EF: Executive Function Viz: Visuospatial

Reasons for missing data are taken from the respondent interview codebook: https://hrs.isr.umich.edu/sites/default/files/meta/hcap/2016/codebook/hc16hp_ri.htm

Table 2: Summary Statistics

	Mean (SD)	Min (Max)	Observations
Demographics			
Age	76.86 (7.53)	65 (103)	3,203
Female	0.60 (0.50)	0 (1)	3,203
White	0.73 (0.44)	0 (1)	3,203
Black	0.15 (0.35)	0 (1)	3,203
Hispanic	0.10 (0.30)	0 (1)	3,203
Mother: HS Dropout	0.55 (0.50)	0 (1)	3,203
Mother: HS Grad	0.31 (0.46)	0 (1)	3,203
Mother: More than HS	0.14 (0.35)	0 (1)	3,203
Schooling Attainment			
Grades of Schooling	12.86 (3.10)	0 (17)	3,203
HS Dropout	0.18 (0.39)	0 (1)	3,203
HS Graduate	0.33 (0.47)	0 (1)	3,203
Some College	0.23 (0.42)	0 (1)	3,203
College Graduate	0.26 (0.44)	0 (1)	3,203
Cognition			
Mini Mental State Exam	26.75 (3.86)	0 (30)	3,072
Memory: Immediate+Delayed	36.35 (14.78)	0 (78)	2,952
Memory: Recognition	29.01 (4.26)	0 (35)	2,976
Verbal Fluency	16.16 (6.39)	0 (43)	3,070
Attention/Speed	45.18 (14.05)	0 (95)	2,908
Executive Function	12.48 (14.05)	0 (17)	3,029
Visuospatial	8.24 (2.31)	0 (11)	3,035

Notes: Standard deviations in parentheses.

Table 3: Average Cognition Scores by Schooling Attainment

Cognition	HS Dropout	HS Grad	Some College	College Grad
Mini Mental State Exam	24.24 (4.61)	26.70 (3.62)	27.49 (3.49)	27.87 (3.06)
Memory: Immediate+Delayed	27.15 (12.87)	34.55 (14.06)	38.68 (13.41)	42.49 (14.63)
Memory: Recognition	26.91 (4.45)	29.01 (4.11)	29.48 (3.95)	29.96 (4.12)
Verbal Fluency	12.76 (5.33)	15.30 (6.04)	16.74 (6.29)	19.07 (7.14)
Attention/Speed	34.29 (13.46)	44.29 (12.68)	47.70 (12.43)	50.85 (13.41)
Executive Function	9.63 (3.83)	12.09 (3.57)	13.18 (3.17)	14.26 (2.95)
Visuospatial	6.92 (2.39)	8.05 (2.21)	8.44 (2.13)	9.20 (2.03)

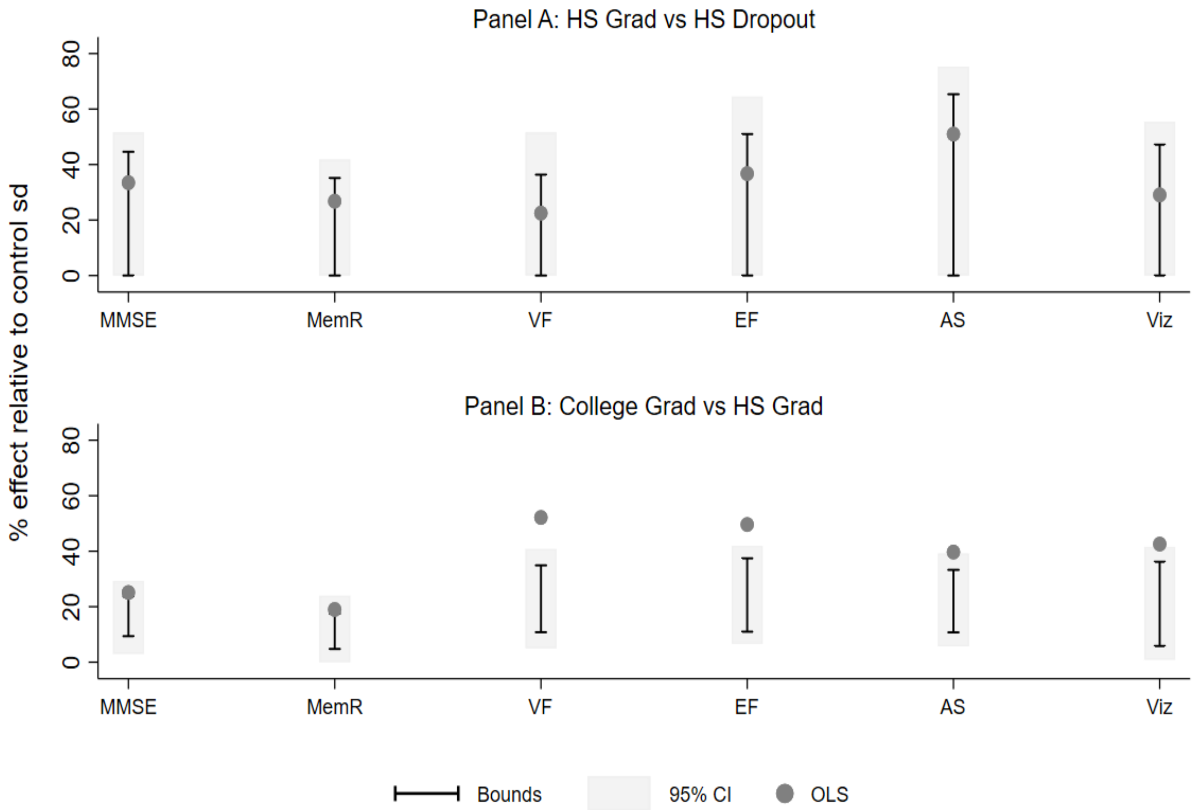
Notes: Standard deviations in parentheses.

Table 4: OLS Estimates and Bounds for the Effect of Schooling on Immediate and Delayed Memory

	Control Mean (SD)	OLS	No Assumption	MTS	MTR	MTR+MTS	MTS+MTR+MIV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A							
HS Grad vs HS Dropout	27.15 (12.87)	4.133*** (.658)	[-59.799, 60.647] (-60.447, 61.290)	[-40.648, 25.152] (-41.240, 26.279)	[0.000, 42.111] (0.000, 42.512)	[0.000, 6.616] (0.000, 7.606)	[0.000, 5.864] (0.000, 7.749)
Panel B							
Some College vs HS Grad	34.55 (14.06)	2.898*** (.623)	[-56.785, 58.494] (-57.464, 59.129)	[-38.691, 19.806] (-39.180, 20.757)	[0.000, 43.330] (0.000, 43.689)	[0.000, 4.642] (0.000, 5.452)	[0.000, 2.361] (0.000, 2.910)
Panel C							
College vs Some College	38.68 (13.41)	3.906*** (.653)	[-59.139, 61.253] (-59.806, 61.863)	[-39.600, 24.482] (-40.188, 25.495)	[0.000, 42.840] (0.000, 43.213)	[0.000, 6.069] (0.000, 6.954)	[0.000, 4.310] (0.000, 5.011)
Panel D							
College vs HS Grad	34.55 (14.06)	6.804*** (.609)	[-54.654, 58.477] (-55.360, 59.090)	[-47.412, 13.407] (-48.088, 14.414)	[0.000, 52.590] (0.000, 53.147)	[0.000, 7.519] (0.000, 8.455)	[1.283, 5.040] (0.604, 5.704)
Panel E							
College vs HS Dropout	27.15 (12.87)	10.937*** (.686)	[-60.444, 65.115] (-61.172, 65.675)	[-60.444, 10.941] (-61.172, 12.049)	[0.000, 65.115] (0.000, 65.675)	[0.000, 10.941] (0.000, 12.070)	[2.043, 9.879] (1.404, 11.807)

Notes: All outcomes were adjusted for age, gender, year of birth and race by calculating the residuals and adding back the global mean. Robust standard errors in (.) in column 2. In columns 3-7 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. The min and max values of the residuals were used in computing the bounds.

Figure 2: MTS+MTR+MIV Bounds for Other Cognition Domains



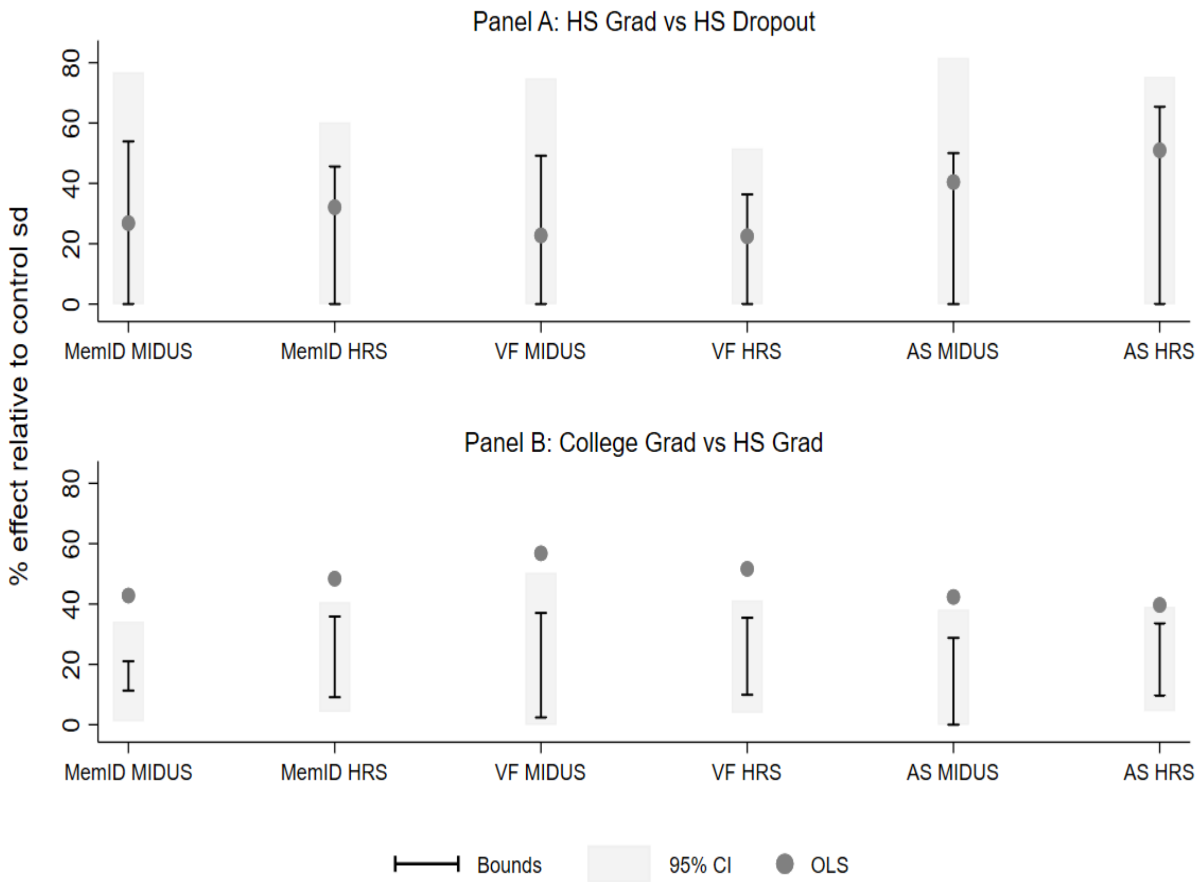
Notes: MMSE:Mini Mental State Examination Memory MemR: Recognition Memory
VF: Verbal Fluency EF:Executive Function AS:Attention/Speed Viz:Visuospatial

Table 5: Summary Statistics for older adults in the MIDUS

	Mean (SD)	Min (Max)	Observations
	(1)	(3)	(5)
Demographics			
Age	72.04 (5.15)	65 (84)	1,016
Female	0.56 (0.50)	0 (1)	1,016
White	0.95 (0.21)	0 (1)	1,016
White Missing	0.04 (0.18)	0 (1)	1,106
Mother: HS Dropout	0.52 (0.50)	0 (1)	1,016
Mother: HS Grad	0.28 (0.45)	0 (1)	1,016
Mother: More than HS Grad	0.19 (0.39)	0 (1)	1,016
Schooling			
Grades of Schooling	13.79 (2.69)	6 (20)	1,016
HS Dropout	0.10 (0.30)	0 (1)	1,016
HS Grad	0.31 (0.46)	0 (1)	1,016
Some College	0.29 (0.45)	0 (1)	1,016
College Grad	0.30 (0.45)	0 (1)	1,016
Cognition			
Immediate+Delayed Memory	9.07 (4.56)	0 (28)	950
Attention/Speed	30.78 (9.23)	0 (100)	1,010
Verbal Fluency	16.24 (5.54)	0 (38)	1,015

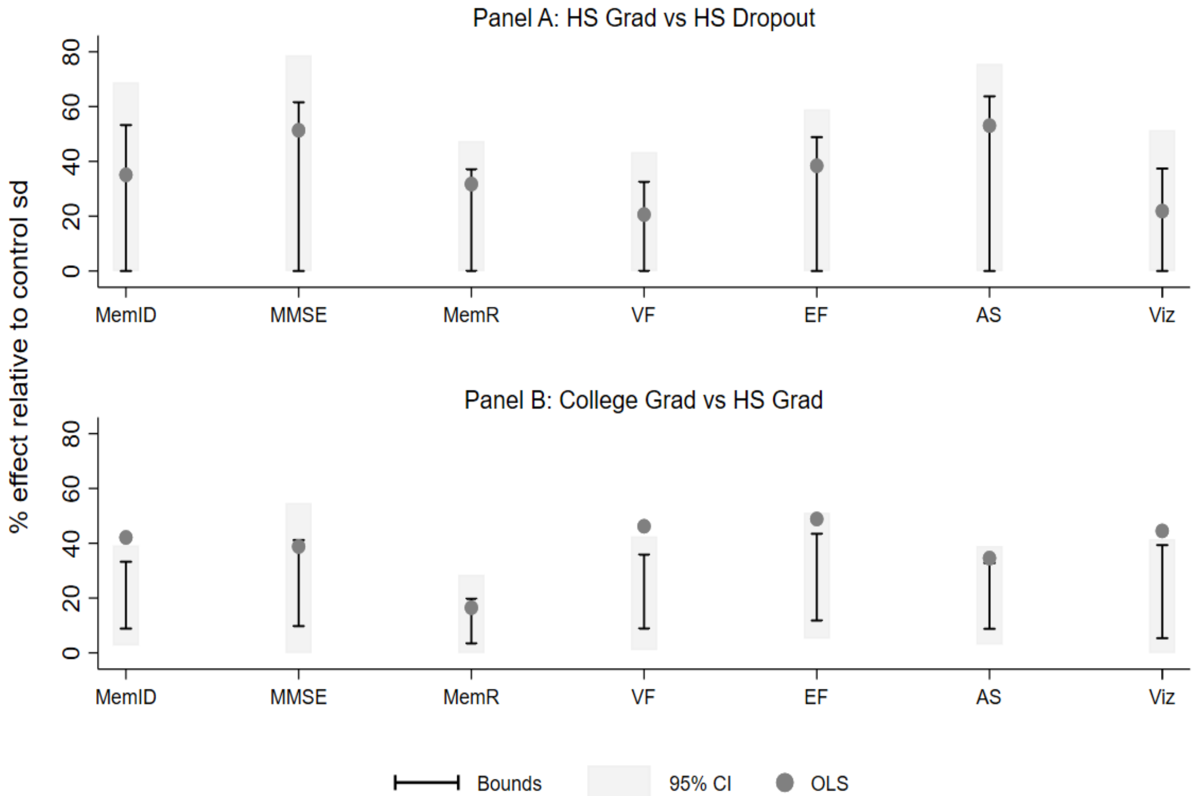
Notes: Standard deviations in parentheses.

Figure 3: Comparison of MTS+MTR+MIV Bounds in HRS and MIDUS



Notes: MemID:Immediate+Delayed Memory VF: Verbal Fluency AS:Attention/Speed

Figure 4: Inverse Probability Weighted MTS+MTR+MIV Bounds and OLS Estimates



Notes: MemID:Immediate+Delayed Memory MMSE:Mini Mental State Examination MemR:Recognition Memory
VF: Verbal Fluency EF:Executive Function AS:Attention/Speed Viz:Visuospatial

Appendix A: Technical Appendix on the CLR Method

This technical appendix provides additional details about the Chernozhukov et al. (2013; CLR) methodology, which we employ to obtain median-bias-corrected estimated bounds and valid confidence intervals for the true parameter value for bounds containing maximum or minimum operators.[11] To provide some intuition on the CLR method, we first make explicit the notion of creating the bins of the MIV. We use below 3 MIV bins $\mathcal{B}_m$, $m = 1, 2, 3$. Then, for instance, the lower bound on $E[Y(t_2)]$ from Equation (11) can be rewritten as:

$$(12) \sum_{m=1}^{3} P(Z \in \mathcal{B}_m) \cdot \max_{m_1 \leq m} LB_{m_1}^1$$

where the $LB_{m_1}^1$ are the MTR-MTS lower bounds in bins m_1 up through m.

Instead of expressions like (12) which comprise 3 different maxima, the CLR method requires that these be rewritten as a set of expressions under a single maximum (or minimum, for upper bounds), with each element inside the max operator called a bounding function. Intuitively, each bounding function represents one of the possible outcomes from evaluating (12) in the data. Finally, the full set of bounding functions is defined for the ATE, so we also perform all necessary subtractions. For example, the final bounding functions for the lower bound on $\Delta(t_1, t_2)$ are created from all possible subtractions of the $E[y(t_1)]$ upper bound bounding functions from the $E[Y(t_2)]$ lower bound bounding functions. In total, each bound on each ATE implies $(2\text{^}\{3 - 1\})\text{^}2 = 16$ bounding functions, denoted $\theta^l(v)$ and $\theta^u(v)$, $v = 1, \ldots, 16$, for the respective lower and upper bounds. Using these bounding functions and denoting the true value of the lower bound of the ATE as θ_0^l and the one for the upper bound as θ_0^u, we can write

$$\theta_0^l = \max_{v \in \mathcal{V}^l} \{\theta^l(v)\}$$

and

$$\theta_0^u = \min_{v \in \mathcal{V}^u} \{\theta^u(v)\},$$

where $\mathcal{V}^l$ and $\mathcal{V}^u$ are the indexing sets for the bounding functions of the lower ($\theta^l(v)$) and upper ($\theta^u(v)$) bounds, respectively.

The key aspect of the CLR procedure is that the steps for estimation of the bounds and for constructing confidence intervals are completed on the individual bounding functions prior to taking the associated maximum (or minimum). This is referred to as the *precision adjustment* and proceeds as follows.[12] Generally, the adjustment involves taking the product of a critical value $\kappa(p)$ and the pointwise standard error $s(v)$ of the bounding function. For lower bounds, this product is subtracted from the estimator $\widehat{\theta^l}(v)$; for upper bounds, it is added to $\widehat{\theta^u}(v)$. Then—depending on the choice of critical value p—the adjustment yields either the half-median unbiased estimator of the lower and upper bounds ($p = 0.5$), or the desired lower and upper limits of the confidence interval (see below). In this way, the CLR method offers the convenience that median-bias correction and inference are carried out within the same procedure. Also, we note that the

[11] This appendix is based on the discussion of the CLR methodology in the context of estimating bounds on the population average treatment effect under the MTS, MTR and MIV assumptions in Germinario et al. (2022), and on a related discussion of the CLR methodology in Flores and Flores-Lagunes (2013).

[12] This process requires that the estimators of $\theta^l(v)$ and $\theta^u(v)$ are consistent and asymptotically normal. Since in our case these estimators are made up of sample means and sample proportions, these conditions are met.

resulting large number of bounding functions makes it crucial to implement the CLR procedure for estimation of the bounds and the construction of valid confidence intervals, as in practice the amount of bias tends to increase with the number of bounding functions (e.g., Germinario et al. 2021).

More specifically, the precision-corrected estimators of the lower (θ_0^l) and upper (θ_0^u) bounds of the average treatment effect are given, respectively, by:

$$(13)\ \widehat{\theta^l}(p) = \max_{v \in \mathcal{V}^l}\{\widehat{\theta^l}(v) - \kappa^l(p) \cdot s^l(v)\}$$

and

$$(14)\ \widehat{\theta^u}(\mathrm{p}) = \min_{v \in \mathcal{V}^u}\{\widehat{\theta^u}(\mathrm{v}) + \kappa^u(\mathrm{p}) \cdot \mathrm{s}^u(\mathrm{v})\}$$

where $\widehat{\theta^l}(v)$ and $\widehat{\theta^u}(v)$ are the unadjusted estimators of the bounding functions, and $s^l(v)$ and $s^u(v)$ are their associated standard errors. The critical values $\kappa^l(p)$ and $\kappa^u(p)$ are computed via simulations as described below. An important feature of the CLR procedure is that the critical values $\kappa^l(p)$ and $\kappa^u(p)$ are computed by simulation not based on the indexing sets $\mathcal{V}^l$ and $\mathcal{V}^u$, but instead based on the preliminary set estimators $\widehat{V^l}$ and $\widehat{V^u}$ of, respectively:

$$V_0^l = \arg\max_{v \in \mathcal{V}^l}\{\theta^l(v)\}$$

and

$$V_0^u = \arg\min_{v \in \mathcal{V}^u}\{\theta^u(v)\}$$

Intuitively, $\widehat{V^l}$ (respectively, $\widehat{V^u}$) selects those bounding functions that are close enough to binding to affect the asymptotic distribution of the estimator of the lower bound $\widehat{\theta^l}(p)$ (upper bound $\widehat{\theta^u}(p)$). This is done because choosing the maximum or minimum over all possible bounding functions by using $\mathcal{V}^l$ and $\mathcal{V}^u$, respectively, leads to asymptotically valid but conservative inference. Below we describe how the preliminary set estimators $\widehat{V^l}$ and $\widehat{V^u}$—which CLR call adaptive inequality selectors—are computed.

First, consider the lower bound, and more specifically, computing $\kappa^l(p)$ and $\widehat{V^l}$. Let $\widehat{\boldsymbol{\gamma}^l}$ be the 16-dimensional column vector of sample analog estimators of all the unadjusted bounding functions for the lower bound, with $\widehat{\boldsymbol{\gamma}^u}$ defined likewise for the upper bounds. An initial step obtains from $B = 999$ bootstrap replications a consistent estimate $\widehat{\Omega}_l$ of the asymptotic variance-covariance matrix of $\sqrt{N}(\widehat{\boldsymbol{\gamma}^l} - \boldsymbol{\gamma}^l)$, where N denotes the sample size (an analogous process is followed for the upper bounds). With $\widehat{\boldsymbol{g}^l}(\boldsymbol{v})'$ the v^{th} row of $\widehat{\Omega}_l^{1/2}$, define $s^l(v) \equiv \frac{||\widehat{g^l}(v)||}{\sqrt{N}}$. Next, following CLR, we simulate $R = 100{,}000$ draws from a $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ distribution, where $\boldsymbol{I}$ is the 16×16 identity matrix. The draws are labelled $\boldsymbol{Z}_r$, $r = 1, \dots, 100{,}000$, and are used to compute $Z_r^*(v) \equiv \widehat{\boldsymbol{g}^l}(\boldsymbol{v})'\boldsymbol{Z}_r / \left|\left|\widehat{g^l}(v)\right|\right|$ for each r and v. In each replication, we select the maximum over the set of $Z_r^*(1), \dots, Z_r^*(16)$. From the resulting R values, we compute $\kappa^l(c)$, which is defined as the c^{th} quantile of these values, where $c \equiv 1 - (0.1/\log N)$. The value of $\kappa^l(c)$ is then used to construct the following set estimator:

$$\widehat{V^l} = \left\{ v \in \mathcal{V}^l : \widehat{\theta^l}(v) \geq \max_{\tilde{v} \in \mathcal{V}^l} \left[\widehat{\theta^l}(\tilde{v}) - \kappa^l(c) \cdot s^l(\tilde{v}) \right] - 2\kappa^l(c) \cdot s^l(v) \right\}$$

From the values $Z_r^*(v)$, we next take the maximum from each replication r, this time restricting the search only to $v \in \widehat{V^l}$ (instead of searching over all the indexes $v \in \mathcal{V}^l$). Lastly, the CLR critical value $\kappa^l(p)$ is taken as the p^{th} quantile of these R values (i.e., as the p^{th} quantile of the R maximums coming from each replication).

Regarding computation of $\kappa^u(p)$ and $\widehat{V^u}$ for the upper bound, the same procedure as above is followed, now defining $\widehat{V^u}$ as:[13]

$$\widehat{V^u} = \left\{ v \in \mathcal{V}^u : \widehat{\theta^u}(v) \leq \min_{\tilde{v} \in \mathcal{V}^u} \left[\widehat{\theta^u}(\tilde{v}) + \kappa^u(c) \cdot s^u(\tilde{v}) \right] + 2\kappa^u(c) \cdot s^u(v) \right\}.$$

Half-median unbiased estimators of the lower and upper bounds of the average treatment effect are obtained by setting $p = 0.5$, computing the critical values $\kappa^l(0.5)$ and $\kappa^u(0.5)$ as described above, and using equations (13) and (14) to compute the half-median unbiased estimates $\widehat{\theta^l}(0.5)$ and $\widehat{\theta^u}(0.5)$.

To obtain $(1-\alpha)\cdot 100\%$ confidence intervals for the true value of the average treatment effect θ_0, we must make one final adjustment which accounts for the width of the identified set. Borrowing notation from CLR (2013), define:

$$\hat{\Gamma} \equiv \widehat{\theta^u}(0.5) - \widehat{\theta^l}(0.5)$$
$$\widehat{\Gamma^+} \equiv \max\{0, \hat{\Gamma}\}$$
$$\rho \equiv \max\{\widehat{\theta^u}(0.75) - \widehat{\theta^u}(0.25), \widehat{\theta^l}(0.25) - \widehat{\theta^l}(0.75)\}$$
$$\tau \equiv 1/(\rho \log N)$$
$$\hat{p} \equiv 1 - \Phi\left(\tau \widehat{\Gamma^+}\right) \cdot \alpha,$$

where $\Phi(\cdot)$ is the standard normal CDF. Note that $\hat{p} \in [1-\alpha, 1-\alpha/2]$, with $\hat{p}$ approaching $1-\alpha$ when $\hat{\Gamma}$ grows large relative to sampling error, and $\hat{p} = 1-\alpha/2$ when $\hat{\Gamma}$ =0. An asymptotically valid $(1-\alpha)\cdot 100\%$ confidence interval for the true value of θ_0 is given by $\left[\widehat{\theta^l}(\hat{p}), \widehat{\theta^u}(\hat{p})\right]$. We report 95% confidence intervals for θ_0 using the critical values $\kappa^l(\hat{p})$ and $\kappa^u(\hat{p})$ with $\alpha = 0.05$ in equations (13) and (14), respectively.

[13] Because of the symmetry of the normal distribution, no changes are needed in any of the other steps.

Appendix B: Description of Cognition Tests

Mini-Mental State Examination (MMSE): This test includes 22 items (10 orientation, 8 language, 1 of each: registration, memory, spelling backward, and construction). The maximum score is 30.

HRS Telephone Interview for Cognition Status (TICS): Participants had to identify two name two words (vocabulary) and naming the President of the United States. Specific questions were: (1) What do people usually use to cut paper? (2) What do you call the kind of prickly plant that grows in the desert? (3) Who is the President of the United States right now? The maximum score is 3.

CERAD Word List Learning and Recall- Immediate: Participants were shown a list of 10 words, two seconds at a time for each word. Participants read each word and after the last word were asked to recall as many words from the list as possible. The score ranges from 0-10.

Animal Naming: Participants were asked to name as many animals as they could within a 60-second time limit. The test score range is 0-43.

Letter Cancellation: Participants had one minute to cross out as many "P" and "W" letters as possible from a large grid of letters.

Backward Counting: Participants had to count backward from 100 as fast as possible in a 30 second time limit.

Community Screening Instrument for Dementia (CSI-D): Participants were asked questions evaluating language, knowledge and the ability to follow directions. The questions/tasks were: (1) point to your elbow; (2) what do you do with a hammer? (3) where is the local market/ local store?; (4) point first to the window and then to the door. The maximum score is 4.

CERAD Word List-Delayed: This is a single trial to recall the list of 10 words from the CERAD Word List Learning and Recall (Immediate) task. Participants are asked to freely recall as many words as possible from that list. The interviewer records all correct responses as well as intrusions (words not on the original list). Respondents are given up to 2 minutes to complete this task.

Story Memory-Immediate: Participants were read one of two from the Wechsler Memory Scale (WMS-IV). participants had to report back on the main parts of the story immediately after it was read.

CERAD Word List-Recognition: Participants were visually presented a series of 20 words, 10 from the CERAD word list and 10 foils. They were asked to identify which words were given on the original list

CERAD Constructional Praxis – Immediate: Participants had to copy geometric figures that varied in difficulty

Symbol-Digit Modalities Test (SDMT): Participants were given random geometric figures and a separate key that paired numbers with each figure. Participants had to substitute a number of each figure, completing as many pairings as possible in the 90-second time limit.

CERAD Constructional Praxis – Delayed: This is a delayed recall of the geometric shaped drawn in the test of CERAD Constructional Praxis – Immediate. Respondents are asked to draw the shapes from earlier in the interview to the best of their memory.

Story Memory-Delayed: Participants were asked to think back to the two stories read to them earlier and recall as much about each story as they can.

Story Memory-Recognition: Participants were given 15 yes/no questions on whether a specified story point was part of the story they were read

HRS Number Series: Participants were presented with a series of numbers with one or two numbers missing. Participants had to identify the missing numbers. The test was not timed and was adaptive such that difficulty level changed depending on the participants' responses. The range is 409-584.

Raven's Standard Progressive Matrices: This test evaluates picture-based pattern reasoning of varying difficulty. Each question presents a geometric picture with a small section that appears to have been cut out. Participants are shown a set of smaller pictures that fit the missing piece and are asked to identify which is the correct one to complete the pattern.

Trail Making Test (A and B): Participants asked to draw lines connecting consecutively numbered circles on a worksheet (part A) and connect consecutively numbered and lettered circles on another worksheet (part B) by alternating between the numbers and letters. The interviewer is instructed to point out errors to the participant and have the participant go back to the previous circle and move on to the next correct one. The score for this test is the number of seconds to complete part A and part B, where the time to correct errors serves to increase the total time to complete the test.

Appendix C: Additional Tables

Table C1: OLS Estimates and Bounds for the Effect of Schooling on the Mini-Mental State Examination

	Control Mean (SD)	OLS	No Assumption	MTS	MTR	MTR+MTS	MTS+MTR+MIV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A							
HS Grad vs HS Dropout	24.24 (4.61)	1.543*** (.206)	[-23.348, 26.672] (-23.682, 26.936)	[-10.597, 5.292] (-10.831, 5.605)	[0.000, 23.280] (0.000, 23.492)	[0.000, 1.899] (0.000, 2.212)	[0.000, 2.056] (0.000, 2.378)
Panel B							
Some College vs HS Grad	26.70 (3.62)	0.540*** (.163)	[-24.847, 23.330] (-25.169, 23.667)	[-16.638, 6.987] (-16.941, 7.334)	[0.000, 17.248] (0.000, 17.533)	[0.000, 0.905] (0.000, 1.138)	[0.000, 0.565] (0.000, 0.743)
Panel C							
College vs Some College	27.49 (3.49)	0.346** (.161)	[-24.930, 25.614] (-25.259, 25.931)	[-21.338, 13.777] (-21.638, 14.202)	[0.000, 12.733] (0.000, 13.000)	[0.000, 0.895] (0.000, 1.129)	[0.068, 0.559] (0.000, 0.767)
Panel D							
College vs HS Grad	26.70 (3.62)	0.886*** (.146)	[-24.068, 23.235] (-24.391, 23.567)	[-22.712, 5.500] (-23.023, 5.863)	[0.000, 18.897] (0.000, 19.231)	[0.000, 1.161] (0.000, 1.400)	[0.321, 0.878] (0.072, 1.080)
Panel E							
College vs HS Dropout	24.24 (4.61)	2.429*** (.205)	[-24.949, 27.440] (-25.296, 27.709)	[-24.949, 2.430] (-25.296, 2.768)	[0.000, 27.440] (0.000, 27.709)	[0.000, 2.430] (0.000, 2.782)	[0.596, 2.607] (0.389, 2.972)

Notes: All outcomes were adjusted for age, gender, year of birth and race by calculating the residuals and adding back the global mean. Robust standard errors in (.) in column 2. In columns 3-7 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. The min and max values of the residuals were used in computing the bounds.

Table C2: OLS Estimates and Bounds for the Effect of Schooling on Verbal Fluency

	Control Mean (SD)	OLS	No Assumption	MTS	MTR	MTR+MTS	MTS+MTR+MIV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A							
HS Grad vs HS Dropout	12.76 (5.33)	1.199*** (.283)	[-33.421, 33.081] (-33.778, 33.451)	[-23.701, 13.405] (-23.955, 13.950)	[0.000, 21.903] (0.000, 22.093)	[0.000, 2.227] (0.000, 2.661)	[0.000, 1.937] (0.000, 2.750)
Panel B							
Some College vs HS Grad	15.30 (6.04)	0.943*** (.285)	[-31.560, 32.488] (-31.927, 32.832)	[-21.783, 10.639] (-22.019, 11.121)	[0.000, 23.570] (0.000, 23.749)	[0.000, 1.721] (0.000, 2.106)	[0.000, 0.758] (0.000, 1.085)
Panel C							
College Grad vs Some College	16.74 (6.29)	2.177*** (.318)	[-33.091, 34.137] (-33.469, 34.485)	[-20.961, 12.699] (-21.247, 13.230)	[0.000, 24.308] (0.000, 24.482)	[0.000, 2.871] (0.000, 3.324)	[0.111, 1.943] (0.000, 2.302)
Panel D							
College Grad vs HS Grad	15.30 (6.04)	3.120*** (.287)	[-30.471, 32.445] (-30.854, 32.779)	[-26.050, 6.643] (-26.410, 7.145)	[0.000, 29.137] (0.000, 29.432)	[0.000, 3.335] (0.000, 3.800)	[0.599, 2.142] (0.242, 2.489)
Panel E							
College Grad vs HS Dropout	12.76 (5.33)	4.319*** (.317)	[-34.024, 35.658] (-34.401, 35.984)	[-34.024, 4.320] (-34.401, 4.847)	[0.000, 35.658] (0.000, 35.984)	[0.000, 4.320] (0.000, 4.863)	[0.808, 3.748] (0.459, 4.675)

Notes: All outcomes were adjusted for age, gender, year of birth and race by calculating the residuals and adding back the global mean. Robust standard errors in (.) in column 2. In columns 3-7 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. The min and max values of the residuals were used in computing the bounds.

Table C3: OLS Estimates and Bounds for the Effect of Schooling on Executive Function

	Control Mean (SD)	OLS	No Assumption	MTS	MTR	MTR+MTS	MTS+MTR+MIV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A							
HS Grad vs HS Dropout	9.63 (3.83)	1.407*** (.184)	[-16.327, 17.713] (-16.527, 17.876)	[-9.274, 5.637] (-9.426, 5.931)	[0.000, 14.147] (0.000, 14.254)	[0.000, 2.071] (0.000, 2.353)	[0.000, 1.952] (0.000, 2.469)
Panel B							
Some College vs HS Grad	12.09 (3.57)	0.829*** (.150)	[-16.439, 16.284] (-16.636, 16.488)	[-10.999, 5.358] (-11.153, 5.623)	[0.000, 12.252] (0.000, 12.383)	[0.000, 1.326] (0.000, 1.530)	[0.000, 0.762] (0.000, 0.880)
Panel C							
College Grad vs Some College	13.18 (3.17)	0.879*** (.142)	[-16.818, 17.470] (-17.017, 17.656)	[-12.945, 8.497] (-13.124, 8.770)	[0.000, 10.512] (0.000, 10.649)	[0.000, 1.539] (0.000, 1.732)	[0.000, 1.066] (0.000, 1.231)
Panel D							
College Grad vs HS Grad	12.09 (3.57)	1.709*** (.135)	[-15.801, 16.298] (-15.998, 16.489)	[-14.304, 4.216] (-14.497, 4.470)	[0.000, 14.042] (0.000, 14.224)	[0.000, 1.960] (0.000, 2.166)	[0.384, 1.309] (0.220, 1.470)
Panel E							
College Grad vs HS Dropout	9.63 (3.83)	3.115*** (.178)	[-16.861, 18.744] (-17.082, 18.908)	[-16.861, 3.135] (-17.082, 3.434)	[0.000, 18.744] (0.000, 18.908)	[0.000, 3.135] (0.000, 3.438)	[0.639, 2.945] (0.475, 3.460)

Notes: All outcomes were adjusted for age, gender, year of birth and race by calculating the residuals and adding back the global mean. Robust standard errors in (.) in column 2. In columns 3-7 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. The min and max values of the residuals were used in computing the bounds.

Table C4: OLS Estimates and Bounds for the Effect of Schooling on Attention/Speed

	Control Mean (SD)	OLS	No Assumption	MTS	MTR	MTR+MTS	MTS+MTR+MIV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A							
HS Grad vs HS Dropout	34.29 (12.46)	6.351*** (.672)	[-69.360, 71.015] (-70.144, 71.754)	[-46.624, 30.489] (-47.157, 31.707)	[0.000, 48.775] (0.000, 49.142)	[0.000, 8.250] (0.000, 9.275)	[0.000, 8.142] (0.000, 9.378)
Panel B							
Some College vs HS Grad	44.29 (12.68)	2.271*** (.587)	[-66.297, 67.609] (-67.086, 68.333)	[-45.663, 21.949] (-46.134, 22.921)	[0.000, 49.679] (0.000, 50.049)	[0.000, 4.019] (0.000, 4.827)	[0.000, 2.618] (0.000, 3.159)
Panel C							
College Grad vs Some College	47.70 (12.43)	2.763*** (.622)	[-69.375, 70.903] (-70.154, 71.642)	[-46.448, 26.566] (-47.021, 27.647)	[0.000, 49.270] (0.000, 49.645)	[0.000, 4.933] (0.000, 5.797)	[0.000, 3.236] (0.000, 3.953)
Panel D							
College Grad vs HS Grad	44.29 (12.68)	5.035*** (.561)	[-64.037, 66.877] (-64.836, 67.578)	[-56.090, 12.491] (-56.838, 13.502)	[0.000, 60.433] (0.000, 61.065)	[0.000, 6.047] (0.000, 6.971)	[1.220, 4.252] (0.579, 4.948)
Panel E							
College Grad vs HS Dropout	34.29 (12.46)	11.385*** (.703)	[-71.126, 75.620] (-71.960, 76.287)	[-71.126, 11.389] (-71.960, 12.551)	[0.000, 75.620] (0.000, 76.287)	[0.000, 11.389] (0.000, 12.573)	[2.244, 11.118] (1.639, 12.521)

Notes: All outcomes were adjusted for age, gender, year of birth and race by calculating the residuals and adding back the global mean. Robust standard errors in (.) in column 2. In columns 3-7 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. The min and max values of the residuals were used in computing the bounds.

Table C5: OLS Estimates and Bounds for the Effect of Schooling on Recognition Memory

	Control Mean (SD)	OLS	No Assumption	MTS	MTR	MTR+MTS	MTS+MTR+MIV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A							
HS Grad vs HS Dropout	26.91 (4.45)	1.192*** (.223)	[-25.486, 28.701] (-25.861, 28.998)	[-11.955, 5.708] (-12.218, 6.048)	[0.000, 24.445] (0.000, 24.674)	[0.000, 1.452] (0.000, 1.795)	[0.000, 1.564] (0.000, 1.861)
Panel B							
Some College vs HS Grad	29.01 (4.11)	0.266 (.188)	[-26.702, 25.178] (-27.034, 25.544)	[-17.950, 7.254] (-18.270, 7.652)	[0.000, 18.523] (0.000, 18.823)	[0.000, 0.599] (0.000, 0.885)	[0.000, 0.412] (0.000, 0.621)
Panel C							
College Grad vs Some College	29.48 (3.95)	0.480** (.194)	[-26.748, 27.487] (-27.104, 27.825)	[-22.354, 14.182] (-22.666, 14.658)	[0.000, 14.126] (0.000, 14.394)	[0.000, 0.820] (0.000, 1.102)	[0.019, 0.496] (0.000, 0.762)
Panel D							
College Grad vs HS Grad	29.01 (4.11)	0.746*** (.183)	[-25.858, 25.074] (-26.199, 25.436)	[-24.297, 5.428] (-24.617, 5.838)	[0.000, 20.599] (0.000, 20.965)	[0.000, 0.953] (0.000, 1.255)	[0.169, 0.715] (0.000, 0.970)
Panel E							
College Grad vs HS Dropout	26.91 (4.45)	1.939*** (.228)	[-27.056, 29.486] (-27.415, 29.785)	[-27.056, 1.940] (-27.415, 2.302)	[0.000, 29.486] (0.000, 29.785)	[0.000, 1.940] (0.000, 2.325)	[0.383, 1.995] (0.173, 2.373)

Notes: All outcomes were adjusted for age, gender, year of birth and race by calculating the residuals and adding back the global mean. Robust standard errors in (.) in column 2. In columns 3-7 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. The min and max values of the residuals were used in computing the bounds.

Table C6: OLS Estimates and Bounds for the Effect of Schooling on Visuospatial

	Control Mean (SD)	OLS	No Assumption	MTS	MTR	MTR+MTS	MTS+MTR+MIV
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Panel A							
HS Grad vs HS Dropout	6.92 (2.39)	0.695*** (.119)	[-9.342, 10.087] (-9.464, 10.187)	[-5.333, 3.099] (-5.433, 3.281)	[0.000, 7.997] (0.000, 8.068)	[0.000, 1.009] (0.000, 1.193)	[0.000, 1.127] (0.000, 1.323)
Panel B							
Some College vs HS Grad	8.05 (2.21)	0.286*** (.103)	[-9.441, 9.259] (-9.551, 9.373)	[-6.392, 2.920] (-6.489, 3.082)	[0.000, 6.920] (0.000, 7.000)	[0.000, 0.582] (0.000, 0.727)	[0.000, 0.323] (0.000, 0.406)
Panel C							
College Grad vs Some College	8.44 (2.13)	0.662*** (.104)	[-9.589, 10.032] (-9.713, 10.138)	[-7.260, 4.937] (-7.378, 5.112)	[0.000, 6.026] (0.000, 6.107)	[0.000, 0.931] (0.000, 1.069)	[0.000, 0.656] (0.000, 0.789)
Panel D							
College Grad vs HS Grad	8.05 (2.21)	0.948*** (.096)	[-9.050, 9.311] (-9.170, 9.426)	[-8.185, 2.391] (-8.296, 2.563)	[0.000, 7.991] (0.000, 8.099)	[0.000, 1.072] (0.000, 1.222)	[0.098, 0.785] (0.000, 0.907)
Panel E							
College Grad vs HS Dropout	6.92 (2.39)	1.644*** (.120)	[-9.673, 10.678] (-9.795, 10.774)	[-9.673, 1.644] (-9.795, 1.839)	[0.000, 10.678] (0.000, 10.774)	[0.000, 1.644] (0.000, 1.844)	[0.230, 1.731] (0.127, 1.949)

Notes: All outcomes were adjusted for age, gender, year of birth and race by calculating the residuals and adding back the global mean. Robust standard errors in (.) in column 2. In columns 3-7 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. The min and max values of the residuals were used in computing the bounds.

Table C7: OLS Estimates and Bounds for the Effect of Schooling on Cognitive Abilities of Older Individuals in the MIDUS

	Control Mean (SD)	OLS	MTS+MTR	MTS+MTR+MIV
Immediate+Delayed Memory				
HS Grad vs HS Dropout	7.15 (3.72)	0.998** (.448)	[0.000, 1.867] (0.000, 2.568)	[0.000, 2.005] (0.000, 2.855)
Some College vs HS Grad	8.43 (4.18)	1.119*** (.348)	[0.000, 1.417] (0.000, 1.919)	[0.011, 0.676] (0.000, 0.963)
College Grad vs Some College	9.63 (4.92)	0.671* (.368)	[0.000, 1.222] (0.000, 1.746)	[0.000, 0.731] (0.000, 1.290)
College Grad vs HS Grad	8.43 (4.18)	1.790*** (.333)	[0.000, 1.885] (0.000, 2.436)	[0.473, 0.879] (0.051, 1.426)
College Grad vs HS Dropout	7.15 (3.72)	2.788*** (.464)	[0.000, 2.790] (0.000, 3.575)	[0.502, 2.732] (0.174, 3.680)
Verbal Fluency				
HS Grad vs HS Dropout	13.91 (5.04)	1.147** (.564)	[0.000, 2.224] (0.000, 3.135)	[0.000, 2.478] (0.000, 3.766)
Some College vs HS Grad	15.26 (5.20)	0.739* (.406)	[0.000, 1.582] (0.000, 2.166)	[0.000, 0.499] (0.000, 0.759)
College Grad vs Some College	15.94 (5.25)	2.215*** (.432)	[0.000, 2.655] (0.000, 3.286)	[0.000, 1.830] (0.000, 2.525)
College Grad vs HS Grad	15.26 (5.20)	2.954*** (.430)	[0.000, 3.131] (0.000, 3.826)	[0.125, 1.926] (0.000, 2.618)
College Grad vs HS Dropout	13.91 (5.04)	4.101*** (.583)	[0.000, 4.104] (0.000, 5.131)	[0.235, 4.071] (0.000, 5.565)
Attention/Speed				
HS Grad vs HS Dropout	26.07 (8.45)	3.417*** (.925)	[0.000, 4.649] (0.000, 6.238)	[0.000, 4.228] (0.000, 6.893)
Some College vs HS Grad	29.83 (8.50)	0.508 (.649)	[0.000, 1.774] (0.000, 2.770)	[0.000, 0.817] (0.000, 1.382)
College Grad vs Some College	30.14 (8.63)	3.093*** (.710)	[0.000, 3.639] (0.000, 4.713)	[0.000, 2.446] (0.000, 3.239)
College Grad vs HS Grad	29.83 (8.50)	3.600*** (.698)	[0.000, 3.939] (0.000, 5.107)	[0.038, 2.792] (0.000, 3.608)
College Grad vs HS Dropout	26.07 (8.45)	7.018*** (.969)	[0.000, 7.023] (0.000, 8.751)	[0.367, 6.380] (0.000, 9.142)

Notes: All outcomes were adjusted for age, gender, year of birth and race by calculating the residuals and adding back the global mean. Robust standard errors in (.) in column 2. In columns 3 and 4 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. The min and max values of the residuals were used in computing the bounds.

Table C8: Inverse Probability Weighted OLS Estimates and Bounds for the Effect of Schooling on Cognition

	Control Mean (SD)	OLS	MTS+MTR	MTS+MTR+MIV
	(1)	(2)	(3)	(4)
Panel A: Immediate+Delayed Memory				
HS Grad vs HS Dropout	27.15 (12.88)	4.355*** (.953)	[0.000, 6.393] (0.000, 7.887)	[0.000, 6.216] (0.000, 7.926)
Some College vs HS Grad	34.55 (14.06)	2.912*** (.730)	[0.000, 4.601] (0.000, 5.628)	[0.000, 2.946] (0.000, 3.594)
College Grad vs Some College	38.67 (13.41)	3.010*** (.771)	[0.000, 5.576] (0.000, 6.665)	[0.000, 3.673] (0.000, 4.523)
College Grad vs HS Grad	34.55 (14.06)	5.922*** (.719)	[0.000, 6.868] (0.000, 8.028)	[1.246, 4.675] (0.382, 5.497)
College Grad vs HS Dropout	27.15 (12.88)	10.276*** (.985)	[0.000, 10.281] (0.000, 11.906)	[2.260, 9.541] (1.417, 11.368)
Panel B: Mini-Mental State Examination				
HS Grad vs HS Dropout	24.23 (4.62)	2.313*** (.507)	[0.000, 2.757] (0.000, 3.595)	[0.000, 2.785] (0.000, 3.559)
Some College vs HS Grad	26.70 (3.62)	0.594* (.312)	[0.000, 1.345] (0.000, 1.860)	[0.000, 1.001] (0.000, 1.485)
College Grad vs Some College	27.49 (3.46)	0.811* (.423)	[0.000, 1.716] (0.000, 2.403)	[0.000, 0.859] (0.000, 1.294)
College Grad vs HS Grad	26.70 (3.62)	1.405*** (.370)	[0.000, 1.967] (0.000, 2.643)	[0.353, 1.490] (0.000, 1.978)
College Grad vs HS Dropout	24.23 (4.62)	3.718*** (.582)	[0.000, 3.721] (0.000, 4.726)	[0.834, 3.629] (0.300, 4.433)
Panel C: Recognition Memory				
HS Grad vs HS Dropout	26.94 (4.40)	1.413*** (.339)	[0.000, 1.639] (0.000, 2.219)	[0.000, 1.642] (0.000, 2.092)
Some College vs HS Grad	29.01 (4.11)	2.961 (.265)	[0.000, 0.698] (0.000, 1.091)	[0.000, 0.674] (0.000, 0.943)
College Grad vs Some College	29.49 (3.96_	0.382 (.276)	[0.000, 0.855] (0.000, 1.267)	[0.000, 0.474] (0.000, 0.827)
College Grad vs HS Grad	29.01 (4.11)	0.678*** (.248)	[0.000, 0.986] (0.000, 1.419)	[0.141, 0.810] (0.000, 1.171)
College Grad vs HS Dropout	26.94 (4.40)	2.091*** (.348)	[0.000, 2.093] (0.000, 2.730)	[0.461, 2.014] (0.128, 2.564)
Panel D: Verbal Fluency				
HS Grad vs HS Dropout	12.74 (5.34)	1.199*** (.339)	[0.000, 2.027] (0.000, 2.561)	[0.000, 1.809] (0.000, 2.451)
Some College vs HS Grad	15.30 (6.04)	0.925*** (.342)	[0.000, 1.653] (0.000, 2.122)	[0.000, 0.969] (0.000, 1.393)
College Grad vs Some College	16.74 (6.29)	1.866*** (.374)	[0.000, 2.692] (0.000, 3.195)	[0.003, 1.868] (0.000, 2.269)
College Grad vs HS Grad	15.30	2.791***	[0.000, 3.083]	[0.535, 2.163]

	(6.04)	(.330)	(0.000, 3.611)	(0.063, 2.563)
College Grad vs HS Dropout	12.74 (5.34)	3.990*** (.370)	[0.000, 3.992] (0.000, 4.639)	[0.840, 3.526] (0.375, 4.265)
Panel E: Executive Function				
HS Grad vs HS Dropout	9.62 (3.83)	1.464*** (.249)	[0.000, 2.066] (0.000, 2.472)	[0.000, 1.857] (0.000, 2.216)
Some College vs HS Grad	12.09 (3.57)	0.867*** (.193)	[0.000, 1.444] (0.000, 1.728)	[0.000, 0.972] (0.000, 1.163)
College Grad vs Some College	13.18 (3.17)	0.877*** (.205)	[0.000, 1.723] (0.000, 2.023)	[0.000, 1.203] (0.000, 1.481)
College Grad vs HS Grad	12.09 (3.57)	1.744*** (.195)	[0.000, 2.101] (0.000, 2.427)	[0.421, 1.551] (0.186, 1.823)
College Grad vs HS Dropout	9.62 (3.83)	3.208*** (.258)	[0.000, 3.228] (0.000, 3.668)	[0.769, 2.963] (0.536, 3.406)
Panel F: Attention/Speed				
HS Grad vs HS Dropout	34.23 (12.48)	6.522*** (1.018)	[0.000, 8.081] (0.000, 9.733)	[0.000, 8.246] (0.000, 9.732)
Some College vs HS Grad	44.29 (12.69)	2.210*** (.646)	[0.000, 4.058] (0.000, 5.009)	[0.000, 3.083] (0.000, 3.729)
College Grad vs Some College	47.70 (12.44)	2.179*** (.694)	[0.000, 4.731] (0.000, 5.707)	[0.000, 2.818] (0.000, 3.600)
College Grad vs HS Grad	44.29 (12.69)	4.389*** (.627)	[0.000, 5.693] (0.000, 6.731)	[1.115, 4.158] (0.389, 4.945)
College Grad vs HS Dropout	34.23 (12.48)	10.912*** (1.0049)	[0.000, 10.917] (0.000, 12.667)	[2.441, 10.755] (1.720, 12.392)
Panel G: Visuospatial				
HS Grad vs HS Dropout	6.92 (2.39)	0.548*** (.171)	[0.000, 0.856] (0.000, 1.137)	[0.000, 0.942] (0.000, 1.229)
Some College vs HS Grad	8.05 (2.22)	0.396*** (.141)	[0.000, 0.668] (0.000, 0.866)	[0.000, 0.402] (0.000, 0.531)
College Grad vs Some College	8.43 (2.12)	0.592*** (.140)	[0.000, 0.953] (0.000, 1.149)	[0.000, 0.731] (0.000, 0.921)
College Grad vs HS Grad	8.05 (2.22)	0.988*** (.138)	[0.000, 1.121] (0.000, 1.343)	[0.119, 0.873] (0.000, 1.049)
College Grad vs HS Dropout	6.92 (2.39)	1.536*** (.165)	[0.000, 1.537] (0.000, 1.832)	[0.268, 1.598] (0.127, 1.914)

Notes: All outcomes were adjusted for age, gender, year of birth and race by calculating the residuals and adding back the global mean. Robust standard errors in (.) in column 2. In columns 3 and 4 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. The min and max values of the residuals were used in computing the bounds.