Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Should Children Do More Enrichment Activities? Leveraging Bunching to Correct for Endogeneity

Carolina Caetano, Gregorio Caetano, and Eric Nielsen

2020-036

Please cite this paper as:
Caetano, Carolina, Gregorio Caetano, and Eric Nielsen (2020). "Should Children Do More Enrichment Activities? Leveraging Bunching to Correct for Endogeneity," Finance and Economics Discussion Series 2020-036. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.036.

Should Children Do More Enrichment Activities? Leveraging Bunching to Correct for Endogeneity

Carolina Caetano[1], Gregorio Caetano[1], and Eric Nielsen[2*]

[1]University of Georgia
[2]Federal Reserve Board

April 2020

Abstract

We study the effects of enrichment activities such as reading, homework, and extracurricular lessons on children's cognitive and non-cognitive skills. We take into consideration that children forgo alternative activities, such as play and socializing, in order to spend time on enrichment. Our study controls for selection on unobservables using a novel approach which leverages the fact that many children spend zero hours per week on enrichment activities. At zero enrichment, confounders vary but enrichment does not, which gives us direct information about the effect of confounders on skills. Using time diary data available in the Panel Study of Income Dynamics (PSID), we find that the net effect of enrichment is zero for cognitive skills and negative for non-cognitive skills, which suggests that enrichment may be crowding out more productive activities on the margin. The negative effects on non-cognitive skills are concentrated in higher-income students in high school, consistent with elevated academic competition related to college admissions. JEL Codes: I21, I2, J01, C24. Keywords: cognitive skills, non-cognitive skills, bunching, enrichment, homework, college, time use, skill development.

1 Introduction

Families spend substantial resources on activities intended to increase children's skills. These "enrichment" activities include homework, tutoring, reading, and extra-curricular

*We thank Michael Ahn and Hannah Hall for excellent research assistance. We also thank seminar participants for feedback, and in particular Matthew Larsen and Peter Hinrichs for very helpful comments as discussants. Hao Teng provided invaluable help with the time diary data. The analysis and conclusions set forth here are those of the authors and do not indicate concurrence by other members of the research staff, the Board of Governors, or the Federal Reserve System. Carolina gratefully acknowledges the support of the Bonbright Foundation.

activities such as music and art lessons. The money and time committed to these activities are substantial and increasing across the socioeconomic spectrum, leading to concerns that they may contribute to cross-sectional and intergenerational inequality (Aguiar and Hurst, 2007; Bianchi, 2000; Ramey and Ramey, 2010; Duncan and Murnane, 2011; Doepke and Zilibotti, 2017, 2019).

Enrichment activities have opportunity costs that go beyond the time and money spent by parents. The time and energy of the child are also limited – an hour spent doing homework is an hour not spent on other activities, such as play or sleep. Moreover, time spent on enrichment could have spillover effects into the remainder of the day. For example, an exhausted child may not want to engage in active play after finishing their homework, preferring more passive activities. A child over-stimulated by an after-school activity may fall asleep later than usual. Yet sleep and play are activities that have direct, positive impacts on skills (Walker, 2017; Gray, 2019). The opportunity costs of enrichment activities might therefore be substantial depending on the activities replaced.

This paper estimates the net effect of enrichment activities on cognitive and non-cognitive skills taking the substitution patterns among different activities, and their potential effects on skills, into account. Using time diary data from the Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID), we find that spending more time on enrichment activities yields a zero net effect on cognitive skills and a sizeable, negative net effect on non-cognitive skills.

Our results appear to contradict the positive effects of various forms of enrichment activities on both cognitive and non-cognitive skills found in the child development literature (e.g. Todd and Wolpin, 2007; Bernal and Keane, 2011; Fiorini and Keane, 2014; Caetano et al., 2019). However, relative to this prior literature, we are effectively identifying a different parameter. In order to control for confounders in the estimation of the effect of enrichment on skills, researchers often rely on detailed model specifications with many control variables. These include variables that may be influenced by enrichment, and hence be post-determined (e.g., time spent on other activities, family expenditures, other family investments). Thus, for example, if we were to control for the amount of time spent on active play, we would be shutting off one of the indirect paths through which homework could negatively impact skills. The estimated positive effects of enrichment when we include other activities as controls would thus reflect only the direct effect of enrichment on skills, but substitution and negative spillovers could counteract these positive effects, possibly even yielding a net negative total effect.

We propose a novel method to control for confounders that allows us to avoid adding post-determined variables entirely. Our method makes use of the fact that enrichment time cannot fall below zero and that many children in our data bunch at this lower limit. We argue that the choice of enrichment is the outcome of a constrained optimization problem that depends on both observed covariates and unobserved confounders, with the constraint that chosen enrichment be non-negative. The group of children at zero enrichment includes those for whom the constraint is barely binding, and those for whom the constraint is binding with great intensity. Consequently, the children who choose zero enrichment are different and more heterogeneous in comparison with the children who choose just a few minutes of enrichment per week.

Consistent with this idea, we show that the children who chose zero enrichment are discontinuously different in every observable way from the children who chose just above zero. Further, we present evidence that the same discontinuities also hold for the unobservables. This creates an opportunity around the bunching point, because although the unobservables are discontinuous there, the treatment itself is not – a few minutes of enrichment per week is not very different from zero minutes of enrichment. Thus, when we control for observables and then compare the skills of the children at zero enrichment and the skills of the children just barely above zero, the difference uncovers the direct effect of the unobservables on skills. We can use this information to build a correction for the selection bias.

Applying our method to the PSID time diary data, we find that the net effect of enrichment on cognitive skills is approximately zero, and that the net effect of enrichment on non-cognitive skills is quite negative and significant. These results are robust to several alternative definitions of enrichment time, alternative constructions of cognitive and non-cognitive skills, and various other sensitivity analyses. Breaking our results down by the child's grade in school, we find that the cognitive effects are also around zero in all grades, while the negative non-cognitive effects are concentrated entirely in high school.

To rationalize these results, we present a simple model arguing that if enrichment time is chosen so as to maximize cognitive skills, we would expect that a marginal increase in enrichment would yield zero net return to cognitive skills. Intuitively, optimality requires substituting enrichment for other activities up to the point where all activities have equal marginal returns to cognitive skills, so that the net effect of a marginal increase in enrichment on cognitive skills would be zero. We argue that plausible deviations from this stylized assumption will still yield cognitive estimates gravitating around zero, which is what we find.

Moreover, it is not generally possible to maximize both cognitive and non-cognitive skills at the same time. Thus, the level of enrichment that maximizes cognitive skills may be past the optimum for non-cognitive skills, leading to negative non-cognitive returns on the margin. Indeed, we find that the composition of enrichment shifts in later grades to activities that may come at the direct expense of non-cognitive skills. Whereas children in earlier grades spend relatively more of their enrichment time on activities with a social component, children in high school spend almost all of their enrichment time on homework, which may generate a sharper trade-off between cognitive and non-cognitive skills, thus explaining why we find negative non-cognitive estimates only for high school.

Further breaking the high school estimates down by household income, we find that the negative non-cognitive effects are particularly large for middle- and high-income youth. The large negative returns for high-income youth may be explained by the higher amount of time that group spends on enrichment, coupled with diminishing returns to enrichment on non-cognitive skills. The even larger negative returns for middle-income youth may be explained by substitution patterns, since enrichment comes at the expense of social activities for this group.

Our results highlight the pitfalls and trade-offs associated with intensive investment in children's human capital. The perception that such activities have high returns drives these investments. Many families strain to invest in an effort to increase the chance of admission to college. The stress that these high and rising investments place on both parents and children is well documented in the child development literature (e.g. Luthar and Becker, 2002; Luthar, 2003; Villaire, 2003; Ginsburg et al., 2007; Gray, 2011; Jarvis et al., 2014; Veiga et al., 2016), has been the subject of many books (e.g. Rosenfeld and Wise, 2000; Anderegg, 2003; Lareau, 2003; Warner, 2005; Gray, 2013; Abeles, 2015; Lukianoff and Haidt, 2018), and has been widely covered in the popular press (see Gray, 2010; Rosin, 2015; Rosen, 2015; Khazan, 2016; Avent, 2017 for recent examples). Yet, we find that many children are spending so much time on enrichment that, on the margin, they are actively harming their non-cognitive skills. This is particularly relevant given the widespread evidence of the importance of non-cognitive skills for key economic outcomes later in life (e.g. Heckman and Rubinstein, 2001; Heckman et al., 2006; Waddell, 2006; Lindqvist and Vestman, 2011; Deming, 2017).

The rest of the paper is organized as follows. Section 2 presents the data. Section 3 presents our identification strategy, followed by the results in Section 4. Section 5 discusses these results. Finally, Section 6 concludes.

2 Data

We use data from the Panel Study of Income Dynamics (PSID) and the 1997, 2002 and 2007 waves of the Child Development Supplement (CDS). The CDS data contain detailed time diary data and extensive measures of cognitive and non-cognitive skills, and is one of only two datasets that can be used for our study.[1] We link the CDS with the PSID, which allows us to build controls related to child, family and environmental characteristics.

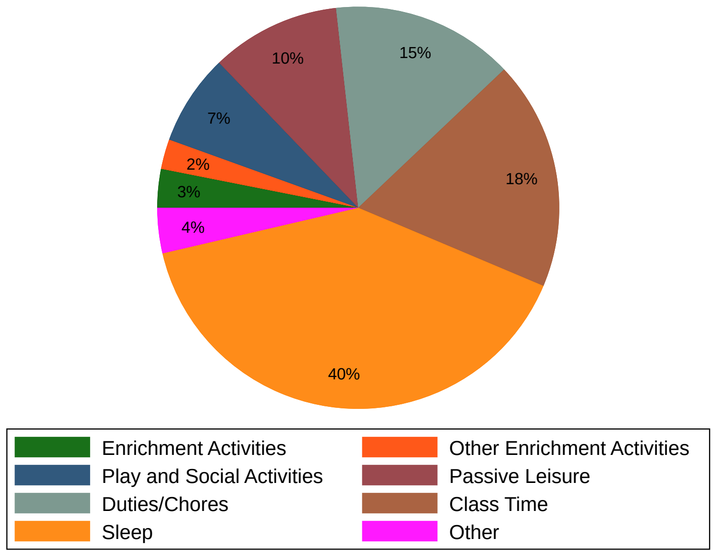
The time diaries in each CDS wave collect data on the full 24-hour breakdown of one random weekday and one random weekend day for each child. The child's activities during the selected days are coded into one of over 300 different categories reported by the child, or by the parent if the child is young, with subsequent editing and help from the PSID interviewer. We exclude cases where the day is described as non-typical, either the weekday or weekend day data is missing, or where the diary does not cover the full 24 hours. However, when the time slots between 10 p.m. and 6 a.m. are missing we do not exclude the observation and instead record that time as "sleeping," consistent with prior literature (Fiorini and Keane, 2014; Caetano et al., 2019). Finally, we aggregate the 300+ primitive time-use categories into eight categories: enrichment activities, other enrichment activities, play and social activities, passive leisure, duties/chores, class time, sleep, and other. Figure 1 shows the proportional breakdown of time among these categories.

Our definition of enrichment intends to capture the kinds of activities that are typically considered to be investments in children's skills. Therefore, our baseline measure includes only those activities that are unambiguously related to skill development over and above class time in school. In a typical week, children on average spend about 3% of their time on this type of enrichment, or roughly 5 hours/week. Figure 2 shows the breakdown of enrichment activity into various sub-categories.

The primary component of this baseline measure is homework, at two-thirds of the total. The next most important component of enrichment is reading a book, at 14% of the total. While 7% of enrichment time is spent on before- or after-school programs, relatively little is spent on each of the remaining categories: other reading (e.g., magazines

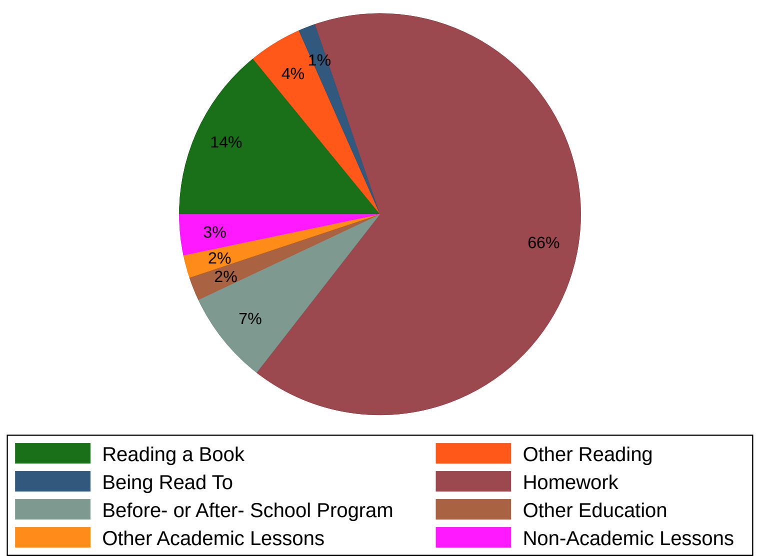
[1] To our knowledge, the only other dataset that has cognitive and non-cognitive skill measurements as well as time inputs spanning the entire day is the Longitudinal Study of Australian Children (LSAC). However, the CDS data contains more detailed time-use information, which is used for the precise categorization of the activities and the study of substitution patterns.

Figure 1: Daily Time Breakdown



Note: Panel plots the average division of enrichment time into different sub-categories over a typical week. The figure pools the 1997, 2002 and 2007 CDS waves.

Figure 2: Enrichment Time Breakdown



Note: Panel plots the average division of enrichment time into different sub-categories over a typical week. The figure pools the 1997, 2002 and 2007 CDS waves.

and newspapers), being read to (e.g., by parents), other academic lessons (e.g., tutoring, academic courses and lectures), non-academic lessons (e.g., piano and soccer lessons), and other education (e.g., driving lessons, military training).

As a robustness check, we also extend the notion of enrichment by including activities that are sometimes considered enrichment but which do not have such a clear connection to academic skills or human capital as traditionally conceived. This extended measure, which we label “broad enrichment,” includes our standard notion of enrichment plus “other enrichment:” making art/music, visiting museums, organized (structured) sports, volunteer work, the educational use of computers, and so forth. Figure 11 in Appendix A presents the breakdown of “other enrichment" into its constituent pieces, demonstrating that about two-thirds of the category is organized sports.

We also define a number of other time aggregates which we will use in Section 5 to assess substitution patterns between enrichment and other activities. Their breakdown can be seen in Figure 11 in Appendix A. First, we define “passive leisure" as activities that do not involve active, face-to-face social participation (e.g., any screen time, computer games, etc.) Two-thirds of passive leisure consists of watching TV. “Play and social activities," by contrast, consists of sports (not through school or in an organized league), social interactive games (e.g., board games, hide and seek), hobbies, socializing, social and church groups, etc. A little less than half of the time spent on this category is spent on social interactive games. We define “duties and chores" as all necessary, non-leisure and non-school activities such as household chores, paid work, travel (e.g., commuting, errands), shopping, personal care (hygiene, medical care, etc.), and meals. Traveling, meals and personal care take the most time within this category. “Class time" is defined as time at school for enrolled children and daycare or nursery care for children not in school. “Sleep" is defined as sleep at night, naps, and, as explained above, missing time slots between 10 pm and 6 am. Altogether, these time use categories are mutually exclusive and exhaustive.

We create our primary cognitive skill measure by applying iterated principle factor analysis to the standardized letter-word, applied problems, and passage comprehension subtests of the Woodcock Johnson Revised Tests of Achievement, Form B, which are available in each CDS wave. We likewise construct our non-cognitive skill measure through iterated principle factor analysis applied to parental assessments captured in 36 questions on the child's behavior. The loading factors for these scales are shown in Table 7 in Appendix A.[2]

Table 1 presents summary statistics for our sample. We have a pooled sample of 4,330 children ranging from 5 to 18 years of age, with an average age of just under 12. While children in our data spend on average a little over five hours per week on enrichment activities, about 30% do not spend any time at all on enrichment. About 40% of the children are black and about 7% are Hispanic. Further, 26% of the children in our sample attend a gifted program, 8% attend a special education program, 1% are home schooled, and 8% attend a private school. Throughout the paper, we classify children as low-, middle- or high-income if their household income falls in the bottom, middle, or top of the sample income terciles, respectively.

We denote by X the vector of observed child, family and environmental characteristics that we use as controls. Care is needed in the specification of X because many of the potential control variables available in our data are likely to be post-determined, and, as discussed in the introduction, including them would change the meaning of our estimates. Our approach therefore is to use only controls that are unambiguously pre-determined. We are able to adopt this parsimonious set of controls because our identification strategy can handle bias stemming from confounding unobservables. Our list of controls includes child's age and squared age (in months), and indicators for: CDS wave (1997, 2002 and 2007), grade (thirteen variables, from kindergarten through grade 12), gender, ethnicity (black, Hispanic and other non-white ethnicity), whether the child has siblings, family income tercile, whether the mother is alive, and whether the father is alive.[3] As a robust-

[2]For robustness, we also use the internalizing and externalizing subscales of the behavior problems index (BPI), a standardized scale included in each CDS wave, as alternative measures of non-cognitive skills. The internalizing scale captures the prevalence of withdrawn behaviors, while the externalizing scale captures outwardly aggressive behaviors (Peterson and Zill, 1986). Also for robustness, we use each component of our cognitive skill measure (applied problems, letter word, and passage comprehension) as separate measures of cognitive skill. Our cognitive and non-cognitive measures are all constructed so that a higher score is better and are all normalized to have a mean of zero and a standard deviation of one.

[3]For some of these control variables, some observations have a missing value (less than 1% of the sample).

Table 1: Summary Statistics

Activities (hours per week)	Mean	Standard Deviation
Enrichment	5.22	6.00
Other Enrichment	4.03	6.21
Play and Social Activities	12.30	10.19
Passive Leisure	17.48	11.94
Duties/Chores	24.68	11.29
Class	30.96	10.78
Sleep	67.20	9.19
Other	5.34	9.08
Other Variables		
Enrichment=0	0.29	0.45
1997 Wave	0.26	0.44
2002 Wave	0.46	0.50
2007 Wave	0.28	0.45
Child is in Grade PreK-5	0.31	0.46
Child is in Grade 6-8	0.33	0.47
Child is in Grade 9-12	0.37	0.48
Child is Female	0.50	0.50
Child is White	0.48	0.50
Child is Black	0.40	0.49
Child is Hispanic	0.07	0.26
Child Has Siblings	0.88	0.33
Child is Low-Income	0.33	0.47
Child is Middle-Income	0.33	0.47
Child is High-Income	0.33	0.47
Child's Father is Alive	0.97	0.16
Child's Mother is Alive	0.99	0.08
Child is in Gifted Program	0.26	0.44
Child is in Special Education Program	0.08	0.27
Child is Home Schooled	0.01	0.11
Child is in Private School	0.08	0.27
Age (years)	11.86	39.85

Note: N=4,330. Activity categories are exhaustive. The 1997, 2002 and 2007 CDS Waves are pooled.

ness check, we also estimate alternative specifications where we add as controls some additional variables that may be post-determined, such as whether the child is in a gifted

In these cases, we include the missing observations in our sample by assigning them a unique value for the relevant control variable and creating an indicator variable for whether that observation had a missing value for that control. We then include these indicators as additional controls. The resulting estimates are very similar to the case where we simply drop all observations with any missing control variables.

program, whether the child is in a special education program, whether the child is home schooled, and whether the child attends a private school. Adding these controls barely changes our estimates. Importantly, we do not include time spent on other activities as controls, since these are determined jointly with enrichment time.

We also do not include lagged test scores as controls, even though these are commonly included in the child development literature. Including lagged controls reduces substantially our sample size, as the child would need to be observed in consecutive waves of the CDS. This would also substantially restrict the age range of the children that we can use in the sample, thus not allowing for the breakdown by grade range that we present below. Moreover, our correction strategy renders the use of lagged skills to control endogeneity less important. Indeed, the ability to credibly estimate causal effects without the use of lagged scores is an advantage of our approach.

3 Identification Strategy

Consider the standard outcome equation

$$S = \beta I + h(X) + \epsilon, \tag{1}$$

where S refers to either cognitive or non-cognitive skill, I refers to enrichment time (I stands for "investment," since enrichment activities are generally undertaken as investments in human capital), X is a vector of observed pre-determined controls, and ϵ is the unobservable error term.

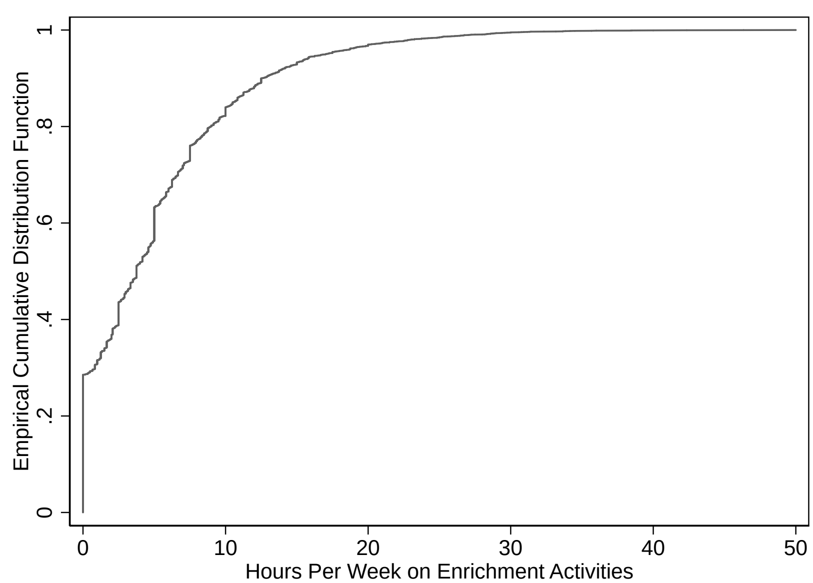
If we ignore any potential endogeneity problem and simply regress Y onto I and X, the coefficient of I may be biased. Indeed, in Appendix B, we apply Caetano (2015)'s test of exogeneity and show that the evidence of endogeneity in the equation above is overwhelming. Moreover, we show evidence there that the bias in the estimation of β is positive, so that a simple regression of S on I and X would over-estimate the effect of enrichment on skills.

Without additional assumptions, the effect of enrichment in the equation above, β, cannot be identified. However, the enrichment variable I is of a peculiar nature that can be leveraged to identify β. Specifically, in Section 3.1 we argue that a substantial fraction of the sample would have chosen a negative amount of enrichment if it were possible but were instead constrained to choose zero. Section 3.2 uses this information to incorporate some structure into the model. Finally, in Section 3.3 we show that β can be identified in the augmented model.

3.1 Enrichment is a constrained choice.

Figure 3 plots the empirical cumulative distribution function of enrichment time in our sample and shows that there is substantial bunching at zero. About 30% of children spend no time on enrichment, while the rest are continuously distributed among different, positive levels.

Figure 3: Evidence of Bunching at Zero Enrichment Time



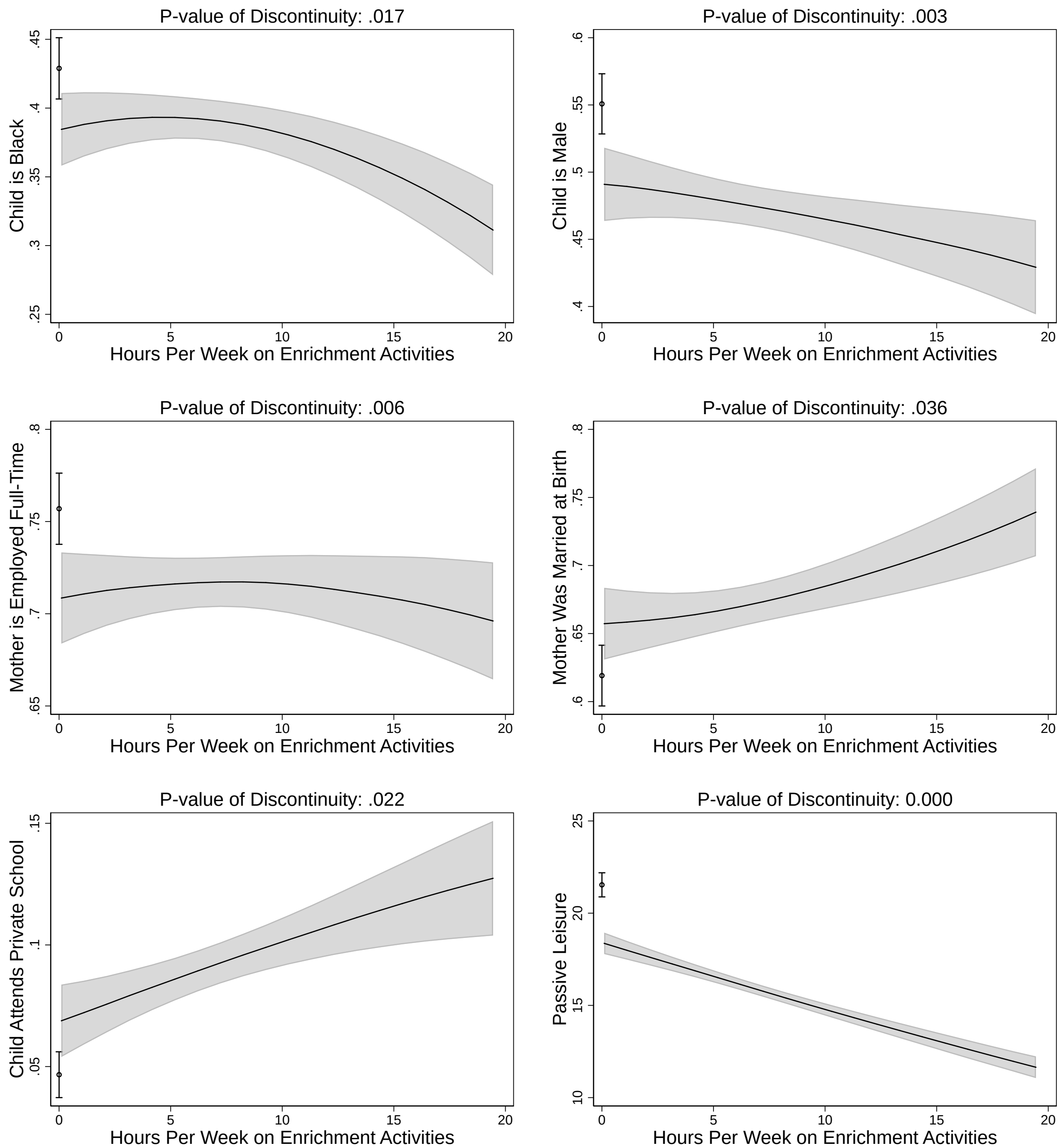
Note: Figure plots the cumulative density function of time spent per week on enrichment activities (in hours) for our full sample.

Why does this bunching happen? To answer this, first note that the children who spend no time on enrichment are discontinuously different in every observable way from the children who spend any positive time at all on enrichment. Figure 4 shows some examples of these discontinuities. The upper left panel of the figure shows a local linear fit of an indicator of whether the child is black conditional on the amount of time the child spends on enrichment, as well as the proportion of children who are black among the children who spend zero time on enrichment. The children at zero are discontinuously more likely to be black than the children who spend marginally positive amounts of time on enrichment. In the header of the panel, we show the p-value of a test of whether the share of black children is continuous at zero enrichment time, and it is clear that we can confidently reject this hypothesis ($p = 0.017$). The other panels of Figure 4 show similar patterns. Children who spend no time on enrichment are discontinuously more likely to be male ($p = 0.003$), to have a mother who works full-time ($p = 0.006$), to have an unmarried mother at birth ($p = 0.036$), to not be enrolled in a private school ($p = 0.022$) and to spend time on passive leisure activities ($p = 0.000$). That is, in each case, we find that the children at zero seem to be negatively selected on observables associated with higher expected achievement.

The last panel in Figure 4 reflects the stark differences between the lives of the children who spend no time on enrichment and everybody else. In particular, note that the children at zero enrichment spend on average four more hours per week on passive leisure than the children at one hour of enrichment. Since the total number of hours in a week is the same for everyone, this means that, relative to the group of children who spend one hour of enrichment, the group of children at zero enrichment is spending one fewer hour on enrichment and three fewer hours in other activities, potentially more productive than passive leisure, such as play, socializing, and sleep.

In Appendix B, we present evidence that this pattern of negative selection at zero-enrollment is also true for unobservables. This pattern – that every characteristic of the child, observable and unobservable, is so starkly different at zero – can be naturally

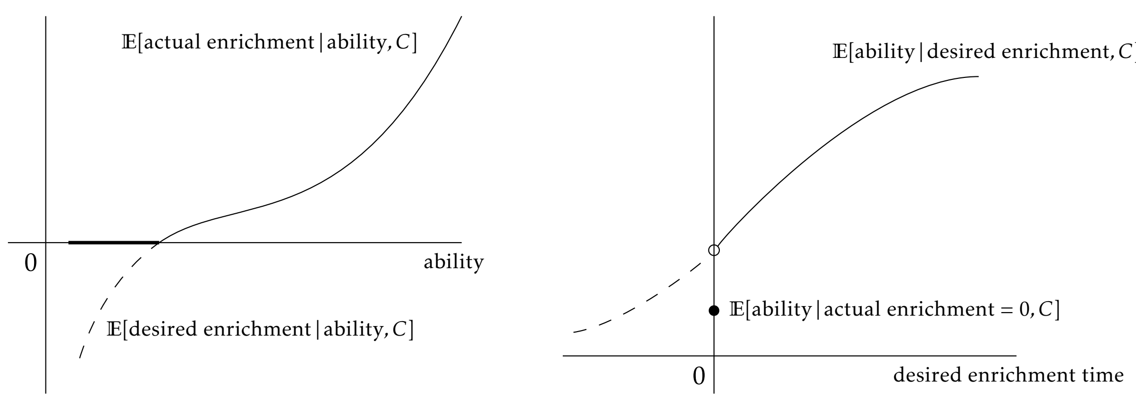
Figure 4: Evidence that Bunching is Selective



Note: Each panel shows a plot of the local linear estimator of the expected value of a variable conditional on enrichment time, along with its 90% confidence interval. The expected value of the variable among the children who spent no time on enrichment is also shown, along with its 90% confidence interval. Finally, the p-value of a test for whether there is discontinuity at zero is shown in the header of each panel.

explained if enrichment is a choice that is constrained to be non-negative. Let us work with this idea. There are two types of enrichment: "desired enrichment," which is the amount a person would like to choose absent the non-negativity constraint, and "actual enrichment," which is the amount they actually choose. In Figure 5 we explore how a single unobservable variable, say "ability," is mapped to this choice. Suppose that we vary ability, but keep every other characteristic fixed (we will denote these other observable and unobservable characteristics as C in the plot). For every level of ability we expect a certain level of desired enrichment. We suppose that higher levels of ability are related to higher levels of desired enrichment, as depicted in the left panel. Whenever desired enrichment is positive, the constraint is not binding, and thus the desired and actual enrichment curves coincide. However, as we move to lower ability levels, the desired enrichment may be negative, as shown in the dashed curve. Meanwhile the actual enrichment choice cannot be negative, and thus the two curves separate. All those who desire a negative amount of enrichment choose actual enrichment equal to zero.

Figure 5: Relationship Between Child's Ability and Enrichment Time



Note: In the left panel, the solid line denotes actual (chosen) enrichment, which is equal to desired enrichment when desired enrichment is non-negative. For negative desired enrichment (dashed line), actual enrichment must be zero. The right panel inverts this relation, showing that this constraint will generate a discontinuity in the expected characteristics of children who do zero enrichment, since that group includes all the children for whom the constraint is binding. C represents all other characteristics that determine enrichment (observed or unobserved).

We can then look at this relationship in an inverse way, by plotting in the right panel the expected ability for every level of enrichment. As explained in the left panel, the group of children who choose zero enrichment include all those for whom the constraint that enrichment must not be negative was binding, and thus the expected ability for these children should reflect the fact that they are very selected and very different from their counterparts who chose marginally positive amounts of enrichment. Indeed, the figure shows that the average ability for those choosing zero enrichment, the solid black dot, is discontinuously lower than the average ability for those choosing small, positive levels of enrichment.

3.2 A model with constrained enrichment choice

Given the discussion in the previous section, we understand that desired enrichment, denoted by I^*, is a function of characteristics both observable, X, and unobservable, η, in the following equation

$$I^* = g(X) + \eta. \tag{2}$$

This equation alone is not an assumption. First, we are not specifying g in any way. In fact, one can think of η as simply the residual of I^* once $g(X)$ is taken out, for any given function g. We are not saying that X and η have a causal relationship with I^*, and we do not require any independence between η and X. We do not even suppose that g is identifiable.

We now introduce some structure based on the discussion of the previous section. We suppose that the actual choice of enrichment is constrained:

$$I = \max\{0, I^*\}. \tag{3}$$

Finally, we add additional structure by opening the error term in equation (1) as $\epsilon = \delta\eta + \varepsilon$:

$$S = \beta I + h(X) + \delta\eta + \varepsilon, \quad \mathbb{E}[\varepsilon|I, X] = 0. \tag{4}$$

This equation makes two assumptions. First, it assumes that all the unobservable confounders are indexed by η. The unobservable ε can be simply understood as the residual $S - \mathbb{E}[S|I, X, \eta]$, which represents the remaining independent heterogeneity. This is a selection on unobservables model which assumes that I is exogenous only if we also condition on the unobservable term η.

Second, and this is our main identifying assumption, equation (4) assumes that η enters linearly, as opposed to entering through a more general nonparametric form such as $f(\eta, X)$. Appendix C discusses this assumption in depth, and provides evidence that it does not seem to play an important role in our empirical conclusions.[4]

3.3 Identifying β

Our model is therefore composed of equations (2), (3) and (4), which together imply

$$\mathbb{E}[S|I, X] = (\beta + \delta)I + \underbrace{h(X) - \delta g(X)}_{m(X)} + \delta \mathbb{E}[I^*|I^* \leq 0, X]\mathbf{1}(I = 0). \tag{5}$$

Equation (5) shows that the expected skill conditional on covariates is discontinuous at zero enrichment. To understand this discontinuity, consider the group of children who chose $I = 0$. Why do we see skill variation in this group? The variation is not due to differences in time spent on enrichment (the first term of the equation) because $I = 0$ for everyone in this group. Part of the variation in skills is explained by variation in the controls, specifically through the term $h(X) - \delta g(X)$. However, even if we condition on the controls, there is further variation in skills due to the differences in the unobservable confounder, η. From equation (2), conditional on X, the variation in η is identical to the variation of I^*.

Thus, if somehow we could identify $\mathbb{E}[I^*|I^* \leq 0, X]$, we could identify δ by relating the variation in skills and the variation in $\mathbb{E}[I^*|I^* \leq 0, X]$ among those who chose zero

[4]Equation (4) also seems to make assumptions about the observables, i.e. linearity in I and separability in X. In reality this model and our identification strategy allow for heterogeneous treatment effects (see Caetano et al. (2020)) and thus the true effect of I on S may be of the general form $\beta(I, X, \varepsilon)$, and the effects reported can be interpreted as averages. In any case, in Appendix C, we find that the treatment effects seem to be uncorrelated with I.

enrichment. We could then extrapolate by assuming that this bias is the same for those who spend positive amounts of time on enrichment, as in the first term of equation (5).

Explicitly, let us rewrite equation (5) as

$$\mathbb{E}[S|I,X] = \beta + m(X) + \delta\left[I + \mathbb{E}[I^*|I^* \leq 0, X]\mathbf{1}(I=0)\right]. \tag{6}$$

Then, if we could identify $\mathbb{E}[I^*|I^* \leq 0, X]$, we could implement a correction by adding the term $I + \mathbb{E}[I^*|I^* \leq 0, X]\mathbf{1}(I=0)$ to the regression as another control. As long as $\mathbb{E}[I^*|I^* \leq 0, X] < 0$ for some values of X in the data, the correction term $I + \mathbb{E}[I^*|I^* \leq 0, X]\mathbf{1}(I=0)$ will be linearly independent of I (note that $\mathbb{E}[I^*|I^* \leq 0, X]\mathbf{1}(I=0)$ is orthogonal to I). This allows us to identify β and δ separately.

How can $\mathbb{E}[I^*|I^* \leq 0, X]$ be identified? Although we do not observe the latent enrichment choice I^* when it is negative, we do observe it when it is positive, since $I^* = I$ when $I > 0$. Our strategy then is to use observations with $I > 0$ to make an out-of-sample prediction of the average desired investment I^* when $I^* \leq 0$. Specifically, we can make assumptions about the shape of the distribution of the confounders η, and relate it to the shape of I^* through equation (2). We explore three assumptions, which are nested and ordered from strongest to weakest:

1. **Tobit:**

   $$\eta|X \sim \mathcal{N}(X'\theta, \sigma^2)$$

   and $m(X) = X'\gamma$.

2. **Heteroskedastic Tobit:**

   $$\eta|X \sim \mathcal{N}(l(X), \sigma^2(X)),$$

   which drops both the linearity of m and of the mean, as well as the homoskedasticity requirements, keeping only the normality assumption.

3. **Heteroskedastic Tail Symmetry:** for all censored quantiles q_0,

$$\eta|X \text{ has symmetric tails below } q_0 \text{ and above } 1-q_0,$$

which drops the normality assumption but keeps the symmetry between the constrained part of the distribution and the corresponding upper tail.

In Appendix D we show how each of the three assumptions can be leveraged to identify and estimate $\mathbb{E}[I^*|I^* \leq 0, X]$, and how each of these assumptions fit our data. To summarize, there is strong evidence of heteroskedasticity, which means that the Tobit assumption is likely not flexible enough. The heteroskedastic Tobit assumption fits the data quite well, although there is some evidence that the tails of the empirical distributions are fatter than normality implies. The heteroskedastic tail symmetry assumption seems to solve this issue. Irrespective of this evidence, below we show the results for corrections based on all three assumptions, and our conclusions hold for all three cases.

4 Empirical Results

4.1 Full-Sample Estimates

Table 2 presents our main results estimated on the full sample. Column (i), which shows the results of simple regressions of skills on enrichment time without controls (equation (6) without either $m(X)$ or the correction term), demonstrates that both cognitive and non-cognitive skills are strongly positively correlated with enrichment time. Column (ii), which adds controls back into the specifications in column (i) (equation (6) without the correction term), shows that while observables seem to explain part of the correlation between enrichment time and skills, the residual relationships remain positive, particularly for cognitive skills.

The discontinuity plots shown in Figure 4, as well as the evidence discussed in Ap-

Table 2: Full-Sample Results: The Effect of Enrichment Time on Skills

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Cognitive	β	0.018**	0.011**	-0.004	-0.007	-0.002
		(0.003)	(0.002)	(0.006)	(0.006)	(0.006)
	δ			0.013**	0.015**	0.010**
				(0.005)	(0.004)	(0.005)
Non-Cognitive	β	0.006**	0.003	-0.015	-0.024**	-0.019*
		(0.003)	(0.003)	(0.010)	(0.009)	(0.010)
	δ			0.015*	0.022**	0.018**
				(0.008)	(0.007)	(0.008)

Note: N=4,330. Bootstrapped standard errors in parentheses (500 iterations). The corrected specifications use 50 clusters (see Appendix D for estimation details and Figure 21 in Appendix E for analogous results with different numbers of clusters.) ** $p<0.05$, * $p<0.1$.

pendix B (see Figure 14 there), suggest that the uncorrected estimates in columns (i) and (ii) are positively biased. The remaining columns in Table 2 show our corrected estimates of β (equation (6)) under the different assumptions on the distribution of $\eta|X$ discussed in Section 3.3 ranging from the strongest to the weakest. Column (iii) shows the results when we implement the Tobit strategy to estimate $\mathbb{E}[I^*|I^* \leq 0, X]$, column (iv) shows the results under the heteroskedastic Tobit strategy, and column (v) shows the results under the heteroskedastic tail symmetry strategy.[5] All standard errors are bootstrapped using 500 iterations.

For cognitive skills, all of the corrected estimates are quite similar – the estimated β's fall from 0.011 standard deviations (s.d.) to around -0.004 s.d. The large differences between column (ii), where the estimate is positive and highly significant, and columns

[5]Note that implementing the heteroskedastic Tobit and heteroskedastic tail symmetry corrections requires that we discretize X in order to estimate $\mathbb{E}[I^*|X^* \leq 0, X]$. We discretize X using hierarchical clustering, in which observations are grouped into clusters based on the similarity of their observables. All the results reported in the paper use 50 clusters in the estimation of $\mathbb{E}[I^*|X^* \leq 0, X]$. Additionally, for the specification of $m(X)$ in equation (6) we use both the non-clustered control variables described in Section 2 as well as the cluster indicators. For details, please refer to Section E in the Appendix. We also show there that our results do not appear to be an artifact of either the particular way we discretize X, or the number of clusters.

(iii)-(v), where the estimates are negative and insignificant, show that our correction method is able to handle endogeneity which was not absorbed by the pre-determined controls. Our most general correction method (symmetry) yields a 90% confidence interval of $[-0.012, 0.008]$.

Correcting for selection has even more dramatic consequences for the non-cognitive estimates – the corrected non-cognitive β's are negative, with point estimates ranging between -0.024 and -0.015 s.d. The point estimate using our preferred method (column (v)) is -0.019, significantly different from zero at 10%. This is about three times larger in magnitude than the unconditional correlation (column (i)).

The non-cognitive estimates in Table 2 are also economically significant. To see this, consider two otherwise similar children: one who engages in zero enrichment and one who spends 12.5 hrs/week, putting her at the 90$^{\text{th}}$ percentile in the full-sample distribution. These 12.5 hours come at the expense of other activities the child could have done instead during that time. The preferred corrected estimates imply that the 90$^{\text{th}}$ percentile child would have 0.19 s.d. lower non-cognitive skills than the child at zero. This is a sizeable difference relative to what is often found in the child development and education literatures.[6]

For both cognitive and non-cognitive skills, the estimated δs are positive and highly significant, confirming the evidence we presented in Appendix B of large amounts of endogeneity bias in the uncorrected estimates. The fact that the β estimates in the "No Controls" (i) column are larger than in the "Uncorrected" (ii) column provides yet further evidence of positive bias.

Note that the standard errors from column (ii) of Table 2 are much smaller than the standard errors from the corrected models, (columns (iii)-(v)). This is a feature of our

[6] By way of comparison, the very sizable black-white gap in cognitive skills is generally found to be around 1 s.d. (Neal and Johnson (1996)). Effect sizes of -0.2 s.d. are large in magnitude relative to the literature on teacher value-added, which typically finds that a standard deviation increase in teacher quality corresponds to an increase of roughly 0.05-0.1 s.d. in student achievement (which relates to our measure of cognitive skills) or student behavior (which relates to our measure of non-cognitive skills). See for example Chetty et al. (2014); Kane and Staiger (2008); Kane et al. (2008); Jackson (2018).

approach, not a bug. The only difference between the corrected and uncorrected models is the presence of the generated regressor $\hat{\mathbb{E}}[I^*|I = 0, X = x]$. Adding one regressor will not generally cause the standard errors in a regression to blow up, so the fact that we see an increase in the standard errors in our application suggests greater underlying uncertainty surrounding the true causal effects of enrichment time on skills once endogeneity is accounted for. Not considering this correction term would lead to overly precise, biased estimates. In turn, this could lead to excessively optimistic and confident expectations of policymakers or families regarding the impact of enrichment activities.

Table 8 in Appendix A shows that our baseline results are robust to plausible alternative measures of cognitive and non-cognitive skills. First, we consider each of the components of our cognitive measure separately. For each component (applied problems, letter-word comprehension, and passage comprehension), we find sizeable, positive uncorrected estimates and statistically insignificant corrected estimates. Next, we consider alternative measures of non-cognitive skills based on the internalizing and externalizing subscales of the behavior problems index (BPI) included in the CDS. Here, the uncorrected estimates suggest significant, positive effects for externalizing problems only, while the corrected estimates for both scales are negative and similar in magnitude to the main non-cognitive estimates reported in Table 2.

Our results are also robust to alternative definitions of enrichment time. First, we consider broad enrichment, which expands the notion of enrichment to include additional activities less directly intended to the development of cognitive skills such as organized sports, arts, and volunteering (see Section 2 for details.) Table 9 in Appendix A shows that using this broader measure yields remarkably similar estimates to the baseline results presented in Table 2. The uncorrected estimates again show significant, positive associations between (broad) enrichment and skills, while the corrected estimates again indicate a null effect for cognitive skills and a significant negative effect for non-cognitive skills. Indeed, the corrected non-cognitive point estimate assuming symmetry is very similar to the baseline estimate and is significant at the 95% level. Conversely, when we do the opposite and restrict enrichment to consist only of homework, we find the same pattern of zero cognitive estimates and even more significant, more negative non-cognitive estimates (Table 10 in Appendix A).

4.2 Estimates by Grade

The full-sample estimates imply that enrichment time, when corrected for selection on unobservables, has roughly no effect on cognitive skills and a significant, negative effect on non-cognitive skills. Here, we break down these results by grade level by applying our method separately for children in different age ranges.

The estimates by grade are presented in Table 3. The uncorrected estimates show that each additional hour of enrichment is associated with a statistically and economically significant increase in cognitive skills for children in middle and high school. Yet, the corrected estimates are all around zero, with some weak evidence of negative effects for high school. The headline result for cognitive skills from the full-sample estimates in Table 2 carries over to each grade range separately: the corrected effect of enrichment on cognitive skills is roughly zero for all grade ranges.

Table 4 repeats the analysis for non-cognitive skills. The uncorrected estimates suggest a significant, positive association between enrichment and non-cognitive skills for high school only. Interestingly, this grade range happens to be exactly the one in which we find the most evidence of endogeneity, as seen by the estimates of δs in columns (iii)-(v). Indeed, the corrected estimates are negative and significant only for high school.

Table 3: Cognitive Estimates by Grade Levels

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
PreK-5	β	0.008*	0.000	0.003	0.002	-0.002
		(0.005)	(0.003)	(0.013)	(0.012)	(0.011)
N=1331	δ			-0.003	-0.002	0.002
				(0.012)	(0.011)	(0.009)
6-8	β	0.020**	0.009**	0.003	-0.001	0.001
		(0.003)	(0.002)	(0.011)	(0.011)	(0.011)
N=1414	δ			0.005	0.008	0.007
				(0.009)	(0.009)	(0.009)
9-12	β	0.027**	0.013**	-0.008	-0.009	-0.008
		(0.003)	(0.002)	(0.008)	(0.008)	(0.009)
N=1585	δ			0.016**	0.017**	0.017**
				(0.006)	(0.006)	(0.007)

Table 4: Non-Cognitive Estimates by Grade Levels

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
PreK-5	β	0.001	-0.001	0.030	0.026	0.023
		(0.005)	(0.005)	(0.024)	(0.024)	(0.021)
N=1331	δ			-0.027	-0.023	-0.020
				(0.021)	(0.022)	(0.018)
6-8	β	0.003	-0.003	0.005	0.000	-0.003
		(0.005)	(0.005)	(0.020)	(0.019)	(0.018)
N=1414	δ			-0.007	-0.003	0.000
				(0.016)	(0.016)	(0.015)
9-12	β	0.012**	0.010**	-0.035**	-0.040**	-0.039**
		(0.003)	(0.004)	(0.012)	(0.011)	(0.014)
N=1585	δ			0.035**	0.039**	0.039**
				(0.008)	(0.008)	(0.010)

Note (Tables 3 and 4): Number of observations (N) for each grade range is shown. Bootstrapped standard errors in parentheses (500 iterations). The corrected specifications use 50 clusters. ** p<0.05, * p<0.1.

4.3 High School Estimates By Household Income

We want to understand why the non-cognitive effects are negative for high school children. To this end, we break down the high school estimates based on the income of the child's household. Table 5 presents the cognitive estimates for high school children by household income tercile. The uncorrected estimates show a significant, positive association between cognitive skills and enrichment for each income tercile. By contrast, the corrected estimates are all negative and indistinguishable from zero.

Table 5: Cognitive Estimates by Income Tercile - Grades 9-12

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Low	β	0.024**	0.013**	-0.002	-0.007	-0.012
		(0.006)	(0.006)	(0.018)	(0.019)	(0.025)
N=468	δ			0.011	0.014	0.020
				(0.013)	(0.014)	(0.020)
Middle	β	0.027**	0.018**	-0.008	-0.006	-0.005
		(0.005)	(0.005)	(0.015)	(0.014)	(0.018)
N=529	δ			0.019*	0.017*	0.019
				(0.011)	(0.010)	(0.014)
High	β	0.015**	0.009**	-0.009	-0.009	-0.006
		(0.003)	(0.003)	(0.012)	(0.012)	(0.011)
N=580	δ			0.015	0.014	0.011
				(0.009)	(0.009)	(0.009)

Note: Number of observations (N) for each household income tercile is shown. Bootstrapped standard errors in parentheses (500 iterations). The corrected specifications include 50 clusters. ** p<0.05, * p<0.1.

Table 6 shows the analogous results for non-cognitive skills. The uncorrected estimates suggest some positive association between non-cognitive skills and enrichment. However, the corrected estimates uniformly indicate negative causal effects, particularly for middle- and high-income children. These negative effects are large. For instance, our preferred middle-income estimates (column v) are over twice the magnitude of the unconditional relationship between enrichment and non-cognitive skills (column i). Correspondingly, the estimated δs are large, positive, and are statistically significant for the top two income terciles, indicating strong positive selection into enrichment within each income group.

Table 6: Non-Cognitive Estimates by Income Tercile - Grades 9-12

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Low	β	0.020**	0.015*	-0.007	-0.008	-0.017
		(0.007)	(0.008)	(0.026)	(0.025)	(0.035)
N=468	δ			0.017	0.017	0.026
				(0.020)	(0.019)	(0.028)
Middle	β	0.022**	0.019*	-0.040	-0.059**	-0.060*
		(0.009)	(0.010)	(0.025)	(0.023)	(0.034)
N=529	δ			0.044**	0.057**	0.064**
				(0.017)	(0.016)	(0.027)
High	β	0.002	0.004	-0.034*	-0.030*	-0.028
		(0.005)	(0.005)	(0.018)	(0.018)	(0.018)
N=580	δ			0.030**	0.027*	0.025*
				(0.015)	(0.014)	(0.014)

Note: Number of observations (N) for each household income tercile is shown. Bootstrapped standard errors in parentheses (500 iterations). The corrected specifications include 50 clusters. ** p<0.05, * p<0.1.

In the next section, we discuss possible reasons why all of our causal estimates of enrichment on cognitive skills gravitate around zero, and why there seems to be a negative effect of enrichment on non-cognitive skills which is particularly concentrated in the high school years for middle- and high-income children.

5 Discussion

5.1 A Simple Model of Enrichment Time Allocation

At first sight, our empirical results are puzzling. How could it be that spending more time on reading, studying, extracurricular lessons, and other such activities does not improve cognitive skills? Further, how could the effects of these activities on non-cognitive skills be negative? Here we discuss one possible rationalization of these results using a stylized model of time allocation.

This model aims to capture two key ideas. First, our empirical approach measures the causal effect of spending more time on enrichment *relative to alternative uses of time*. Every additional hour that a child spends on enrichment is an hour not spent on some other activity. If the activities foregone in favor of enrichment have, on the margin, greater returns than enrichment, then the net effect of spending more time on enrichment will be negative. The second idea is that children and families will not generally be able to choose enrichment so as to simultaneously maximize both cognitive and non-cognitive skills. Thus, if improving cognitive skills tends to be the objective, the resulting choice of enrichment might be beyond the non-cognitive optimum, leading to negative net non-cognitive returns on the margin.

Suppose that the total time budget is normalized to 1 and there are only two possible activities: enrichment time I and leisure time $L = 1 - I$. Skills are produced from enrichment and leisure according to

$$S_c = f_c(I, L), \quad S_{nc} = f_{nc}(I, L), \tag{7}$$

where S_c and S_{nc} denote cognitive and non-cognitive skills, respectively.

5.1.1 How can the cognitive estimates be zero?

Consider a hypothetical, stylized scenario where families choose enrichment so as to maximize cognitive skills. Assuming differentiability and an interior solution, the optimal allocation of enrichment time for cognitive skills, I_c, will satisfy the first order condition

$$\frac{\partial}{\partial I} f_c(I_c, 1 - I_c) = \frac{\partial}{\partial L} f_c(I_c, 1 - I_c). \tag{8}$$

Equation (8) states that the optimal enrichment choice equalizes the marginal return to enrichment and leisure activities *for cognitive skills*. Intuitively, if the marginal return of an additional hour of enrichment is higher than the marginal return for leisure, then it is worth it to substitute one hour from leisure to enrichment. Therefore, if children and families chose enrichment so as to maximize cognitive skills, the marginal effect of enrichment on cognitive skill will be

$$\begin{aligned} \frac{dS_c}{dI} &= \frac{\partial}{\partial I} f_c(I_c, 1 - I_c) + \frac{\partial}{\partial L} f_c(I_c, 1 - I_c) \cdot \frac{d(1 - I)}{dI} \\ &= \frac{\partial}{\partial I} f_c(I_c, 1 - I_c) - \frac{\partial}{\partial L} f_c(I_c, 1 - I_c) \\ &= 0. \end{aligned}$$

At the optimum, the marginal effect of investment on cognitive skills should be zero. This result might explain why our causal cognitive estimates are close to zero in all cases. If children and families choose enrichment so as to maximize cognitive skills, causal estimates around zero are exactly what we would expect to find.

This example considers only two activities, enrichment and leisure. However, a similar result holds when there are more than two activities. This conclusion also continues to hold approximately even if some children bunch at $I_c = 0$, provided that sufficiently many choose interior solutions ($I_c > 0$).[7]

In practice, the exact returns of each activity are not precisely known, and thus en-

[7] In fact, in this scenario we would expect cognitive estimates to be slightly negative and small, exactly as we are finding them to be. Indeed, while the net cognitive returns to enrichment would be zero for those in interior solutions, it would be negative for those at the corner solution. If net cognitive returns to enrichment were not negative at zero for the bunched children, we would expect them to choose positive values of enrichment instead.

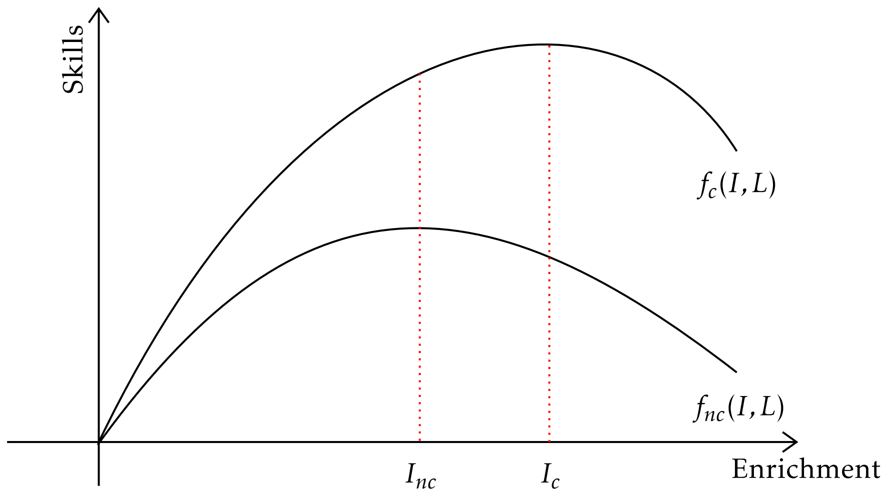
richment is likely not chosen to be exactly I_c. One may even wonder how realistic it is to suppose that families and children maximize cognitive skills when choosing enrichment. Over 70% of enrichment is composed of activities that strongly target academic cognitive skills (mainly homework). In such cases, grades and test scores are likely the main optimization objective, and these outcomes align closely with our definition of cognitive skills. However, the choice of other enrichment activities such as extra-curricular classes and reading may be more complex than the simple maximization of cognitive skills. For example, families may be optimizing something else entirely, such as college acceptance, or a combination of cognitive skills and other considerations, such as the enjoyment of the activity, caving to social pressure, or belonging to a social group. Nevertheless, the model conclusions will still hold approximately in all these cases provided cognitive skills are given enough consideration in the objective function of children and families.

5.1.2 How can the non-cognitive estimates be negative?

Next, we discuss why enrichment activities may yield a negative effect on non-cognitive skills, particularly for middle- and high-income children in high school. We begin by arguing that the optimal amount of enrichment for cognitive skills, I_c, will generally be different from the optimal amount of enrichment for non-cognitive skills, I_{nc}. Figure 6 explains this point. It plots hypothetical cognitive and non-cognitive production functions as a function of enrichment for later grades. The cognitive-maximizing point, I_c, lies to the right of the non-cognitive-maximizing point, I_{nc}. Around I_c, the marginal return to enrichment is close to zero for cognitive skills and negative for non-cognitive skills, reflecting the findings of Section 4 for high school.

To see this, consider someone who spends $I_c - 1$ hours on enrichment and is considering doing one extra hour of enrichment, say homework. Hypothetically, this extra hour of homework could have a positive direct effect on cognitive skills and a small negative indirect effect through the foregone substituted activities (e.g., play, sleep) for a total

Figure 6: Cognitive and Non-Cognitive Skills Production - High School



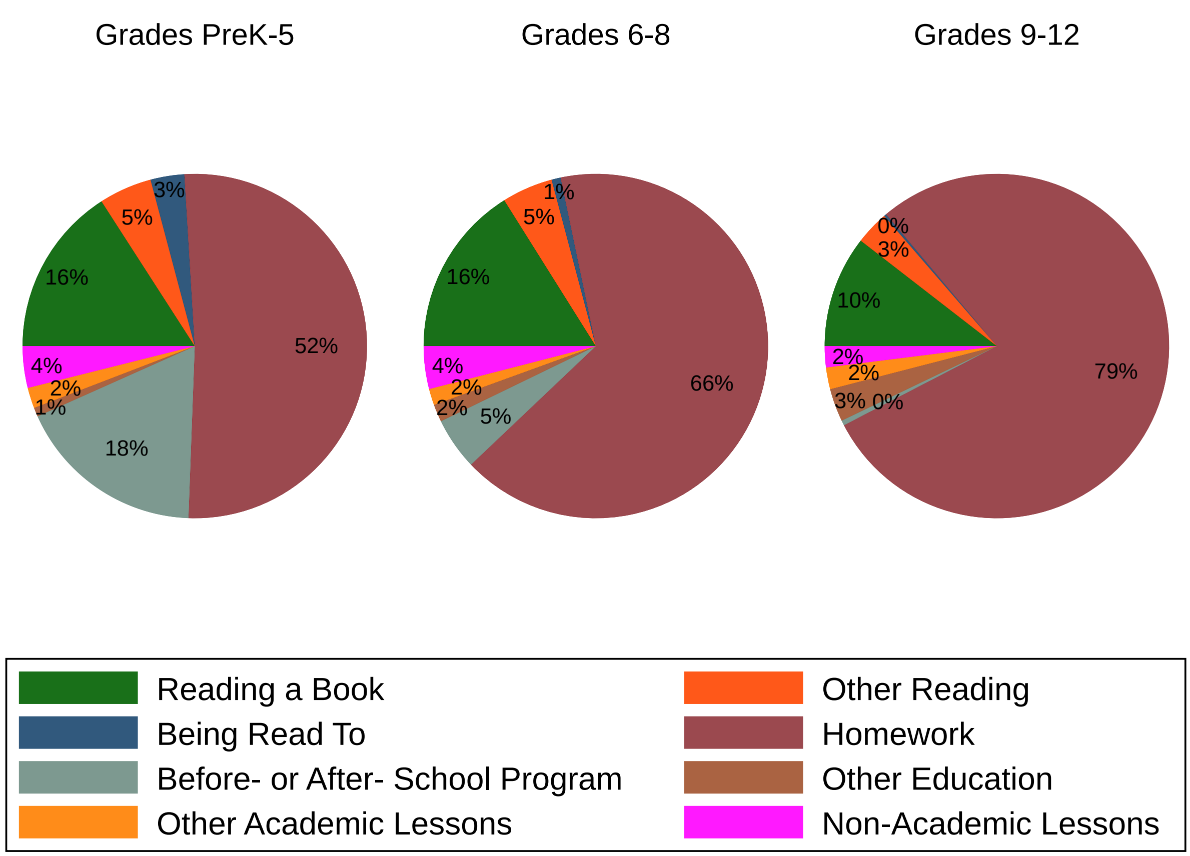
Note: This figure illustrates a potential explanation for our findings of zero net effects on cognitive skills and negative net effects on non-cognitive skills for youth in high school. The top curve shows how cognitive skills vary causally with enrichment. The lower curve shows the analogous relationship for non-cognitive skills. I_c is the level of enrichment that maximizes cognitive skills, and I_{nc} is the level of enrichment that maximizes non-cognitive skills. Around I_c, the net effect of enrichment on cognitive skills is close to zero and its corresponding effect on non-cognitive skills is negative.

positive net effect. However, homework may have only a small positive direct effect on non-cognitive skills while having a very negative indirect effect through the foregone activities, for a net negative effect. Therefore to maximize cognitive skills one may want to spend one more hour on homework, while to maximize non-cognitive skill one may not.

The trade-off between maximizing cognitive and non-cognitive skills depends entirely on which activities we are considering. While the trade-off in the case of homework may be high, the trade-off in the case of other activities with a higher social component, for example, may not be as pronounced. This may explain why the effect of enrichment on non-cognitive skills is more negative in high school. If the composition of enrichment changes from activities with low trade-offs (social extra-curriculars) to activities with high trade-offs (homework), we would expect the effect of enrichment on non-cognitive skills to become more negative in higher grades.

Indeed, this seems to be the case. Figure 7 shows that the composition of enrichment time in higher grades is more focused on activities that tend to provide a high direct effect on cognitive skills while having little or no effect on non-cognitive skills. In particular, the average share of enrichment time devoted to homework increases notably, from 52% in the PreK-5 group to 79% in grades 9-12.[8] At the same time, the average share devoted to reading books falls from 16% to 10%, and the time spent in before- and after-school programs (which often involves socializing with other children) declines precipitously from 18% to 0%.

Figure 7: Enrichment Time Breakdowns by Grade Level



Note: Panels plot the average division of time into different categories over a typical week for each grade level. The figure pools the 1997, 2002 and 2007 CDS waves.

How can we explain the breakdown of the high school results by household income presented in Section 4.3? There we see that low-income youth in high-school have a

[8]Note that this comparison is likely to understate the true disparity, since the nature of the homework across grades is different as well, with high school homework likely being less associated with non-cognitive skills (and more associated with cognitive skills) than homework in earlier grades.

slightly negative effect for non-cognitive skills, while the middle- and high-income youth have very negative effects. One possible explanation for this pattern could be that the composition of enrichment time in high school might differ by household income, as it does by age. However, we do not find evidence to support this hypothesis – Figure 12 in Appendix A shows that the composition of enrichment time in high school is on average very similar across the different income terciles. All three groups spend about the same share on each enrichment activity, including homework.

Figure 8 may help explain the difference in the non-cognitive estimates between high- and low-income youth. It shows that high-income children spend considerably more time on enrichment than middle- and low-income children. Thus, since the composition of enrichment is the same, we might expect high-income children to have relatively more negative non-cognitive estimates simply due to diminishing returns.[9]

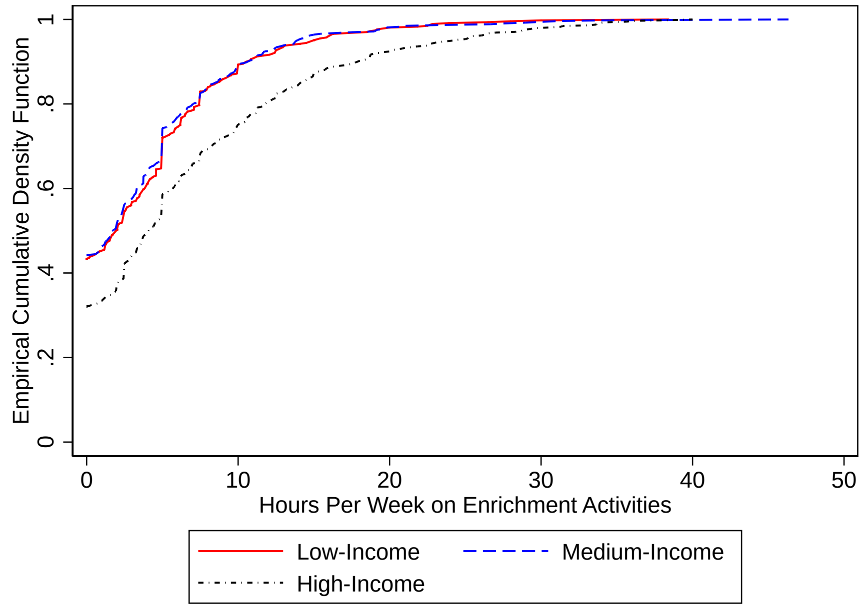
We still need to understand the difference in the non-cognitive estimates between medium- and low-income youth, since both the amount of time spent on enrichment, as well as the composition of enrichment, is the same across those terciles. In the next section, we argue that substitution patterns may explain the difference – at the margin, low- and middle-income children spend time on enrichment at the expense of different activities.

5.2 Which activities are crowded out by enrichment?

In this section, we attempt to gain some insight into the substitution patterns between enrichment and other activities. When high-schoolers do an additional hour of enrichment, from which activities is that hour taken? In particular, we ask if these activities are different depending on household income. A detailed analysis aimed at obtaining the exact extent of substitution between enrichment and each alternative activity would

[9]Note that cognitive estimates are all near zero across income terciles, which is consistent with the idea that all income groups are choosing levels of enrichment near their corresponding I_c. The fact that I_c for high-income children is larger than I_c for their lower-income counterparts may be due to differences in the production function. Indeed, it is plausible that high-income children have access to additional resources that might be complementary to enrichment activities (e.g., smaller class sizes, better teachers, etc).

Figure 8: Enrichment Time by Income Tercile: Grades 9-12



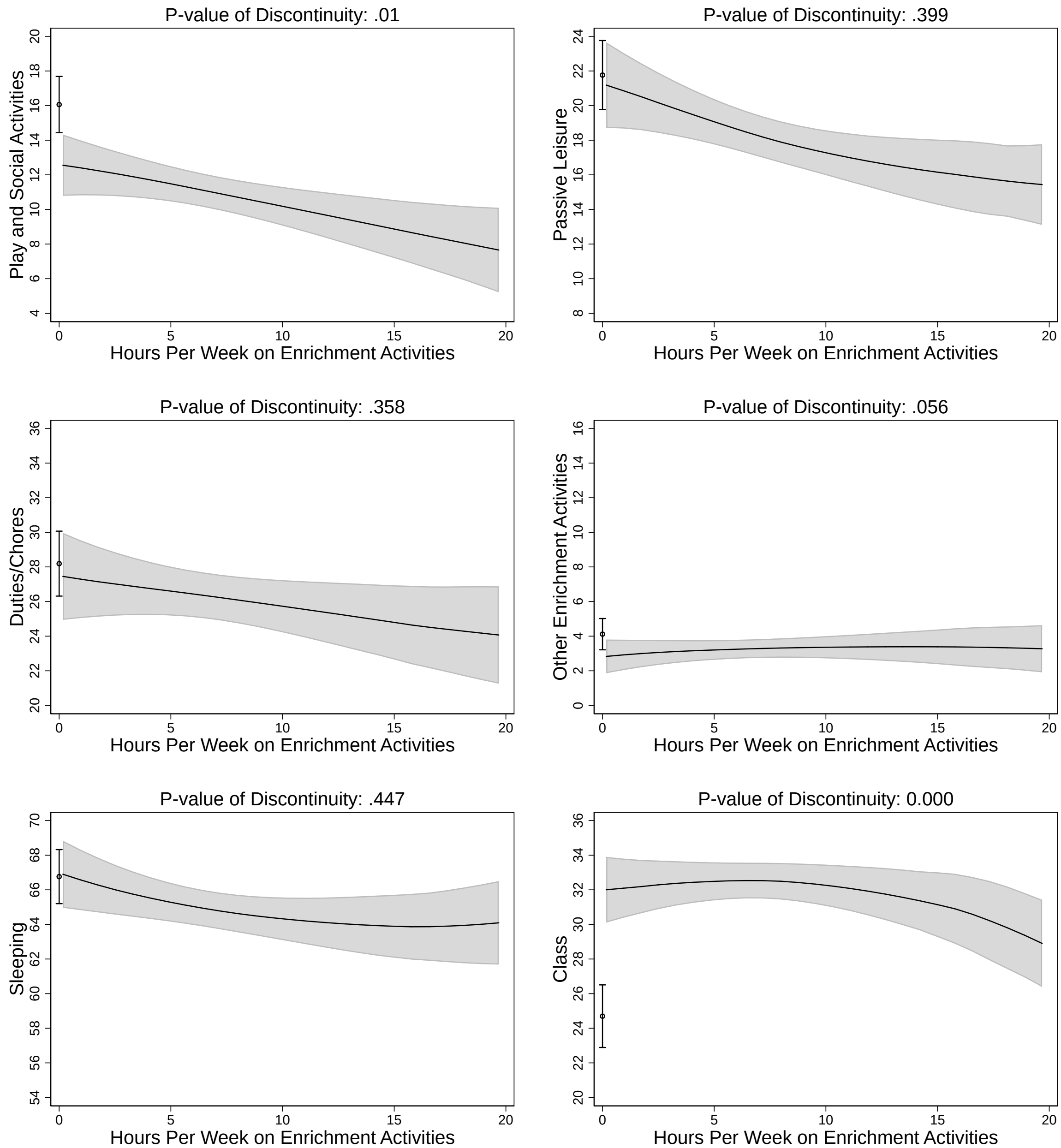
Note: This Figure shows the empirical cumulative distribution functions of enrichment time for each household income tercile, among high school children.

require the identification of causal substitution effects and is beyond the scope of this paper. Nevertheless, here we use some of the ideas explored so far in this paper to provide suggestive evidence that children in high school substitute enrichment away from different activities depending on their income. We find that middle-income children forego activities that would have more benefits to non-cognitive skills than low- and high-income children.

We begin with the low-income tercile in Figure 9. Each panel plots the average time spent on the other activity categories described in Section 2 for each level of enrichment. The figure shows that as low-income children spend more time on enrichment, they tend to spend less time on play and social activities, passive leisure, duties/chores, sleep, and class time. The relationship between enrichment and other enrichment activities (activities included in the broad enrichment category but not included in the baseline enrichment category) is roughly flat.

Of course, these relationships are not necessarily causal, so they may not reflect actual

Figure 9: Child Time Use by Enrichment Time: Low-Income Children in Grades 9-12



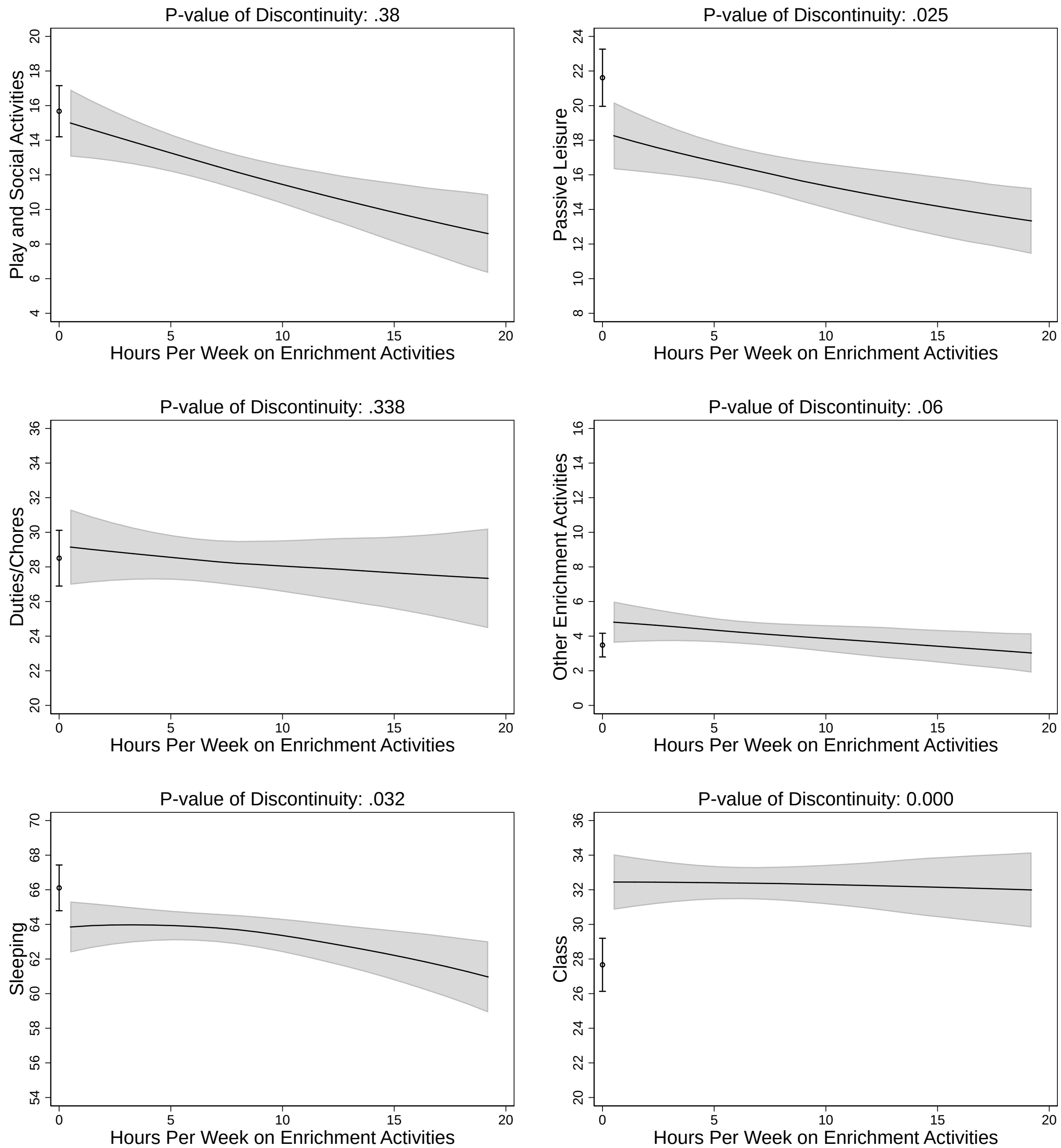
Note: Each panel shows a plot of the local linear estimator of the expected value of a variable for a given amount of time spent on enrichment, along with its 90% confidence interval. The expected value of the variable for the children who spent no time on enrichment is also shown, along with its 90% confidence interval. Finally, the *p*-value of a test for whether there is discontinuity at zero time on enrichment is also shown in the header of each panel.

substitution. However, we can use Caetano (2015)'s test of exogeneity (see Section 3 and Appendix B) to gain insight into whether endogeneity is likely to play a major role in these observed raw correlations. For instance, the average amount of play and social activities for the zero-enrichment children is starkly different from the average for those children who do just one or two hours of enrichment per week. This suggests that the raw correlation between play and social activities and enrichment is not necessarily causal – the discontinuity is evidence of uncontrolled-for endogeneity. By contrast, there is no evidence of a discontinuity at zero (p=0.399) for passive leisure, suggesting that the very negative gradient between passive leisure and enrichment may be causal, and thus imply some substitution.

Extending this logic to the other panels, we conclude that there is some evidence that high school children from low-income households substitute toward enrichment away from passive leisure, duties/chores, and sleep. The gradient between enrichment and passive leisure is much steeper than the analogous gradients for sleep and duties/chores, suggesting that the evidence of substitution away from passive leisure is the strongest. Analogously, Figure 10 shows some evidence that high school children from middle-income households substitute toward enrichment away from play/social activities and duties/chores. The strongest evidence points to play/social activities, which has a much steeper slope.

The apparent differences in substitution patterns by household income may help explain why the non-cognitive effect for the middle-income group is particularly negative relative to the low-income group despite their similar levels and compositions of enrichment. Middle-income high school students tend to substitute toward enrichment away from play/social activities, while low-income high school students tend to substitute toward enrichment away from passive leisure (screen time, consisting mostly of watching TV). Clearly, the opportunity cost of enrichment is higher for middle-income students, as play/social activities are known to be beneficial to non-cognitive skills relative to watching TV ((Lukianoff and Haidt, 2018)).

Figure 10: Child Time Use by Enrichment Time: Middle-Income Children in Grades 9-12



Note: Each panel shows a plot of the local linear polynomial estimator of the expected value of a variable for a given amount of time spent on enrichment, along with its 90% confidence interval. The expected value of the variable for the children who spent no time on enrichment is also shown, along with its 90% confidence interval. Finally, the p-value of a test for whether there is discontinuity at zero time on enrichment is also shown in the header of each panel.

Figure 13 in Appendix A presents the analogous results for high-income children, suggesting that they substitute toward enrichment away from passive leisure, play/social activities, duties/chores and other enrichment, with the strongest evidence for passive leisure due to its much steeper slope. It seems that differences in enrichment totals explain most of the differences in the non-cognitive returns to enrichment between high- and low-income youth, as the substitution patterns seem similar.

In sum, differences in substitution seem to explain most of the difference in the non-cognitive effects between low- and medium-income children, while the differences in total time spent on enrichment seem to explain most of the difference in the non-cognitive effects between low- and high-income children.

6 Conclusion

In this paper, we estimate the total effect of time spent on enrichment activities on cognitive and non-cognitive skills. We propose an endogeneity correction which leverages the bunching at zero enrichment generated by the constraint that the choice of enrichment cannot be negative.

Our results suggest that the sizable, positive correlations observed between enrichment time and childhood skills are mostly driven by unobservables. Correcting for the bias introduced by these unobservables, we find that the net causal effect of enrichment activities is negligible and may even be negative for cognitive skills. Regarding non-cognitive skills, the corrected estimates are also negligible in earlier grades, but quite negative and very significant in high school. The negative high school effects for non-cognitive skills are particularly large for middle- and high-income children.

We interpret our results through the lens of a model of time allocation and skill production. We argue that if parents and children put a lot of weight on cognitive skill production when choosing their level of enrichment, we would expect marginal cognitive returns to enrichment to gravitate towards zero. However, parents and children cannot maximize cognitive and non-cognitive skills at the same time. If there is a practical trade-off between maximizing cognitive skills and non-cognitive skills when choosing the level of enrichment, then the non-cognitive returns might be negative.

This model may also explain why the non-cognitive effects are quite negative in high school. Intensifying competition for college admissions means that high school may be a time when enrichment is especially geared towards cognitive skills and away from non-cognitive skills. We show that this is indeed the case, as enrichment shifts towards more homework and less social activities as children get older.

The more negative effects for high-income in comparison to low-income youth in high school may be explained by the fact that high-income youth spend substantially more time on enrichment. The more negative effects for middle-income in comparison to low-income youth in high school may be explained by the substitution patterns: middle-income youth tend to choose their last hour of enrichment at the expense of play and social activities. In contrast, low-income youth tend to choose their last hour of enrichment at the expense of TV. Social activities are likely more beneficial for non-cognitive skills than TV, so the opportunity cost of enrichment for middle-income children is higher.

Finally, we call attention to the need for the development of further, larger data sources connecting time use and skills. Currently, the question posed in this paper can only be studied with two datasets, both with limited sample sizes (CDS-PSID and LSAC). Larger datasets would not only improve the precision of the estimates but would also allow a more complete study of causal substitution patterns.

References

Abeles, V. (2015). *Beyond Measure, Rescuing an Overscheduled, Overtested, Underestimated Generation.* Simon and Schuster, New York, NY.

Aguiar, M. and Hurst, E. (2007). Measuring Trends in Leisure: The Allocation of Time Over Five Decades. *Quarterly Journal of Economics*, 122:969–1006.

Anderegg, D. (2003). *Worried All the Time: Rediscovering the Joy in Parenthood in an Age of Anxiety*. Free Press, New York, NY.

Avent, R. (2017). High-Pressure Parenting. *The Economist*, Feb/Mar 2017.

Bernal, R. and Keane, M. P. (2011). Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers. *Journal of Labor Economics*, 29(3):459–512.

Bianchi, S. M. (2000). Maternal Employment and Time with Children: Dramatic Change or Surprising Continuity? *Demography*, 37:401–414.

Bonhomme, S., Lamadon, T., and Manresa, E. (2017). Discretizing Unobserved Heterogeneity. Working Paper.

Bonhomme, S. and Manresa, E. (2015). Grouped Patterns of Heterogeneity in Panel Data. *Econometrica*, 83(3):1147–1184.

Caetano, C. (2015). A Test of Exogeneity Without Instrumental Variables in Models With Bunching. *Econometrica*, 83(4):1581–1600.

Caetano, C., Caetano, G., and Nielsen, E. (2020). Correcting Endogeneity Bias in Models with Bunching. *Working Paper*. Available here.

Caetano, G., Kinsler, J., and Teng, H. (2019). Towards causal estimates of children's time allocation on skill development. *Journal of Applied Econometrics*, 34(4):588–605.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9):2593–32.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics*, 132(4):1593–1640.

Doepke, M. and Zilibotti, F. (2017). Parenting With Style: Altruism and Paternalism in Intergenerational Preference Transmission. *Econometrica*, 85:1331–1371.

Doepke, M. and Zilibotti, F. (2019). *Love, Money, and Parenting: How Economics Explains the Way We Raise Our Kids*. Princeton University Press.

Duncan, G. and Murnane, R. (2011). *Restoring Opportunity: The Crisis of Inequality and the Challenge for American Education*. Russell Sage, New York, NY.

Fiorini, M. and Keane, M. P. (2014). How the Allocation of Children's Time Affects Cognitive and Noncognitive Development. *Journal of Labor Economics*, 32(4):787–836.

Ginsburg, K. R. et al. (2007). The Importance of Play in Promoting Healthy Child Development and Maintaining Strong Parent-Child Bonds. *Pediatrics*, 119(1):182–191.

Gray, P. (2010). The Decline of Play and Rise in Children's Mental Disorders. *Psychology Today*, Jan 2010.

Gray, P. (2011). The Decline of Play and the Rise of Psychopathology in Children and Adolescents. *American Journal of Play*, 3(4):443–463.

Gray, P. (2013). *Free to Learn: Why Unleashing the Instinct to Play Will Make Our Children Happier, More Self Reliant and Better Students for Life.* Basic Books, New York, NY.

Gray, P. (2019). Evolutionary Functions of Play: Practice, Resilience, Innovation, and Cooperation. In Smith, P. K., editor, *The Cambridge Handbook of Play: Developmental and Disciplinary Perspectives*, pages 84–102. Cambridge University Press, Cambridge.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153–161.

Heckman, J. J. and Rubinstein, Y. (2001). The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *American Economic Review*, 91(2):145–149.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3):411–482.

Jackson, C. K. (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes. *Journal of Political Economy*, 126(5):2072–2107.

Jarvis, P., Newman, S., and Swiniarski, L. (2014). On 'Becoming Social': The Importance of Collaborative Free Play in Childhood. *International Journal of Play*, 3(1):53–68.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6):615 – 631.

Kane, T. J. and Staiger, D. O. (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. Working Paper 14607, National Bureau of Economic Research.

Khazan, O. (2016). Ending Extracurricular Privilege. *The Atlantic*, Dec 2016.

Lareau, A. (2003). *Unequal Childhoods: Class, Race, and Family Life*. University of California Press, Berkeley, CA.

Lin, C.-C. and Ng, S. (2012). Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown. *Journal of Econometric Methods*, 1(1):42–55.

Lindqvist, E. and Vestman, R. (2011). The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment. *American Economic Journal: Applied Economics*, 3(1):101–28.

Lukianoff, G. and Haidt, J. (2018). *The Coddling of the American Mind: How Good Intentions and Bad Ideas are Setting up a Generation for Failure*. Penguin.

Luthar, S. S. (2003). The Culture of Affluence: Psychological Costs of Material Wealth. *Child Development*, 74(6):1581–1593.

Luthar, S. S. and Becker, B. E. (2002). Privileged But Pressured? A Study of Affluent Youth. *Child Development*, 73(5):1593–1610.

Neal, D. A. and Johnson, W. R. (1996). The Role of Premarket Factors in Black-White Wage Differences. *The Journal of Political Economy*, 104:869–895.

Peterson, J. and Zill, N. (1986). Marital Disruption, Parent-Child Relationships, and Behavioral Problems in Children. *Journal of Marriage and the Family*, 48:295–307.

Ramey, G. and Ramey, V. (2010). The Rug Rat Race. Working Paper.

Rosen, R. (2015). Why Affluent Parents Put So Much Pressure on Their Kids. *The Atlantic*, Dec 2015.

Rosenfeld, A. A. and Wise, N. (2000). *The Over-Scheduled Child: Avoiding the Hyper-parenting Trap*. St Martin's Griffin, New York, NY.

Rosin, H. (2015). The Silicon Valley Suicides. *The Atlantic*.

Todd, P. and Wolpin, K. (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, 1(1):91–136.

Veiga, G., Neto, C., and Rieffe, C. (2016). Preschoolers' Free Play: Connections with Emotional and Social Functioning. *International Journal of Emotional Education*, 8(1):48–62.

Villaire, T. (2003). Families on the Go: Active or Hyperactive? *Our Children*, 28(5):4–5.

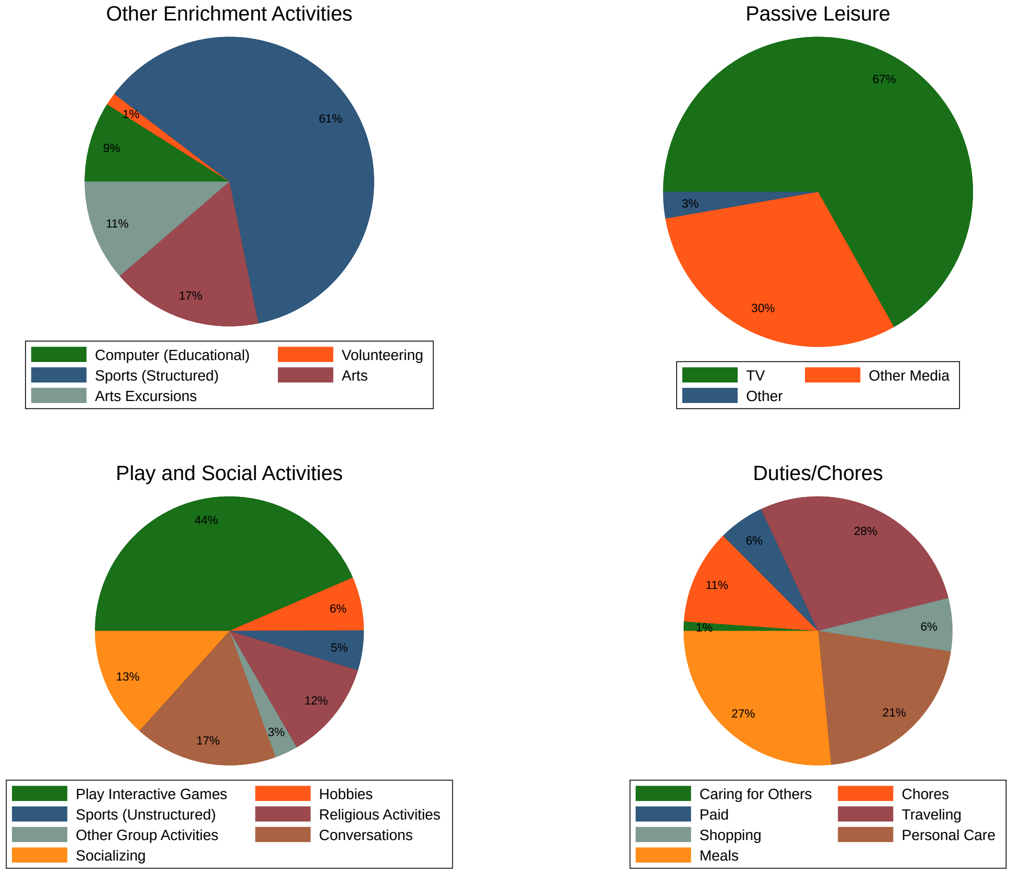
Waddell, G. R. (2006). Labor-Market Consequences of Poor Attitude and Low Self-Esteem in Youth. *Economic Inquiry*, 44(1):69–97.

Walker, M. (2017). *Why We Sleep: Unlocking the Power of Sleep and Dreams*. Simon and Schuster.

Warner, J. (2005). *Perfect Madness: Motherhood in the Age of Anxiety*. Riverhead Books, New York, NY.

A Supporting Tables and Figures

Figure 11: Time Breakdowns - Other Time Aggregates



Note: Panels plot the average division of time into different categories over a typical week for our full CDS sample. The figure pools the 1997, 2002 and 2007 CDS waves.

Table 7: Cognitive and Non-Cognitive Factor Loadings

Cognitive Skills	1997	2002	2007
Letter Word	0.95	0.94	0.85
Applied Problems	0.89	0.89	0.76
Passage Comprehension	0.96	0.96	0.90
Non-Cognitive Skills			
Cheat or tells lies	0.46	0.52	0.56
Bullies or mean to others	0.55	0.56	0.51
Feels no regret after misbehaving	0.41	0.45	0.43
Breaks things on purpose	0.46	0.48	0.47
Has sudden changes in mood	0.55	0.56	0.58
Feels no love	0.49	0.52	0.57
Too fearful or anxious	0.41	0.47	0.50
Feels worthless or inferior	0.48	0.53	0.64
Sad or depressed	0.52	0.55	0.64
Cries too much	0.42	0.36	0.38
Easily confused	0.50	0.53	0.53
Has obsessions	0.51	0.51	0.60
Rather high strung, tense and nervous	0.48	0.54	0.53
Argues too much	0.60	0.59	0.59
Disobedient	0.51	0.58	0.57
Stubborn, sullen, or irritable	0.61	0.61	0.64
Has a very strong temper	0.59	0.65	0.64
Has difficulty concentrating	0.57	0.59	0.59
Impulsive, or acts without thinking	0.62	0.62	0.62
Restless or overly active	0.55	0.52	0.49
Has trouble getting allong with other children	0.59	0.59	0.59
Not liked by other children	0.44	0.43	0.50
Withdrawn, does not get involved with others	0.37	0.43	0.45
Clings to adults	0.32	0.31	0.27
Demands a lot of attention	0.58	0.53	0.54
Too dependent on others	0.43	0.46	0.49
Thinks before acting, not impulsive	0.52	0.52	0.58
Generally well behaved, does what adults request	0.53	0.59	0.60
Can get over being upset quickly	0.42	0.44	0.51
Waits turns in games and other activities	0.47	0.52	0.49
Gets along well with other children	0.60	0.62	0.61
Admired by other children	0.55	0.55	0.57
Cheerful, happy	0.42	0.48	0.58
Tries things for himself/herself	0.35	0.34	0.46
Does neat, careful work	0.39	0.41	0.49
Curious and exploring, likes new experiences	0.12	0.21	0.26

Note: Cognitive and non-cognitive factor loadings are shown for each CDS wave.

Table 8: Uncorrected and Corrected Results – Alternative Skill Measures

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Cognitive						
Applied	β	0.013**	0.008**	0.000	-0.004	0.000
Problems		(0.003)	(0.002)	(0.006)	(0.005)	(0.006)
	δ			0.007	0.010**	0.007
				(0.005)	(0.004)	(0.005)
Letter Word	β	0.012**	0.008**	-0.001	-0.003	-0.001
		(0.003)	(0.001)	(0.006)	(0.005)	(0.006)
	δ			0.007	0.009**	0.007
				(0.005)	(0.004)	(0.005)
Passage	β	0.014**	0.009**	-0.001	-0.003	0.000
Comprehension		(0.003)	(0.001)	(0.006)	(0.005)	(0.006)
	δ			0.009*	0.010**	0.008*
				(0.005)	(0.004)	(0.005)
Non-Cognitive						
External	β	0.010**	0.006**	-0.014	-0.023**	-0.017*
		(0.002)	(0.002)	(0.009)	(0.008)	(0.009)
	δ			0.016**	0.023**	0.018**
				(0.008)	(0.007)	(0.008)
Internal	β	0.002	-0.001	-0.021**	-0.026**	-0.017*
		(0.003)	(0.003)	(0.010)	(0.009)	(0.010)
	δ			0.016**	0.020**	0.013
				(0.008)	(0.007)	(0.008)

Note: N=4,330. Bootstrapped standard errors in parentheses (500 iterations). The corrected specifications use 50 clusters. ** p<0.05, * p<0.1.

Table 9: Uncorrected and Corrected Results – Broad Enrichment

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Cognitive	β	0.024**	0.010**	-0.001	-0.003	-0.002
		(0.002)	(0.001)	(0.007)	(0.006)	(0.006)
	δ			0.011*	0.012**	0.011**
				(0.006)	(0.006)	(0.005)
Non-Cognitive	β	0.009**	0.007**	-0.018*	-0.018**	-0.017**
		(0.002)	(0.002)	(0.011)	(0.009)	(0.008)
	δ			0.023**	0.023**	0.021**
				(0.009)	(0.008)	(0.007)

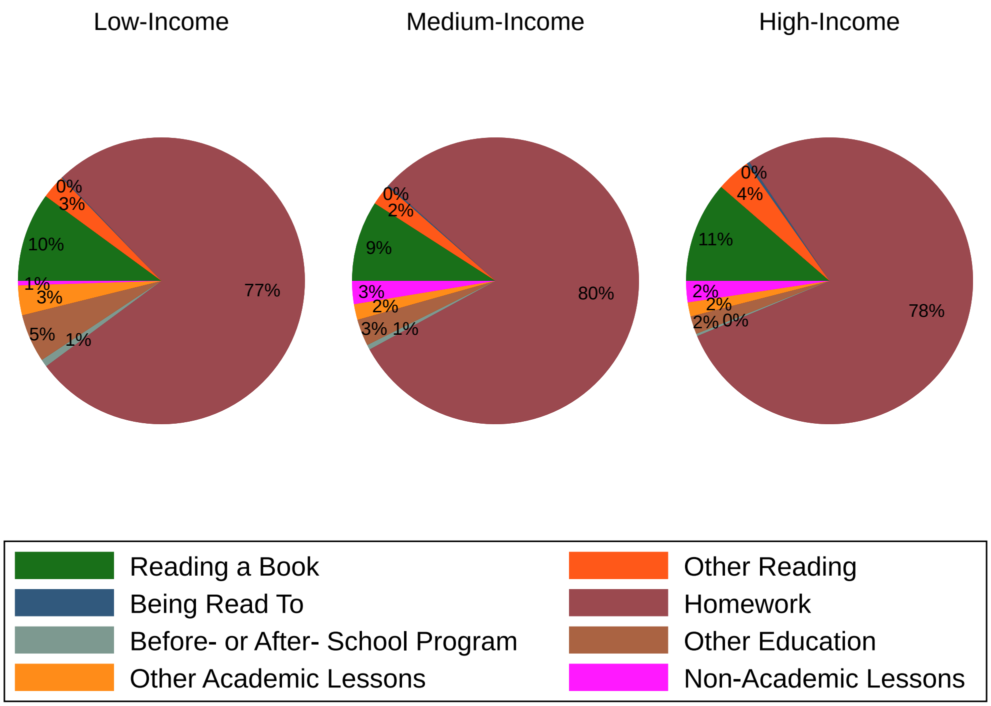
Note: N=4,330. Bootstrapped standard errors in parentheses (500 iterations). The corrected specifications use 50 clusters. ** p<0.05, * p<0.1.

Table 10: Uncorrected and Corrected Results – Homework Only

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Cognitive	β	0.032**	0.010**	0.007	0.002	0.003
		(0.003)	(0.002)	(0.007)	(0.007)	(0.008)
	δ			0.003	0.006	0.006
				(0.005)	(0.005)	(0.006)
Non-Cognitive	β	0.011**	0.006**	-0.020*	-0.032**	-0.029**
		(0.003)	(0.003)	(0.011)	(0.010)	(0.013)
	δ			0.020**	0.028**	0.028**
				(0.008)	(0.007)	(0.010)

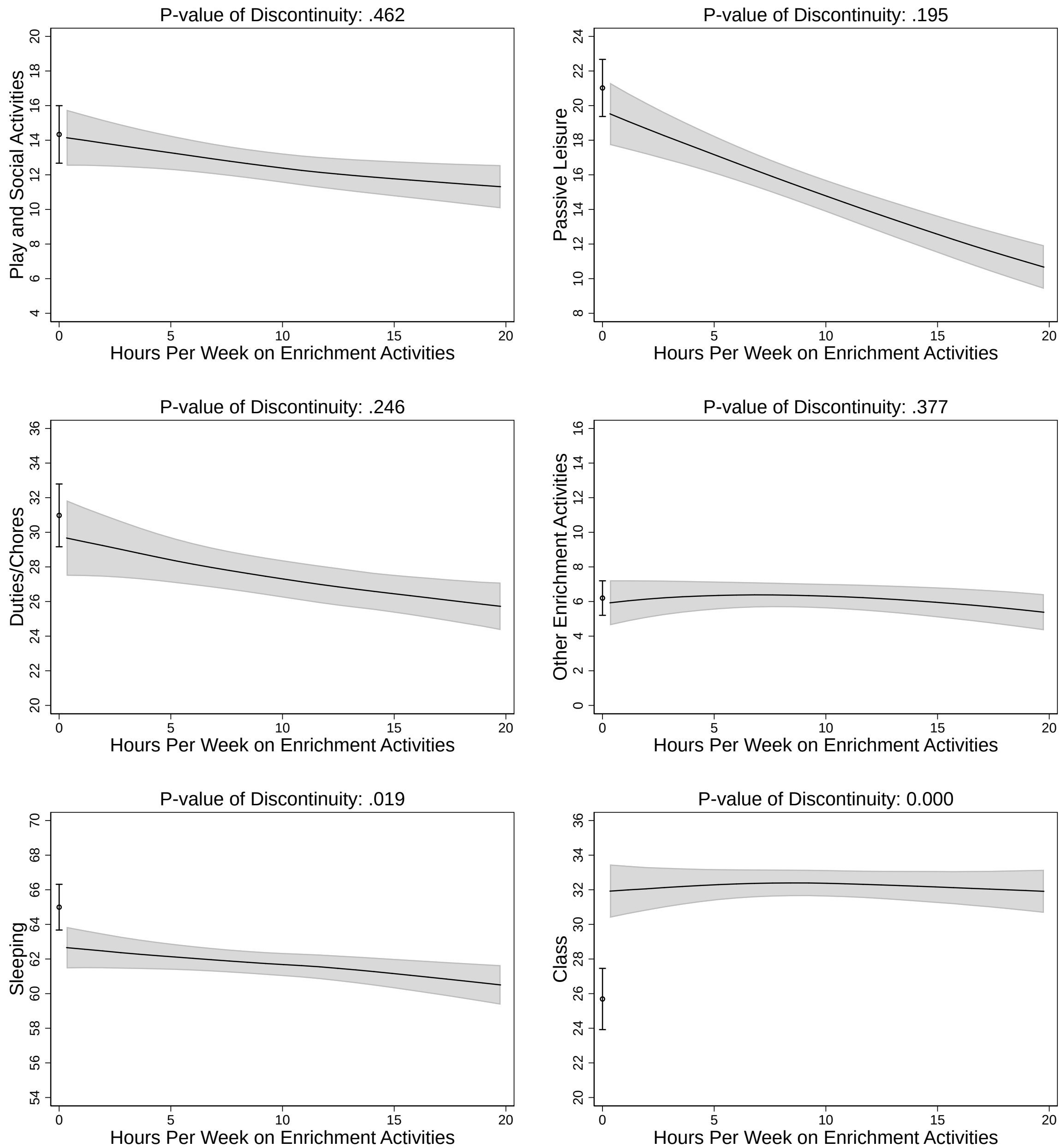
Note: N=4,330. Bootstrapped standard errors in parentheses (500 iterations). The corrected specifications use 50 clusters. ** p<0.05, * p<0.1.

Figure 12: Enrichment Time Breakdowns - High School, By Income Tercile



Note: Panels plot the average division of time into different categories over a typical week for each income tercile among those in grades 9-12. The figure pools the 1997, 2002 and 2007 CDS waves.

Figure 13: Child Activities by Enrichment Time: High-Income Children in Grades 9-12



Note: Each panel shows a plot of the local linear estimator of the expected value of a variable for a given amount of time spent on enrichment, along with its 90% confidence interval. The expected value of the variable for the children who spent no time on enrichment is also shown, along with its 90% confidence interval. Finally, the *p*-value of a test for whether there is discontinuity at zero time on enrichment is also shown in the header of each panel.

B The uncorrected estimates are positively biased.

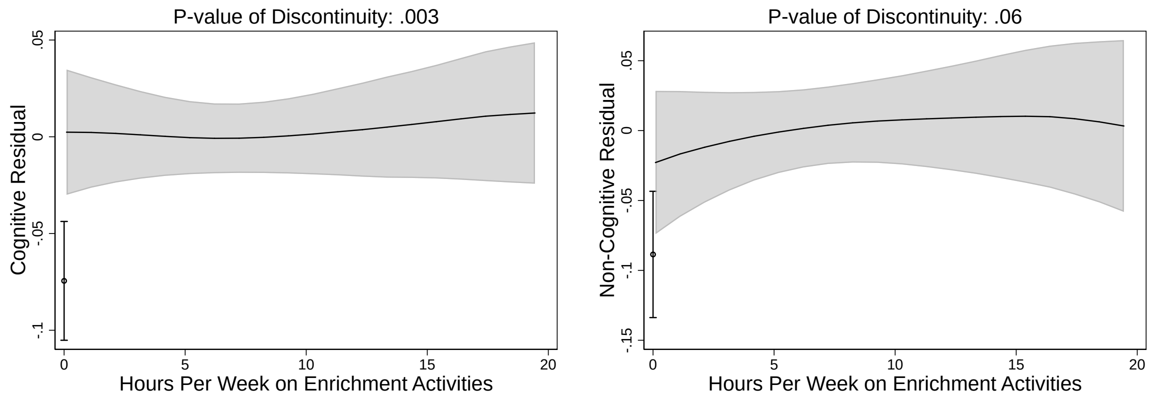
In Section 3, we claim that I is endogenous in the uncorrected model given by equation (1). In fact, we argue further that the bias resulting from this endogeneity is positive. We provide additional evidence in support of these claims here.

We do this using Caetano (2015)'s test of exogeneity. The test exploits the fact that if I is exogenous, then $\mathbb{E}[\epsilon|I,X] = 0$, and therefore $\mathbb{E}[S|I,X]$ must be continuous in I at zero. The idea is thus to estimate $\mathbb{E}[S|I,X]$ using only observations for which $I > 0$. If I is exogenous, then the limit of $\hat{\mathbb{E}}[S|I,X]$ as I approaches zero should be equal to $\hat{\mathbb{E}}[S|I = 0,X]$. We perform this test using the full list of controls included in our main analysis in Section 4. We find strong evidence that $\mathbb{E}[S|I = 0,X]$ is discontinuous at $I = 0$. Thus, we conclude that I is endogenous and the uncorrected estimator of β must be biased.

In order to display this multivariate result in a two-dimensional plot, Figure 14 shows the residuals of regression (1) applied to cognitive skills (left panel) and non-cognitive skills (right panel) when we use only observations such that $I > 0$ to estimate the coefficients. The solid lines represent local linear fits of the residuals of these regressions conditional on enrichment time. The plots also show the average residuals at zero enrichment along with their 90% confidence intervals.

Because we are conditioning on all controls X when we run regression (1), the local linear fits already incorporate all discontinuities in the controls. Therefore, the fact that the residuals at zero enrichment are discontinuously lower than the residuals just above zero for both cognitive and non-cognitive skills is direct evidence that the unobserved confounders are also discontinuous at zero enrichment. Moreover, because the discontinuity in the residuals is positive, we conclude that the unobservables that contribute positively to enrichment also directly contribute positively to skills, and thus the OLS estimator of β in equation (1) is biased upward for both types of skill.

Figure 14: Evidence that Standard Estimates May be Biased Upward



Note: Each panel shows a plot of the local linear polynomial estimator of the expected value of the residuals from equation (1), estimated on the positive enrichment subsample, conditional on enrichment time, along with its 90% confidence interval. The expected value of the residuals for the children who spent no time on enrichment is also shown, along with its 90% confidence interval. Finally, the *p*-value of a test for whether there is discontinuity at zero is also shown in the header of each panel.

C Are our findings an artifact of assuming linearity in η?

The linear specification of the error term in equation (4) can be relaxed in a more flexible, non-parametric model, as discussed in Caetano et al. (2020). Ultimately, however, we can never escape the fact that in order to use the discontinuities in skill at zero enrichment to make inferences about endogeneity globally, we must be willing to accept that the effect of the confounders on skills at zero enrichment is informative about the corresponding effect in the rest of our sample.

Effectively, our method estimates the average treatment effect corrected for the endogeneity from the variables that are correlated with enrichment around zero. Thus, it is important to acknowledge that the effect of confounders on skills at zero may not be representative of the effect in the whole sample, which may lead us to either over-correct or under-correct for bias. In this section we show evidence that, if anything, we might be under-correcting for the positive bias in our application. Our main empirical findings therefore do not seem to be an artifact of this linearity assumption.

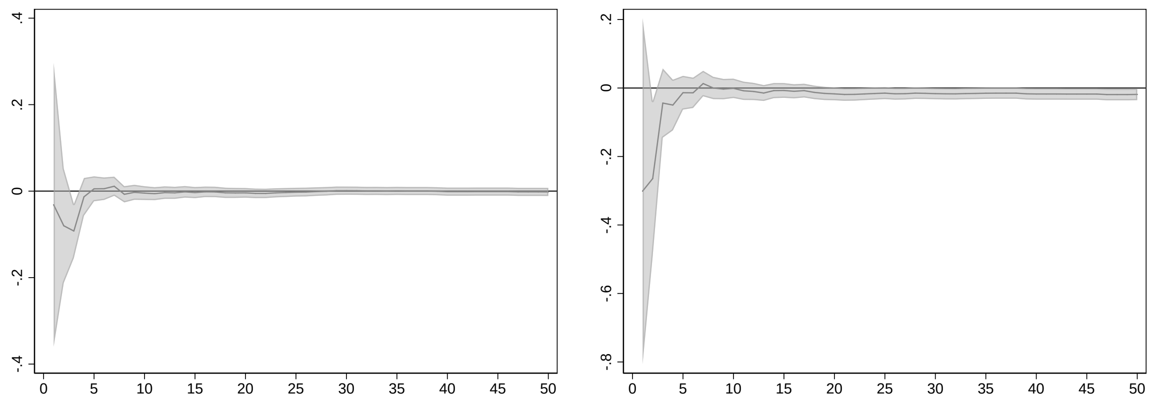
To fix ideas, we consider two extreme scenarios: one where people only spend either 0 or 1 hours per week on enrichment, and one where people spend between 0 and 50 hours of enrichment per week. The linearity assumption is more plausible in the first scenario than in the second scenario. Intuitively, the effect of the confounders at $I = 0$ is more plausibly similar to the effect of the confounders at $I = 1$ than at $I = 50$.

We build on this idea by restricting the sample to reflect the first scenario, and then we progressively expand the sample until it reaches the second scenario. In Figure 15, we show how our main estimate $\hat{\beta}$ for cognitive (left panel) and non-cognitive skills (right panel) changes for different truncations of our sample depending on the maximum allowed enrichment value ($I \leq I_{\max}$, ranging from $I_{\max} = 1$ to $I_{\max} = 50$).[10] As the maximum hours per week spent on enrichment in our full sample is 50, the estimates in the far right of

[10]To keep everything else constant irrespective of $I_{\max}$, we maintain the same estimate of $\mathbb{E}[I^*|I^* \leq 0, X]$ using our preferred tail symmetry approach. Note that the identification of $\mathbb{E}[I^*|I^* \leq 0, X]$ does not depend on the assumption of linearity of the structural equation on η, which is what we are trying to test here.

each panel are the estimates reported in Table 2.

Figure 15: Estimates for Different Sub-samples of the Data - Full Sample



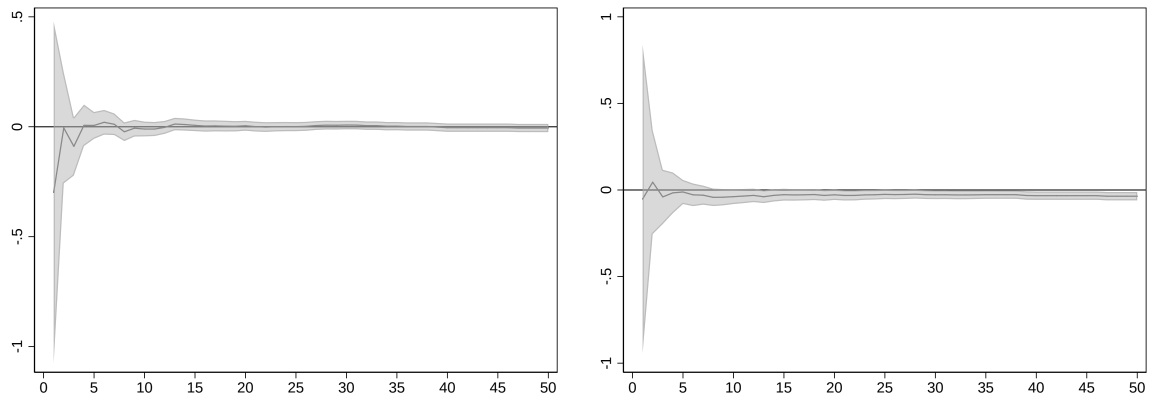
Note: Each panel shows the estimate of β for cognitive (left panel) or non-cognitive skills (right panel) restricting the sample to only children whose enrichment hours are lower than or equal to $I_{\max}$ for values of $I_{\max}$ ranging from 1 (only those who chose $I = 0$ or $I = 1$) to 50 (everyone). These plots suggest that our main findings are not an artifact of the linearity in η assumption.

We find that the estimates of β are mostly similar to the main estimates from Table 2, except for very small values of $I_{\max}$ where the estimates are more negative (albeit with substantially wider confidence intervals).

Note that this is in fact a joint test of the linearity of η as well as of whether the treatment effects vary with I (see footnote 4 in Section 3.2.) If the effect of I on S varies with I (be it because it is a function of I or because it is a function of X, which is itself correlated with I) we should also find variations in our estimates when we restrict the sample as in Figure 15 above. The fact that our estimates are constant as we increase $I_{\max}$ indicates that the treatment effects are likely to be uncorrelated with I. Additionally, this approach allows us to understand whether our conclusions are robust to the elimination of enrichment outliers from our sample, which indeed seems to be the case.

For completeness, Figure 16 shows the analogous plots for high school age children. The findings are similar. We conclude that our main findings in the paper - negative but insignificant cognitive estimates and negative significant non-cognitive estimates in high school - are likely not an artifact of our assumption of linearity in η.

Figure 16: Estimates for Different Sub-samples of the Data - Grades 9-12



Note: Each panel shows the analogous estimates to Figure 15 but for the sub-sample of children in high school.

D Identification of $\mathbb{E}[I^*|I^* \leq 0, X]$

This section discusses three strategies for the identification of $\mathbb{E}[I^*|I^* \leq 0, X]$, presented in increasing order of generality. For convenience we repeat here the assumptions of each method described in Section 3.3. The technical details of all these methods can be found in Caetano et al. (2020).

Tobit

Our first strategy is to identify $\mathbb{E}[I^*|I^* \leq 0, X]$ in a structure similar to Heckman (1979). We call this the "Tobit" strategy because it assumes that I^* satisfies a Tobit model. Specifically, in this strategy, we assume that we can write $g(X) = X'\gamma$ and that

$$\eta|X \sim \mathcal{N}(X'\theta, \sigma^2). \tag{9}$$

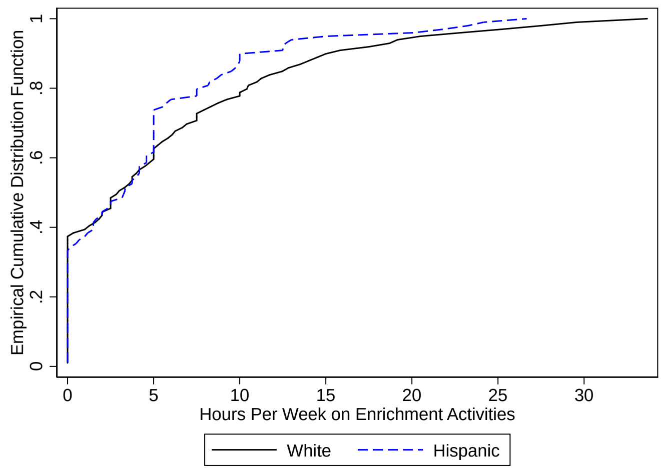
By assuming normality and homoskedasticity, the Tobit strategy constrains the shape of the entire distribution of the unobservable confounders η. Thus, the mean and the variance of this distribution can be identified simply by looking at any portion of the distribution of enrichment time above zero. When equation (9) holds,

$$\mathbb{E}[I^*|I^* \leq 0, X] = X'\pi - \sigma\lambda(-X'\pi/\sigma) \tag{10}$$

where $\pi = \gamma + \theta$ and $\lambda(\cdot)$ is the inverse Mill's ratio. The parameters π and σ can be estimated straightforwardly via a Tobit regression of I on X and plugged into equation (10) to build an estimator of the correction term.

In practice, the Tobit strategy turns out to be too restrictive in our context. To illustrate this, Figure 17 plots the empirical conditional cumulative distribution function of enrichment time I for white and Hispanic high school students. Because we observe the positive quantiles $F_{I^*|X}(q)$ for different values of q, we can infer the shape of part of the distribution of $\eta|X$ per equation (2). Indeed, the figure shows that the homoskedasticity assumption clearly does not hold, as the variance of I is different for different values of X.

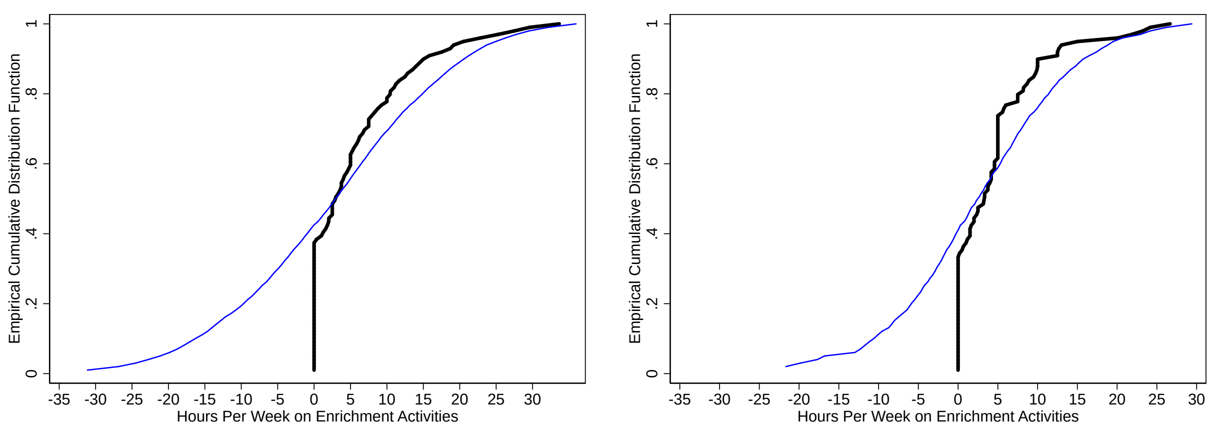
Figure 17: Evidence of heteroskedasticity on the distribution of $I|X$



Note: Each curve depicts the CDF of I for white and Hispanic high school students. The curves show evidence of heteroskedasticity in the distribution of desired enrichment for different values of the controls.

Figure 18 shows the homoskedastic Tobit fit for white (left panel) and Hispanic (right panel) high school students, along with their corresponding empirical CDF of I. It is evident that in both cases the fit on the positive side of I is not satisfactory, which would

Figure 18: Homoskedastic Tobit Fit



Note: Each panel depicts the CDF of enrichment (I) for white (left panel) and Hispanic (right panel) high school students presented in Figure 17 (thick curve) along with the corresponding homoskedastic Tobit fit (thin curve). The plots show evidence that the homoskedastic normal fit for positive values of enrichment is not satisfactory, which may lead us to over-estimate the magnitude of $\mathbb{E}[I^*|I^* \leq 0, X]$ and underestimate the magnitude of δ.

lead us to over-estimate the magnitude of $\mathbb{E}[I^*|I^* \leq 0, X]$ and thus under-estimate the magnitude of the bias. Indeed, this is what we find in Section 4.

Heteroskedastic Tobit

Next, we relax the linear mean and homoskedasticity requirements while maintaining the assumption that $\eta|X$ follows a normal distribution. Specifically, we suppose that

$$\eta|X \sim \mathcal{N}(l(X), \sigma^2(X)). \tag{11}$$

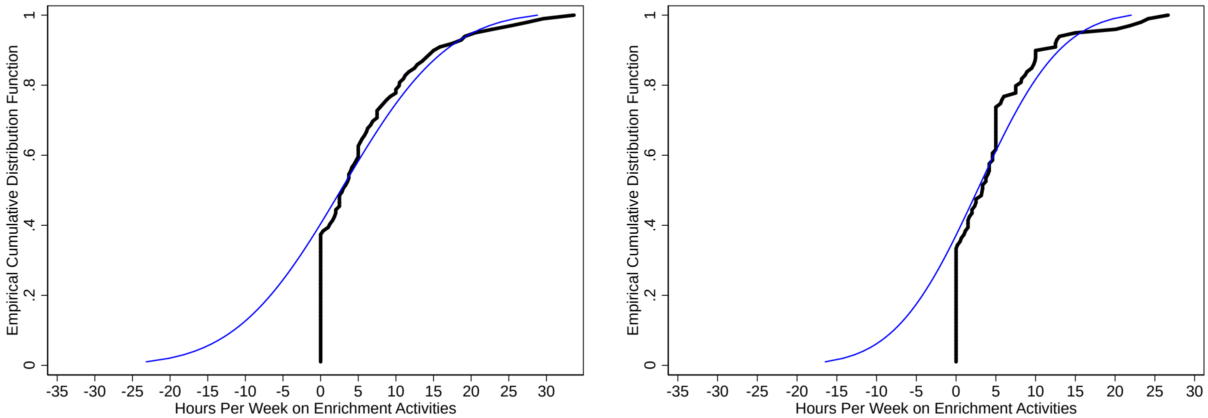
This assumption allows the mean and the variance of η to vary with X in an unrestricted way, but retains the requirement that η be normal separately for each value of X. In this case,

$$\mathbb{E}[I^*|I^* \leq 0, X] = g(X) + l(X) - \sigma(X)\lambda\left(-(g(X) + l(X))/\sigma(X)\right). \tag{12}$$

If X is discrete (or can be discretized, as is the case in our setting, see Appendix E) we can estimate $g(X) + l(X)$ and $\sigma(X)$ separately for each X by running a Tobit regression of I on a constant using only the observations with controls equal to X.

Figure 19 plots the empirical conditional cumulative distribution function I for the same two values of controls (white and Hispanic high school students) in Figure 17, along with the corresponding heteroskedastic Tobit fits for each. The fits are clearly superior to the homoskedastic fits presented in Figure 18. Nonetheless, it seems that the upper tail of the data is fatter than the upper tail implied by a normal distribution for both values of X. These plots are not an exception – we observe a similar pattern for many other values of X in the PSID data. If the upper tail is any indication of what is happening in the lower tail, this suggests that the heteroskedastic Tobit model will tend to under-estimate the magnitude of $\mathbb{E}[I^*|I^* \leq 0, X]$, thus over-estimating the magnitude of δ. Indeed, this is consistent with what we find in Section 4.

Figure 19: Heteroskedastic Tobit Fit



Note: Each panel depicts the CDF of enrichment (I) for white (left panel) and Hispanic (right panel) high school students presented in Figure 17 (thick curve) along with the corresponding heteroskedastic Tobit fit (thin curve). The plots show evidence of a better fit than the homoskedastic case for positive values of enrichment. The tails seem to be fatter in the empirical distribution in comparison with the fit, which may lead us to somewhat under-estimate the magnitude of $\mathbb{E}[I^*|I^* \leq 0, X]$ and over-estimate the magnitude of δ.

Heteroskedastic Tail Symmetry

Finally, we drop the normality assumption entirely and require only tail symmetry: for all censored quantiles q_0,

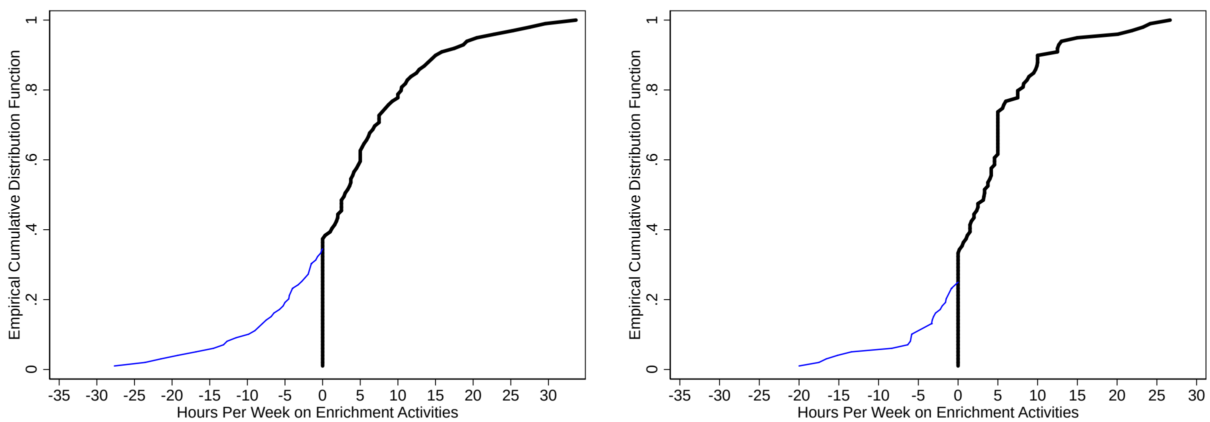
$$\eta|X \text{ has symmetric tails below } q_0 \text{ and above } 1 - q_0. \tag{13}$$

Tail symmetry requires only that the lower tail of $\eta|X$ below the censoring point and the corresponding upper tail be symmetric, a weaker assumption than symmetry of the entire distribution, which in turn is weaker than normality. Tail symmetry allows us to infer the behavior of $I^*|X$ when $I^* < 0$ by looking at the shape of the upper tail of $I|X$.[11]

To see this assumption in action, Figure 20 provides the corresponding plots shown in Figure 19 under tail symmetry. For quantiles below the bunching threshold, the fitted values follow the mirror image of the corresponding upper tail.

[11]This approach can only be implemented for values of X such that the proportion of children who are bunched at zero enrichment is less than half of the sample ($\mathbb{P}(I = 0|X) < 0.5$). In our sample, this is true for almost all values of X (over 97% of the observations). For values of X such that the proportion of children bunched at zero is over 50%, we can estimate $\mathbb{E}[I^*|I^* \leq 0, X]$ with the heteroskedastic Tobit approach, which is feasible with any amount of bunching.

Figure 20: Heteroskedastic Symmetric Fit



Note: Each panel depicts the CDF of I for white (left panel) and Hispanic (right panel) high school students presented in Figure 17 (thick curve) along with the corresponding heteroskedastic Symmetric fit (thin curve).

Under tail symmetry,

$$\mathbb{E}[I^*|I^* \leq 0, X] = F^{-1}_{I|X}(1 - F_{I|X}(0)) - \mathbb{E}[I|I \geq F^{-1}_{I|X}(1 - F_{I|X}(0)), X], \tag{14}$$

where $F_{I|X}(\cdot)$ is the cumulative distribution function of I conditional on X. We carry out the estimation of $\mathbb{E}[I^*|I^* \leq 0, X]$ using equation (14) in three steps. For each value of X, we first estimate the probability of bunching at zero enrichment, $F_{I|X}(0)$. Then we estimate the quantile of I in the upper tail that corresponds to the mirror image of $I = 0$, $F^{-1}_{I|X}(1-F_{I|X}(0))$. Finally we estimate the mean of $I|X$ at the upper tail, $\mathbb{E}[I|I \geq F^{-1}_{I|X}(1 - F_{I|X}(0)), X]$.

E Clustering and the discretization of X

In Section 3.3 and Appendix D, we discuss two strategies for the estimation of $\mathbb{E}[I^*|I^* \leq 0, X]$, heteroskedastic Tobit and heteroskedastic tail symmetry, which require the distribution of X to have a discrete support. In our setting, there are some important controls that have continuous support. Therefore, we want to be able to discretize X in a non-arbitrary way such that the discretized covariates naturally reflect the joint distribution of the original X. We discuss here how we discretize X using clustering methods. We show that our results are not an artifact of the specific way in which we implement the discretization nor of the number of clusters we use.

Our approach classifies observations with similar observed controls into discrete clusters and uses the corresponding cluster membership indicators as discretized versions of X. The classification attempts to maximize the similarities in X among observations in the same cluster – two children in the same cluster have by construction more similar controls than two children in different clusters. Formally, let k_i be the cluster to which child i belongs. The underlining assumption in our method is that $\mathbb{E}[I^*|I^* \leq 0, X = X_i] = \mathbb{E}[I^*|I^* \leq 0, X \in k_i]$ – we require that the heterogeneity in η conditional on X can be entirely explained by the clusters. The larger the number of clusters, the more similar are the controls of the children within the same cluster and thus the weaker this assumption becomes.

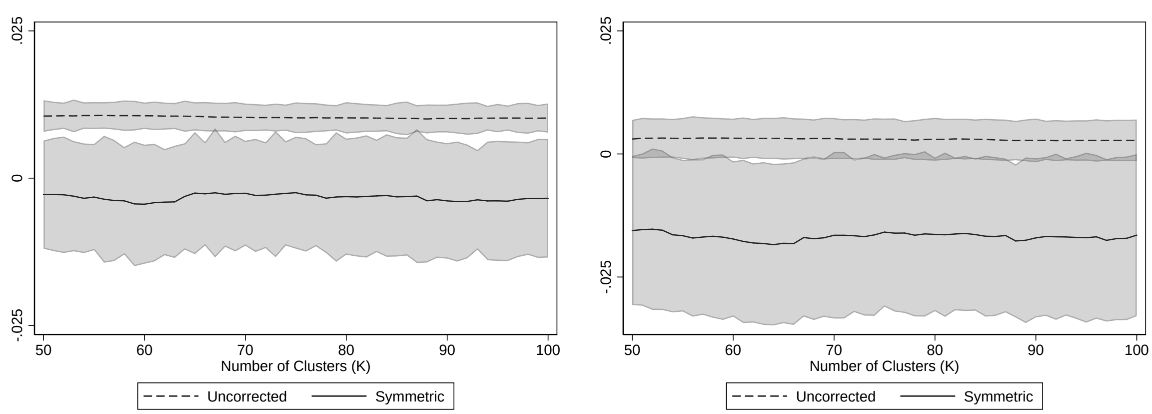
While there are many different clustering methods available, we report results based on hierarchical clustering because it produces clusters that are nested: if two children are in the same cluster when there are K clusters, then they will also be in the same cluster whenever there are $K' < K$ clusters. Moreover, the move from K to $K+1$ clusters always consists of splitting one (and only one) cluster into two smaller clusters. The nested nature of the hierarchical clusters provides some desirable discipline in the comparison of estimates for different numbers of clusters. If the estimate of β using K total clusters is different from the estimate using $K+1$ total clusters, the difference is due only to the cluster which was split. If the estimates of β remain close to constant as the number of clusters increases, this raises our confidence on the assumption that the clusters adequately capture differences in the conditional distributions $\eta|X$ (so that the assumption $\mathbb{E}[I^*|I^* \leq 0, X = X_i] = \mathbb{E}[I^*|I^* \leq 0, X \in k_i]$ is approximately valid).

We have experimented with different linkage methods (average, complete, and Ward's) and different dissimilarity measures (Gower, L_1, L_2, and correlation), and the results are uniformly very similar across these different cases. We report results with Ward's linkage method, in which the criterion at each step is to merge two clusters so as to achieve the minimum total within-cluster variance, and the Gower dissimilarity measure, as this measure works well when there is a combination of continuous and discrete variables.

All the results in the paper including the plots in this section use the clusters to estimate $\mathbb{E}[I^*|I^* \leq 0, X = X_i]$, but include both the original vector X and the cluster indicators in the main regression, equation (6). Specifically, we specify $m(X) = X'\tau + \sum_{k=1}^{K} \alpha_k \mathbf{1}(X \in k)$ in equation (5), in order to account for potential non-linearities (all estimates are robust to the exclusions of the cluster indicators).

Figure 21 shows the analogous results to column (v) (tail symmetry) of Table 2 for different numbers of clusters K. Clearly, adding more clusters than 50 (the number of clusters used in all tables in the paper) do not change the estimates meaningfully. This

Figure 21: Uncorrected and Corrected Cognitive and Non-Cognitive Estimates



Note: Left figure shows cognitive estimates, and right figure shows non-cognitive estimates. Shaded areas depict the 90% confidence intervals. All standard errors are bootstrapped using 500 iterations.

suggests that our results are not an artifact of our discretization of X.

Note that as the number of clusters increases, so does the list of controls used in $m(X)$. Because the clusters are nested, as we add clusters, we increase the flexibility of m. Therefore, Figure 21 can be seen as an illustration of the results of a sequence of traditional omitted variable bias tests (e.g. Ramsey RESET test) in the specification of equation (6). The near constancy of the estimates in Figure 21 from $K = 50$ to $K = 100$ confirms that our approach is able to control for all confounders, including those due to mispecification of the function of controls $m(X)$.

A growing literature within economics explores the use of clustering techniques applied to group fixed effects estimators in panel settings (Lin and Ng, 2012; Bonhomme and Manresa, 2015; Bonhomme et al., 2017). Our use of clustering differs from these applications. We do not cluster on the outcome variable, and we do not use the clusters to handle endogeneity – that is accomplished through our correction term. Rather, the role of the clusters in our setting is to allow the distribution of unobservables to change with observables in a flexible yet tractable way.