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 of Governors

February 5th, 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 unstructured play and socializing, in order to spend time and attention on enrichment. Our study controls for selection on unobservables using a novel approach which leverages the fact that many children spend zero minutes per week on enrichment activities. These bunched observations allow us to build an endogeneity correction similar to Heckman (1979). Using time diary data available in the Panel Study of Income Dynamics (PSID), we find that the net effect of enrichment activities 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 particularly pronounced for middle-income youth in high school, consistent with increased academic competition related to college admissions. Codes: I21, I2, J01, C24

1 Introduction

Families spend substantial amounts of resources on activities intended to increase children's skills. These "enrichment" activities include homework, tutoring, lessons, and extra-curriculars such as music and art lessons. The money and time committed to these

*We thank Michael Ahn and Hannah Hall for excellent research assistance. 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 are substantial and increasing across the socioeconomic spectrum, leading to concerns that they may contribute to the growing 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, homework could have spillover effects into how a child spends the remainder of her day. For example, an exhausted child may not want to engage in active play after finishing their homework, preferring more passive activities, and the over-stimulation may lead them to fall asleep later than usual. Yet sleep and play are activities that generate a direct positive impact on the child's skills (Walker, 2017; Gray, 2019). The opportunity costs of enrichment activities might therefore be substantial depending on which activities they replace.

This paper is concerned with the estimation of the total effect of enrichment activities on cognitive and non-cognitive skills. That is, we seek to estimate the net effect of enrichment activities on 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 total effect on cognitive skills and a sizeable negative total 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., 2018). 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 control for the amount of time spent on active play, we shut off one of the indirect paths through which homework could negatively impact skills. The estimated positive effects of enrichment reflect only the direct effect of enrichment on skill, but the substitution and negative spillovers could cancel the positive effects and even net a negative total effect.

We propose a novel method to control for confounders which 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 ask why such bunching happens, and we argue that the choice of enrichment is the result of a constrained optimization problem which depends on both observed covariates and unobserved confounders, with the constraint that enrichment be non-negative. Therefore, the children who choose zero enrichment are different from the children who choose just a few minutes of enrichment per week because the group at zero includes all the children for whom the constraint is binding for any reason. That is, the group at zero includes children who for whom the constraint barely binds, as well as those for whom it binds with great force, depending on their observable and unobservable characteristics.

Consistent with this story, we show that the children at zero are discontinuously different in every observable way from the children just above zero. The same discontinuities should also hold for the unobservables. This creates an interesting situation, because although the unobservables are discontinuous, 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 the 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.

We find that enrichment time has a negative effect on cognitive skills, but we cannot reject a causal effect of zero. By contrast, our estimates for non-cognitive skills are quite negative and significant. These results are robust to several alternative definitions of enrichment time, cognitive and non-cognitive skills, as well as other sensitivity analyses. We also provide suggestive evidence that, at the margin, spending more time on enrichment activities has contributed to children playing and socializing less, which corroborates the findings in other studies (e.g. Gray, 2011; Lukianoff and Haidt, 2018).

To rationalize these results, we present a stylized model, inspired by Becker (1965), 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 returns 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. However, it is not generally possible to maximize both cognitive skills 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. Since enrichment activities tend to be chosen with the objective of improving cognitive skills, this model may explain our results.

We also explore how these results break down by household income and the child's grade in school. We find that spending more time on enrichment yields roughly zero returns to cognitive skills irrespective of income or grade, which is consistent with the idea that families attempt to maximize only cognitive skills when choosing enrichment. For the non-cognitive estimates, we find that at the margin enrichment activities are especially negative during high school for middle-income households. High-income children also seem to experience some negative effects in high school, but these are much smaller compared to middle-income children. This suggests that the cognitive/non-cognitive skill trade-off discussed above is especially important during the college preparation years, although high-income families may have additional resources to alleviate it.

Our results highlight the pitfalls and trade-offs associated with intensive investment in children's human capital. The perception among parents that such activities have high returns drives these investments. Many families, particularly those in the middle class with children in high school, strain to invest in an effort to increase the chance of admission to the best college possible. The stress that these high and rising investments place on both parents and children has been widely covered in the popular press.[1] Yet, we find that many children are spending so much time on enrichment activities that, on the margin, they are actively harming their non-cognitive skills. This is particularly relevant given widespread evidence of the importance of non-cognitive skills for the future development of both cognitive and non-cognitive skills and for key socioeconomic outcomes later in life (e.g. Cunha et al., 2006; Deming, 2009; Cunha et al., 2010; Heckman and Rubinstein, 2001).

The rest of the paper is organized as follows. Section 2 presents the data. Section 3 discusses the choice of enrichment and how this impacts our model. Section 4 presents our identification and estimation strategy, followed by the results in Section 5. Section 6 provides a model to help interpret the results. Finally, Section 7 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 are ideal for our purposes because they contain detailed time diary data and extensive measures of cognitive and non-cognitive skills. Additionally, the CDS data, when linked to the PSID, allow 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

[1] For recent examples see Avent, 2017; Rosin, 2015; Rosen, 2015; Khazan, 2016; Gray, 2010.

during the selected days are coded into one of over 300 different categories. These diaries are filled out 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, cases where either the weekday or weekend day data is missing, and cases 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., 2018). Finally, we aggregate the 300+ primitive time-use categories into enrichment time and a number of other natural categories (e.g. sleep, play and social activities, passive leisure, etc). Figure 1 shows the average breakdown of all the 24-hours per day among all the categories: enrichment activities, other enrichment activities, play and social activities, passive leisure, duties/chores, class time, sleep, and other.

Our definition of enrichment time is intended to capture the kinds of activities that the child development literature considers as investments in children's skills. Our baseline measure of enrichment time therefore includes only those activities that are unambiguously related to skill development. On average, children spend about 3% of their typical week on this type of enrichment. Figure 2 shows the breakdown of enrichment activity into various sub-categories.

The primary component of this baseline measure is homework, which is two-thirds of the total time spent on enrichment. The next most important component of enrichment is reading of any variety not associated with school, at 20% of the total. Our measure of reading includes reading different types of materials (books, newspapers, magazines) and also includes being read to by someone else (e.g., a parent). While 9% of enrichment time is spent on other academic activities not in a lesson format (before- and after-school programs, and other academic activities such as watching a documentary), 5% is spent in a lesson format. About half of the latter is spent on lessons more connected to academics (tutoring, academic courses and lectures), while the other half includes less academic

Figure 1: Daily Time Breakdown

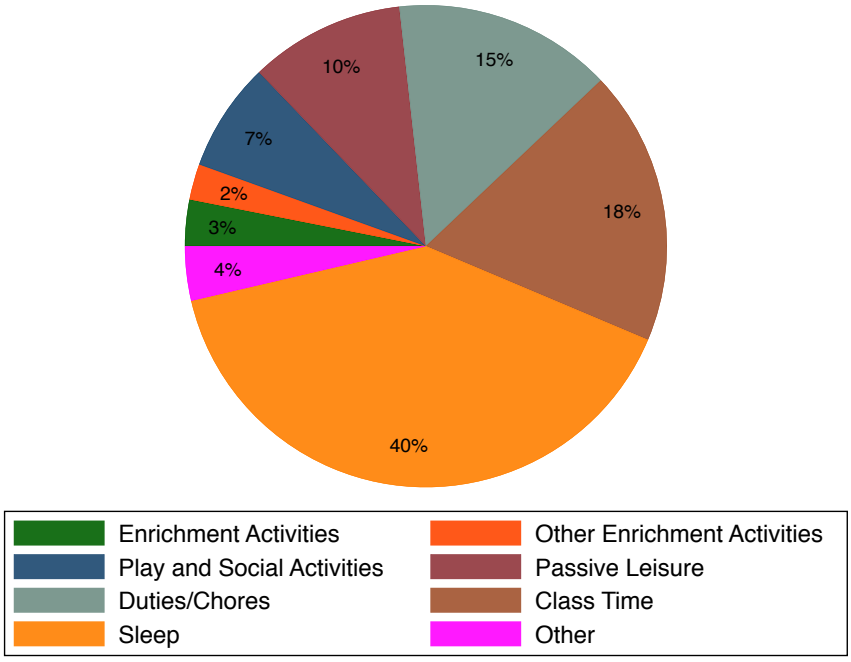
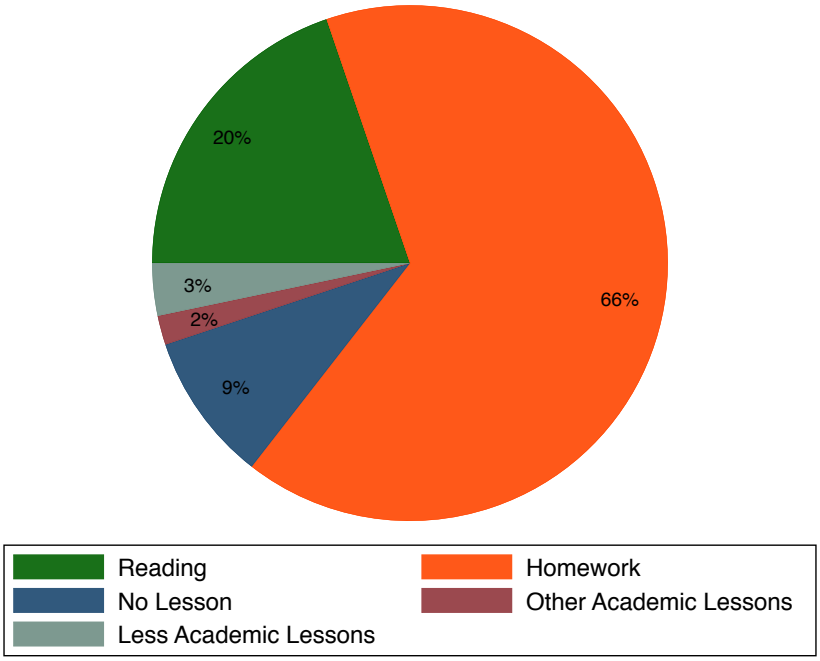


Figure 2: Enrichment Time Breakdown



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.

lessons outside of school hours: music, art, computers, etc.

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 16 presents the breakdown of "other enrichment" into its constituent pieces (Appendix, page 49).

We also define a number of other time aggregates which we will use in Section 6 to assess substitution patterns between enrichment and other activities. First, we define passive leisure as activities that do not involve active, interactive 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), interactive games (e.g., board games, hide and seek), hobbies, socializing, social and church groups, etc. 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. 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. The breakdowns of these categories can be seen in Figure 16 (Appendix, page 49).

We create the 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 all three CDS waves. We likewise construct our non-cognitive skill measure through iterated principle factor analysis applied to parental assessments picked from 36 questions

on the child's behavior. We also use the internalizing and externalizing subscales of the CDS-included behavior problems index (BPI) 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). Both the cognitive and non-cognitive measures are constructed so that a higher score is better and both are normalized to have a mean of zero and a standard deviation of one. Loading factors for these iterated principle factor analyses are shown in Table 6 in Appendix C.

Table 1 presents summary statistics for our analysis sample. We are left with a pooled sample of 4,330 observations with an average age of just under 12. Not surprisingly, across a typical week, sleep is the single biggest use of time, followed by class time, duties/chores, passive leisure (mostly watching TV) and play. 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 in our sample 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, 8% attend a private school, and 1% are home schooled.

Throughout the paper, we denote by X the vector of observed child, family and environmental characteristics which we will 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 strategy because the selection correction we develop, explained in subsequent sections, can handle bias stemming from endogenous unobservables. Our list of controls includes child's age (in months) and squared age, and indicators for: CDS wave (1997, 2002 and 2007), grade (thirteen variables, from kindergarten until grade 12), gender, ethnicity (black, Hispanic

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 is in Gifted Program	0.26	0.43
Child is in Special Education Program	0.08	0.27
Child's Father is Alive	0.97	0.16
Child's Mother is Alive	0.99	0.08
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.

and other non-white ethnicity), whether the child has siblings, tercile of the family income, whether child is in a gifted program, whether child is in a special education program, whether the mother is alive, whether the father is alive, whether child is home schooled, and whether child attends a private school. 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. Moreover, our correction strategy renders the use of lagged skills to control endogeneity unnecessary. Indeed, the ability to credibly estimate causal effects without the use of lagged scores is a major advantage of our approach.

Finally, we note that implementing our preferred (most general) selection correction requires that we discretize X. We discretize X using hierarchical clustering, in which observations are grouped into clusters based on the similarity of their observables. For details, please refer to section B in the Appendix. We also show there that our results do not appear to be an artifact of the discretization of X.

3 Modeling the Effects of Enrichment on Skills

Consider the standard regression model

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

where S refers to the child skill (cognitive or non-cognitive), 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 A (p. 45), we apply Caetano (2015)'s test of exogeneity and show that the evidence of endogeneity of I in the model above is overwhelming. Moreover, we show that the bias in the estimation of β is likely to be positive, so that a naive regression of S on I and X would over-estimate the effect of enrichment on skills.

Without additional structure or assumptions, the effect of enrichment in the model above, β, cannot be identified. However, the enrichment time variable I is of a peculiar nature, and its characteristics can be leveraged to identify β. Specifically, in the remainder of this section, we will show that I has substantial bunching at zero (Section 3.1) and that this bunching is likely to be the result of a constrained choice problem (Section 3.2). In other words, 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. We thus add some structure to the model above to reflect this constrained choice and resulting bunching (Section 3.3). In Section 4 we show how β can be identified in the augmented model under three progressively weaker assumptions on the shape of the distribution of the confounders.

3.1 Enrichment is bunched at zero.

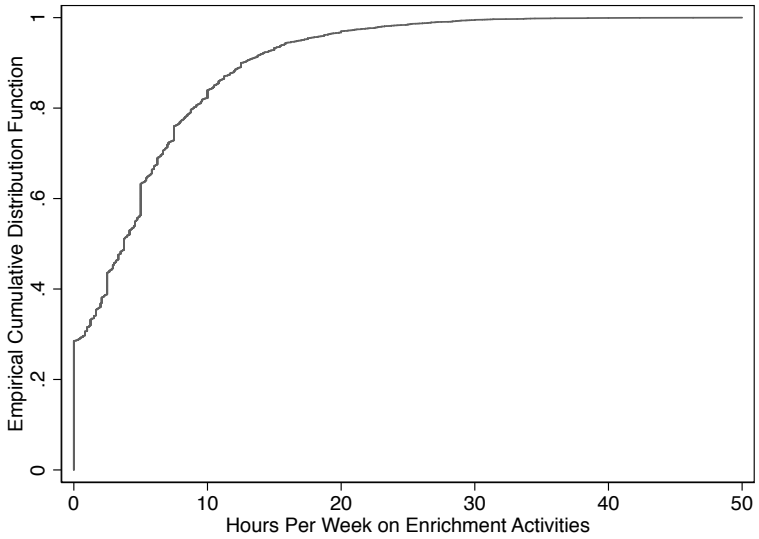
Figure 3 plots the empirical cumulative distribution function of enrichment time in our sample and shows that there is substantial bunching at zero. About 30% of children spend no time on enrichment, while the rest are continuously distributed among different, positive levels of enrichment.

3.2 Enrichment is a constrained choice.

We explain this bunching by considering that enrichment time is a choice variable. If enrichment time is chosen as a function of the child and family characteristics, both observable and unobservable, the optimization process may yield an optimal negative choice for some families. However, since negative choices are impossible, such families will be constrained to spend zero time on enrichment.

We are able to test that the constraint is binding for some families while simultaneously furthering our understanding of the enrichment variable. To develop insight into our approach, consider a potential confounder (a variable correlated with I and S). We will

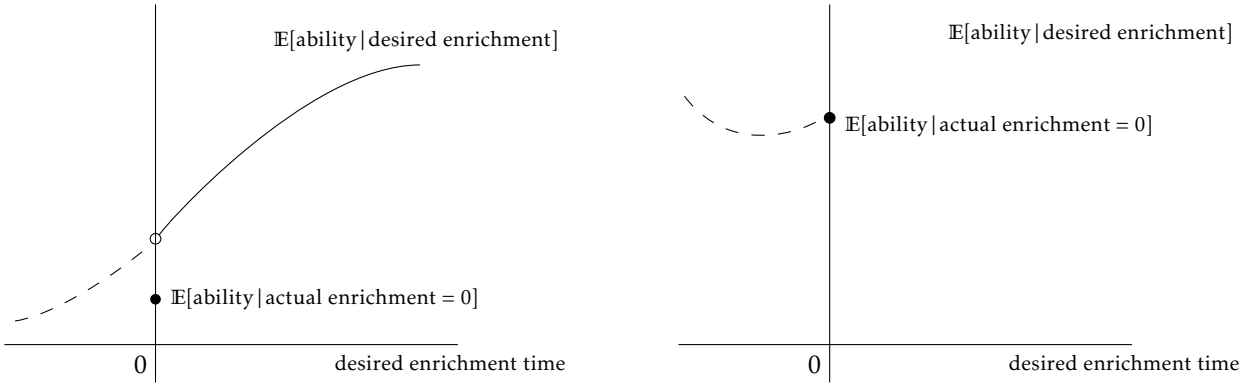
Figure 3: Evidence of Bunching at Zero Enrichment Time



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

refer to this variable as the child's "ability" to put things in more concrete language, but one could substitute any possible confounder in this problem. The left panel of Figure 4 illustrates the relationship between ability and the *desired* enrichment time.

Figure 4: Hypothetical Relationship between Child's Ability and Enrichment Time



Note: The solid and dashed lines show the expected ability of the child for each level of desired time spent on enrichment activities. For the solid line the desired enrichment equals the actual enrichment chosen. However, for the negative amounts of desired enrichment the actual choice of enrichment is zero, and the expected ability of the entire group with negative desired enrichment is the black dot. The right panel shows that if ability is a confounder at zero and the constraint is binding, the only way for the expected ability to be continuous at zero is if the relationship between desired enrichment and ability changes derivative and by coincidence averages to exactly the value of the intercept.

The solid curve and the black circle at zero enrichment represent the expected ability

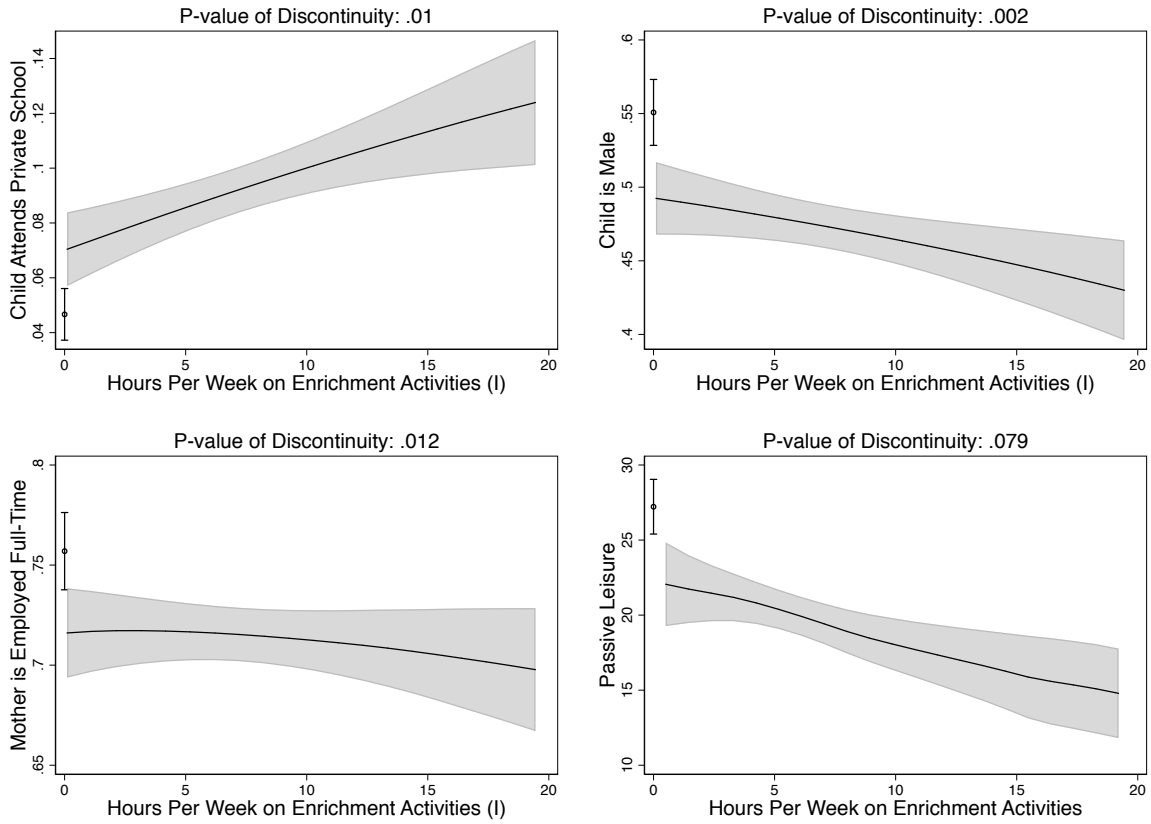
among the children who chose each level of enrichment. Whenever enrichment time is positive, the desired and the actual levels of enrichment are the same. The dashed curve, which represents the expected ability conditional on desired enrichment among those with negative desired enrichment, shows a smooth continuation of the pattern on the positive side (the solid curve). The idea is that if ability is an input in the choice of enrichment, lower ability levels would be associated with lower, and even negative, choices of enrichment.

Because of the constraint that enrichment cannot be negative, the equivalence between actual enrichment and desired enrichment is broken at zero. The black circle is the weighted average of the height of the entire dashed curve on the negative side. Therefore, if the constraint that enrichment cannot be negative is binding for some children, we should find the expected ability to be discontinuous at zero enrichment.

Such a discontinuity should be expected for nearly all potential confounders. If the non-negativity restriction is binding, the expected value of ability will be continuous at zero enrichment only in two special cases. The first case is if the relationship between ability and enrichment is flat for negative enrichment. In this case, although ability is a confounder for positive values of enrichment, it is not a confounder for negative values of enrichment. The second case is if the relationship between ability and enrichment is non-monotonic for negative enrichment in precisely such a way that the average ability among the children who desire negative enrichment is exactly the same as the intercept. The right panel of Figure 4 depicts this second case. Both of these cases are likely to be exceptions; we therefore would expect to see a discontinuity at zero for almost all confounders.

Consistent with this discussion, we find empirically that children who spend no time on enrichment are discontinuously different from children who spend any positive time at all. The upper left panel in Figure 5 shows a local linear fit of the child's private school enrollment conditional on the amount of time the child spends on enrichment activities

Figure 5: Evidence that Bunching is Selective



Note: Each panel shows a plot of the local linear polynomial estimator of the expected value of a variable conditional on enrichment time, along with its 90% confidence interval (bandwidth=15 hours). 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.

(analogous to the solid curve in Figure 4), as well as the private school enrollment rate among the children who spend zero time on enrichment (analogous to the black circle in Figure 4). The children at zero are discontinuously less likely to be attending a private school 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 private school attendance is continuous at zero enrichment time, and it is clear that we can confidently reject this hypothesis ($p = 0.01$). The other panels on Figure 5 show similar patterns. Children who spend no time on enrichment are discontinuously more likely to be male ($p = 0.002$), to have a full-time working mother ($p = 0.012$), and they spend discontinuously more time

on passive leisure activities ($p = 0.079$).[2]

The last panel in Figure 5 reflects the stark difference between the lives of the children who spend no time on enrichment and everybody else. In particular, begin by noting that the children at zero enrichment spend five more hours per week on passive leisure than children who do one hour of enrichment. Since the total number of hours in a week is the same for everyone, this means that children at zero enrichment are spending five more hours of passive leisure while spending five fewer hours on other activities, of which one hour is enrichment. The remaining four hours could be devoted to anything, including potentially more productive activities such as sports, socializing, and sleep.

3.3 A model with constrained enrichment choice

Consider the original model given by equation (1). We now separate the error term into two two components: η, which indexes all the unobservable confounders, and ε, which indexes all non-confounders (i.e. any variables which affect S but are not correlated with I conditional on X),

$$S = \beta I + h(X) + \delta\eta + \varepsilon, \quad \mathbb{E}[\varepsilon|I, X] = 0. \tag{2}$$

This is a selection on unobservables model which assumes that I is exogenous only if we condition on the unobservable term η as well as observed controls X.

Motivated by the previous discussion, we specify the desired enrichment choice I^* as a function of observable, X, and unobservable, η, factors

$$I^* = g(X) + \eta. \tag{3}$$

Note that we are not making any assumptions at this stage about the relationship between η and X, nor are we supposing that g has any causal meaning or is identified in any way.

[2] In Appendix Section A, we present additional evidence using the test introduced in Caetano (2015) that I in the naive model is endogenous and that the resulting bias is positive.

Finally, the actual enrichment choice I is constrained to be non-negative:

$$I = \max\{0, I^*\}. \tag{4}$$

4 Identification Strategy

Equations (2), (3) and (4) 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) clarifies the intuition of our strategy. Suppose that there exists a group of children for whom the constraint is truly binding, so that $I = 0$, but $I^* < 0$. Why do we see skill variation in this group? Logically, the variation is not due to differences in time spent on enrichment activities 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 $h(X) - \delta g(X)$. Importantly, however, another part of the skill variation is due to differences in unobservables, specifically through $\delta\mathbb{E}[I^*|I^* \leq 0, X]\mathbf{1}(I = 0)$. 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 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, 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, similarly to the selection correction in Heckman (1979). 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.

Our identification strategy therefore relies on two key assumptions. First, we assume that η enters linearly in equation (2), allowing us to extrapolate the role of the confounder in the OLS bias from $I = 0$ (where we are able to identify it) to $I > 0$. Section 4.1 discusses this first assumption and how it might be relaxed.

Second, we assume that we can identify $\mathbb{E}[I^*|I^* \leq 0, X]$. How can this be achieved? 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 $I^* \leq 0$. The key idea is that under some assumptions about the shape of the distribution of the confounders η, the observations with $I^* > 0$ will be informative about the observations with $I^* < 0$. In Section 4.2, we discuss three alternative strategies for the identification of $\mathbb{E}[I^*|I^* \leq 0, X]$ that require progressively weaker assumptions on the shape of the distribution of the confounders.

4.1 The confounders affect skills linearly.

Our model builds off equation (1), which already specifies a general linear structure. Note that formally η is defined in equation (3) without loss of generality as the residual of I^* when we take out $h(X)$. We then assume that η enters linearly in the structural equation. Thus, our error term in the structural equation is $\delta\eta + \varepsilon$ as opposed to $\delta f(\eta) + \epsilon$ or some other more general expression.

Both the linear specification of equation of (1) as well as the linear specification of the error term ϵ can be relaxed in a more flexible non-parametric model, as discussed in Caetano et al. (2019). 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 confounders are similar in nature and their effects also follow a stable pattern at zero and everywhere else.

We are therefore not controlling for any confounders that only affect children who do

positive amounts of enrichment. If such confounders have positive effects (so that the confounders associated with more enrichment are also associated with more skills), then our estimates represent upper bounds on the true effects. Considering that our findings are that the effects of enrichment on skills are either zero or negative, the existence of this kind of endogeneity would only strengthen our conclusion.

Similarly, if there are confounders with positive effects that only affect children around zero enrichment, but not children who spend higher positive amounts of time on enrichment, then our results would be over-correcting. In this case, enrichment would have an average effect that is higher than what we estimate.

Effectively, our method estimates the average treatment effects corrected for the endogeneity which appears around zero enrichment.

4.2 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. The technical details of all these methods can be found in Caetano et al. (2019).

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 $g(X) = X'\pi_1$ and that

$$\eta|X \sim \mathcal{N}(X'\pi_2, \sigma^2). \tag{6}$$

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 determined simply by looking at any portion of the

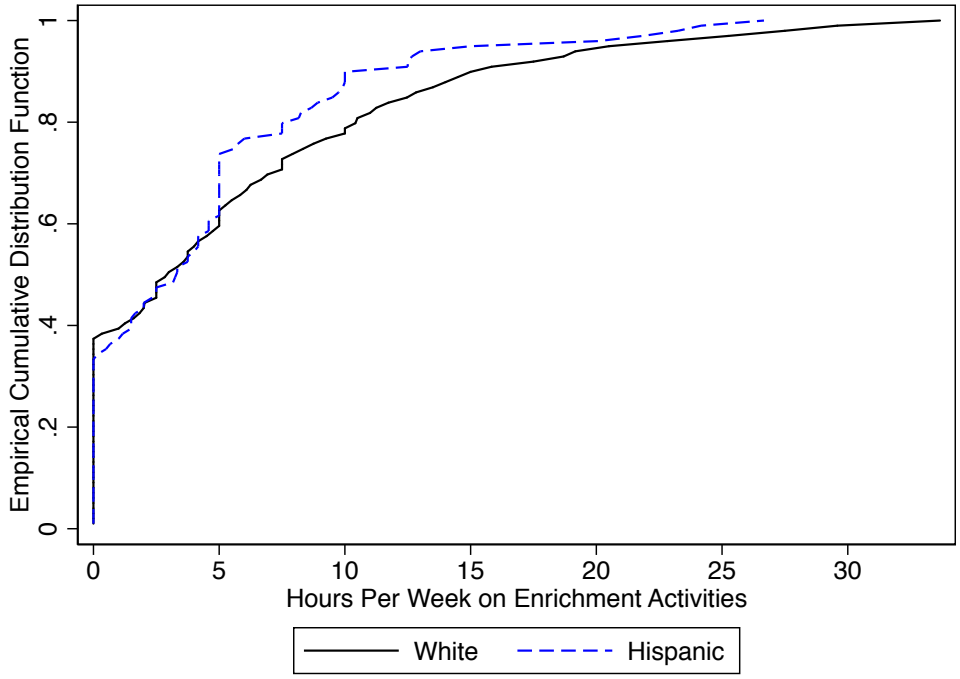
distribution of enrichment time above zero. When equation (6) holds,

$$\mathbb{E}[I^*|I^* \leq 0, X] = X'\pi - \sigma\lambda(-X'\pi/\sigma) \tag{7}$$

where $\pi = \pi_1 + \pi_2$ 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 (7) 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 6 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 the distribution of $\eta|X$ per equation (3). Indeed, the figure shows that the homoskedasticity assumption clearly does not hold, as the variance of I is different for different values of X.

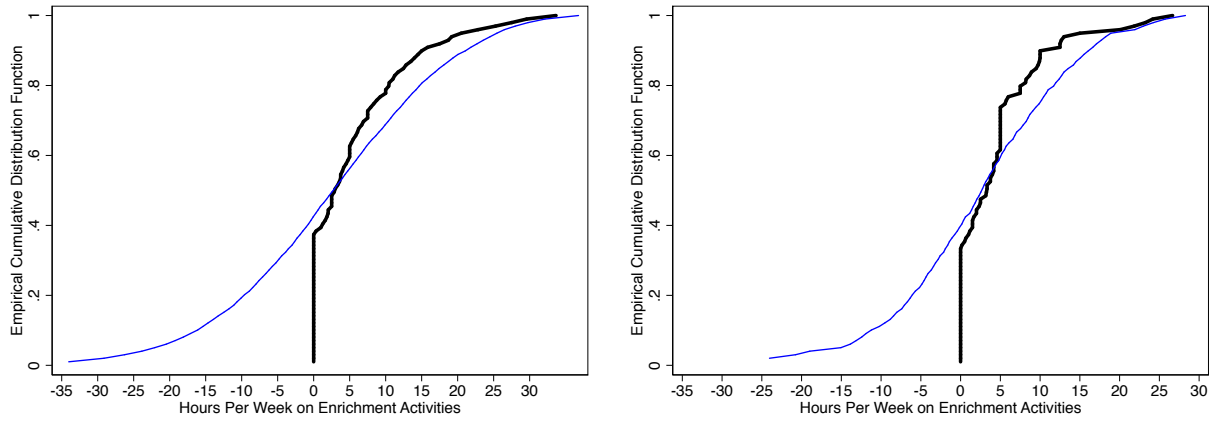
Figure 6: Evidence of heteroskedasticity on the distribution of $I|X$



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 7 shows the homoskedastic Tobit fit for White (left panel) and Hispanic (right panel) high school students, along with their corresponding empirical CDF of I. We are not satisfied with either fit on the positive side of I, which would lead us to over-estimate the magnitude of $\mathbb{E}[I^*|I^* \leq 0, X]$.

Figure 7: 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 6 (thick curve) along with the corresponding homoskedastic Tobit fit (thin curve). 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 $\mathbb{E}[I^*|I^* \leq 0, X]$.

Heteroskedastic Tobit

Here we relax the homoskedasticity requirement while maintaining the assumption that $\eta|X$ follows a normal distribution. Specifically, we suppose that

$$\eta|X \sim \mathcal{N}(l(X), \sigma^2(X)). \tag{8}$$

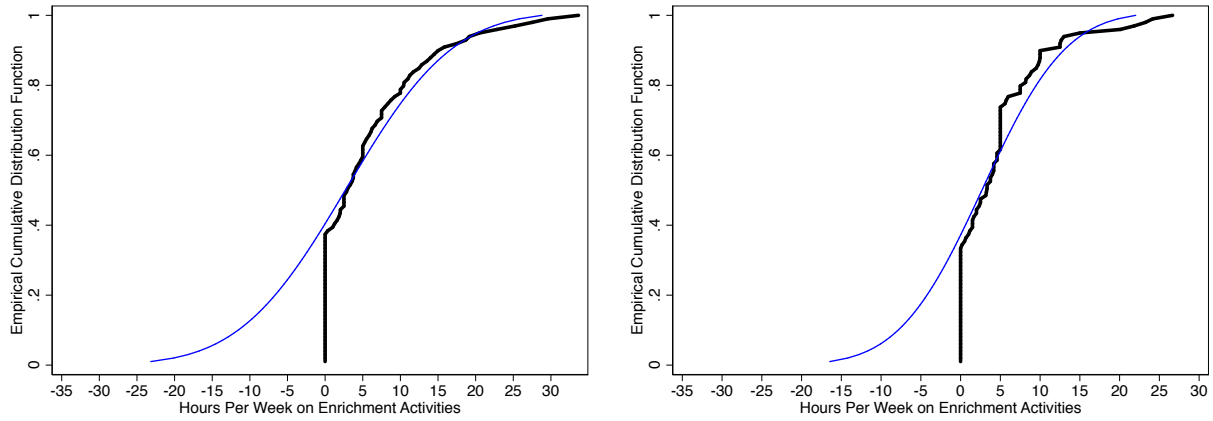
This assumption allows the mean and the variance of η to vary with X in an unrestricted way, but retains the requirement that η be normal separately for each value of X. In this case,

$$\mathbb{E}[I^*|I^* \leq 0, X] = g(X) + l(X) - \sigma(X)\lambda\left(-(g(X) + l(X))/\sigma(X)\right). \tag{9}$$

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

Figure 8 plots the empirical conditional cumulative distribution function I for the same two values of controls (White and Hispanic high school students) in Figure 6, along with the corresponding heteroskedastic Tobit fits for each. The fits are satisfactory and are clearly superior to the homoskedastic fits presented previously. 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 observed 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 underestimate $\mathbb{E}[I^*|I^* \leq 0, X]$, thus overestimating δ and underestimating β in our setting.

Figure 8: 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 6 (thick curve) along with the corresponding heteroskedastic Tobit fit (thin curve). Plots show evidence of a better fit than the homoskedastic case for positive values of enrichment, but still there seem to be fatter tails in the empirical distribution compared to the heteroskedastic normal fit, which may lead us to under-estimate the magnitude of $\mathbb{E}[I^*|I^* \leq 0, X]$.

Heteroskedastic Symmetric

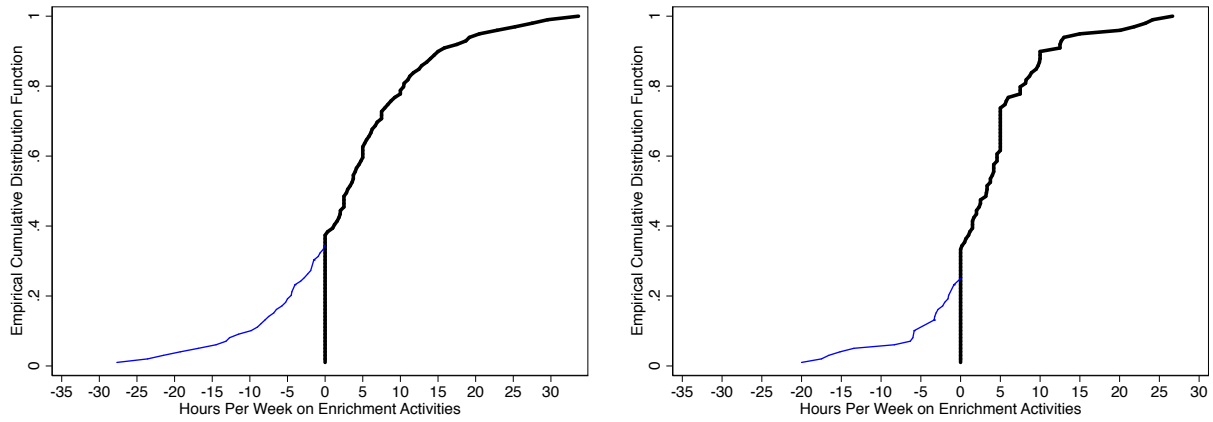
Finally, we drop the normality assumption entirely and require only tail symmetry: for all censored quantiles q_0,

$$\eta|X \text{ has symmetric tails below } q_0 \text{ and above } 1-q_0. \tag{10}$$

Symmetry in the tails simply requires that the lower tail of $\eta|X$ below the censoring point and the corresponding upper tail are 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 $I|X$.[3]

To see this assumption in action, Figure 9 provides the corresponding plots shown in Figure 8 under tail symmetry. For quantiles below the bunching threshold, the fitted values follow the mirror image of the corresponding upper tail.

Figure 9: 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 6 (thick curve) along with the corresponding heteroskedastic Symmetric fit (thin curve).

[3]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. 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.

Under symmetry on the tails,

$$\mathbb{E}[I^*|I^* \leq 0, X] = F^{-1}_{I|X}(1 - F_{I|X}(0)) - \mathbb{E}[I|I \geq F^{-1}_{I|X}(1 - F_{I|X}(0)), X], \tag{11}$$

where $F_{I|X}(\cdot)$ is the cumulative density function of I conditional on X. We carry out the estimation of $\mathbb{E}[I^*|I^* \leq 0, X]$ using equation (11) in three steps. For each value of X, we first estimate the probability of bunching at zero enrichment (i.e., $F_{I|X}(0)$), then we estimate the value of I in the upper tail that corresponds to the mirror image of $I = 0$ (i.e., $F^{-1}_{I|X}(1 - F_{I|X}(0))$), and finally we estimate the mean of $I|X$ at the upper tail (i.e., $\mathbb{E}[I|I \geq F^{-1}_{I|X}(1 - F_{I|X}(0)), X]$).

5 Empirical Results

5.1 Main Estimates

Table 2 presents our main results. Column (i), which shows the results of simple regressions of skills on enrichment time without controls (equation (1) without Xs), demonstrates that both cognitive and non-cognitive skills are strongly positively correlated with enrichment time. Column (ii) adds controls (equation (1)). While the controls seem to explain part of the correlation between enrichment time and skills, the residual relationships remain positive, particularly for cognitive skills.

The evidence discussed in Appendix A (see Figure 14, p. 46), as well as economic intuition about the likely sources of confounders, 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 under the different assumptions on the distribution of $\eta|X$ discussed in Section 4.2. Column (iii) shows the results when we implement the endogeneity correction using the Tobit strategy, which assumes that $\eta|X$ is normally distributed with a mean that is linear in X and a variance which does not vary in X. Column (iv) shows

Table 2: Main 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***	0.008***	-0.007	-0.006	-0.008
		(0.002)	(0.001)	(0.006)	(0.005)	(0.005)
	δ			0.012**	0.011***	0.013***
				(0.005)	(0.004)	(0.004)
Non-Cognitive	β	0.006***	0.001	-0.015	-0.022**	-0.019**
		(0.002)	(0.002)	(0.010)	(0.009)	(0.009)
	δ			0.013*	0.019**	0.017**
				(0.008)	(0.007)	(0.007)

Note: N=4,330. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications use 50 clusters. See Figure 15 in Appendix B for analogous results with different numbers of clusters. *** p<0.01, ** p<0.05, * p<0.1.

the results when $\eta|X$ is assumed to be normal with a mean and variance that depend arbitrarily on X. Finally, column (v) shows the results under the weakest set of assumptions in the paper, which assumes that the tails of $\eta|X$ are symmetric. All standard errors are bootstrapped using 1,000 iterations.

For cognitive skills, all the corrected estimates are quite similar – the estimated β's fall from 0.008 standard deviations (s.d.) to around -0.007 s.d. and change from being highly statistically significant to insignificant. The large differences between column (ii) and columns (iii)-(v) show that our correction method is able to handle endogeneity which was not absorbed by the pre-determined controls. After correcting for selection on unobservables, our most general method yields a 95% confidence interval of $[-0.018, 0.002]$. Therefore there is no evidence that enrichment has more than a negligible positive causal effect on cognitive skills and some evidence that the effect may actually be modestly negative.

Correcting for selection has even more dramatic consequences for the non-cognitive estimates – the corrected non-cognitive β's are negative and statistically significant, with

point estimates ranging between -0.022 and -0.015 s.d. Our preferred method (column (v)) yields a 95% confidence interval of $[-0.037, -0.001]$, which is evidence of a decidedly negative effect, with some evidence that the effect may be substantially negative.

For both cognitive and non-cognitive skills, the estimated δ is positive and highly significant for our preferred specification (last column), confirming the evidence we presented in Section 3 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 further evidence of positive bias.

Note that the uncorrected standard errors from Table 2 (column (ii)) are much smaller than the corrected ones (columns (iii)-(v)). This is a feature of our approach, not a problem. 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 to blow up, so the fact that we see an increase in the standard errors suggests greater underlying uncertainty surrounding the true causal effects of enrichment time on skills. In contrast, not considering this correction term would lead to not only biased estimates, but also to unrealistically precise conclusions, leading to overly-optimistic expectations of policymakers or families regarding the impact of enrichment activities.

Table 7 in p. 52 of Appendix C 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 skill (applied problems, letter-word comprehension, and passage comprehension), we find sizeable, positive uncorrected estimates and negative, statistically insignificant corrected estimates. Next, we consider alternative measures of non-cognitive skills based on the internalizing and externalizing subscales of the CDS-included behavior problems index (BPI).[4] Here, the uncorrected estimates suggest significant, positive effects for externalizing problems only, while the

[4]The internalizing scale measures inward-facing behaviors such as dependence on adults, crying easily, and social withdrawal. The externalizing scale measures outward-facing behaviors such as lying, bullying, breaking things, hyperactivity, and impulsivity.

corrected estimates for both scales are negative and similar in magnitude.

5.2 Estimates by Parental Income

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 family income. As shown in Figure 17 in p.51 of Appendix C, richer families tend to spend more time on enrichment activities, a finding consistent with prior literature (Duncan and Murnane, 2011; Doepke and Zilibotti, 2019). Further, not only do richer families spend more time on enrichment, their time use may be more efficient, yielding higher returns in skills per hour spent. These higher returns could come from differences in "parenting style" (Doepke and Zilibotti, 2019), or through complementarities with parental human capital (better educated parents may be more efficient at helping their children learn the material as they are doing their own homework), or through other channels (mentorship networks, neighborhood safety, stress, etc.). Additionally, depending on the level of income, children may substitute the last hour of enrichment for different activities, some perhaps more productive than others. These considerations suggest that the average treatment effects may be different for families in different parts of the income distribution.

In this section, we test whether the estimated effects vary with income by applying our method separately for children in each tercile of the family income distribution. To summarize the findings we will discuss in the following pages, we do not find strong evidence of differential effects of enrichment time on cognitive skills by family income. In contrast, we do find that the negative effects of enrichment time on non-cognitive skills are concentrated away from the bottom, with particularly negative effects in the middle tercile.

The estimates of the effect of enrichment on cognitive skills for each tercile of the income distribution are presented in Table 3. If we only considered the uncorrected

estimates, we would think that children in the bottom income tercile gain the most from enrichment time – each additional hour of enrichment in this group is associated with a statistically significant increase in cognitive skills of 0.011 s.d. The top two income terciles also show significant, but much smaller, effects of around 0.004 s.d. These results are consistent with the idea that the returns to skills investments may be greater for more disadvantaged youth.

However, the corrected estimates in Table 3 provide little evidence of differential cognitive returns to enrichment by family income. The corrected bottom tercile effects are negative and not statistically significant – lower-income youth do not seem to have a positive return to enrichment on the margin. Similarly, the corrected estimates for the top two terciles are likewise negative and not distinguishable from 0. It seems that the larger apparent effects of enrichment for lower-income youth are due entirely to differential selection on unobservables. If anything, the corrected effects at the bottom of the income distribution are more negative, although the differences by tercile are not statistically significant.

Table 3: Cognitive Estimates by Family Income Levels

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Low Income	β	0.008	0.012***	-0.012	-0.012	-0.009
		(0.005)	(0.003)	(0.011)	(0.012)	(0.013)
N=1434	δ			0.019**	0.020**	0.017
				(0.009)	(0.010)	(0.011)
Middle Income	β	0.006	0.007**	-0.001	-0.005	-0.013
		(0.004)	(0.003)	(0.009)	(0.010)	(0.010)
N=1434	δ			0.006	0.009	0.015*
				(0.007)	(0.008)	(0.008)
High Income	β	0.017***	0.005***	-0.005	-0.009	-0.008
		(0.003)	(0.002)	(0.010)	(0.008)	(0.007)
N=1434	δ			0.009	0.011*	0.011*
				(0.009)	(0.007)	(0.006)

Note: Number of observations (N) for each tercile of income is shown. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications use 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.

Table 4 repeats the analysis for non-cognitive skills. As with cognitive skills, the uncorrected estimates are consistent with the idea that the returns are highest for the lowest income group – the top two terciles have uncorrected estimates of 0, while the bottom tercile estimate, at 0.010 s.d., is positive and significant at the 5% level. After correcting for endogeneity, this bottom-tercile effect remains positive but is no longer significantly different from zero. Correcting for selection also reduces the non-cognitive estimates for the top two terciles substantially. At the margin, middle-income children seem to have significantly negative effects of enrichment time on non-cognitive skills, with effects of about -0.05 s.d. per hour. The high-income estimates are 0 or perhaps modestly negative. These results are in contrast to the low-income group, whose effect seems to be zero or perhaps even modestly positive.

Table 4: Non-Cognitive Estimates by Family Income Levels

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Low Income	β	0.009*	0.010**	0.016	0.007	0.009
		(0.005)	(0.005)	(0.018)	(0.018)	(0.019)
N=1434	δ			-0.004	0.003	0.001
				(0.014)	(0.014)	(0.016)
Middle Income	β	0.004	-0.001	-0.032*	-0.055***	-0.055***
		(0.005)	(0.005)	(0.019)	(0.019)	(0.020)
N=1434	δ			0.025*	0.044***	0.044***
				(0.015)	(0.015)	(0.017)
High Income	β	0.000	-0.002	-0.008	-0.014	-0.009
		(0.003)	(0.004)	(0.015)	(0.013)	(0.012)
N=1434	δ			0.005	0.010	0.005
				(0.013)	(0.011)	(0.009)

Note: Number of observations (N) for each tercile of income is shown. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications use 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.

5.3 Estimates By Grade

We want to understand why the non-cognitive effects are so pronounced for the middle-income children. For this, we further break down the estimates based on the grade the child is attending. Table 5 presents the non-cognitive estimates for children of middle income families attending grades preK-5, 6-8 and 9-12 respectively. The disaggregated results reveal an interesting pattern: the negative non-cognitive effect for middle-income children documented in the previous section is driven mostly by a very negative effect in grades 9-12. We also find negative effects for this income group in other grade ranges, but those estimates are insignificant.

Tables 8 and 9 in Appendix C show the analogous estimates for the low- and high-income groups, respectively. We find positive effects for high-income children attending

grades ranging from Pre-K to grade 5, with negative but insignificant results for later grades. We also find positive but insignificant results for low-income families.

Tables 10-12 in Appendix C show the associated tables for cognitive skills for the low-, middle- and high-income groups. All cognitive estimates are insignificant irrespective of the combination of income and grade.

Table 5: Non-Cognitive Estimates by Grade - Middle Income

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
PreK-5	β	-0.011	-0.008	-0.011	-0.018	-0.013
		(0.009)	(0.009)	(0.046)	(0.040)	(0.032)
N=449	δ			0.002	0.008	0.004
				(0.039)	(0.034)	(0.026)
6-8	β	-0.002	-0.011	-0.029	-0.015	-0.030
		(0.009)	(0.010)	(0.038)	(0.038)	(0.035)
N=456	δ			0.015	0.003	0.015
				(0.031)	(0.032)	(0.029)
9-12	β	0.022**	0.013	-0.021	-0.074***	-0.080**
		(0.009)	(0.008)	(0.026)	(0.027)	(0.032)
N=529	δ			0.025	0.064***	0.073***
				(0.018)	(0.020)	(0.025)

Note: Number of observations (N) for each grade level is shown. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications include 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.

6 Discussion

6.1 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? Although we do not attempt to answer these questions definitively here, we discuss one possible rationalization using a stylized model of time allocation inspired by Becker (1965).

This model aims at capturing 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, a greater return than enrichment, then the net effect of spending more time on enrichment would be negative.

The second idea is that families will not generally be able to choose enrichment time so as to jointly maximize both cognitive and non-cognitive skills.

So, 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 relate to enrichment and leisure according to

$$S_c = f_c(I, L), \quad S_{nc} = f_{nc}(I, L), \tag{12}$$

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

Now consider a 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

$$\frac{\partial}{\partial I} f_c(I_c, 1 - I_c) = \frac{\partial}{\partial L} f_c(I_c, 1 - I_c). \tag{13}$$

Equation (13) just 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 families chose enrichment so as to maximize cognitive skills, the marginal effect of enrichment on

cognitive skill should be

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

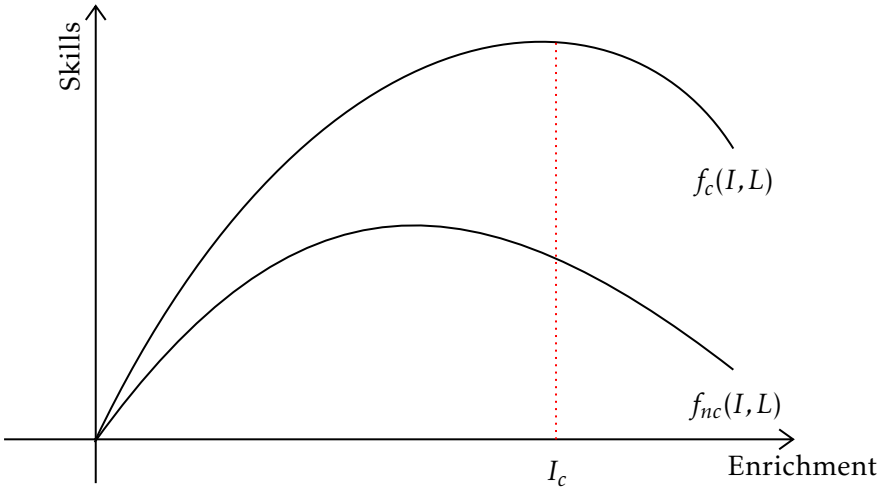
This example considers only two activities, enrichment and leisure. However, a similar result holds when there are more than two activities. This conclusion also continues to approximately hold even if some families bunch at $I_c = 0$, provided that enough families choose $I_c > 0$.

It is clear that the optimal amount of enrichment for cognitive skills, I_c will be generally different from the optimal amount of enrichment for non-cognitive skills, I_{nc}. In fact, considering how enrichment activities tend to be more directly associated with the improvement of cognitive skills, it is likely that $I_c > I_{nc}$. Therefore, if non-cognitive skills are a concave function on the enrichment activities, we expect that the level of enrichment that maximizes cognitive skills will be at the range of negative returns in the non-cognitive skills curve. Figure 10 illustrates this point. It plots the production functions for cognitive and non-cognitive skills as a function of enrichment time. The cognitive-maximizing point, I_c, shown with a red vertical dotted line, lies to the right of the non-cognitive-maximizing point. At I_c, the marginal return to enrichment is zero for cognitive skills and negative for non-cognitive skills, reflecting the findings of Section 5.

This model can also explain why our non-cognitive estimates are most negative for middle-income youth in high school. Because of competition for grades and college placement, this group might feel the strongest pressure to excel on cognitive skills (perhaps even on dimensions not captured by test scores), at the expense of noncognitive skills.

This stylized model depends on two key assumptions: (a) parents choose enrichment activities so as to maximize cognitive skills and (b) the net returns to enrichment decline

Figure 10: Cognitive and Non-cognitive Skills Production



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. The top curve shows how cognitive skills vary with enrichment time, all else constant. The lower curve shows the analogous relationship for non-cognitive skills. I_c is the level of enrichment that maximizes cognitive skills. At that optimum, the net effect of enrichment on cognitive skills is zero and its corresponding net effect on non-cognitive skills is negative.

more quickly with enrichment time for non-cognitive skills than for cognitive skills. Are these assumptions reasonable? Many of the activities that comprise enrichment are explicitly oriented toward the development of cognitive skills. To test these assumptions, we extended the notion of enrichment to also include other activities less obviously intended for the development of cognitive skills such as computer (educational), organized sports, arts and volunteering (see top left panel from Figure 16 in p.49 of the Appendix C). The idea is that while families likely choose homework so as to maximize cognitive skill, they may choose art classes to maximize something else, perhaps even non-cognitive skill. Thus, if (a) and (b) hold for the original definition of enrichment, we would expect the effects of enrichment on non-cognitive skills using the broader definition of enrichment to be less negative. Table 13 in the appendix provides the analogous baseline results from Table 2. Indeed, the non-cognitive estimates are still negative but smaller in magnitude, thus no longer significant.

In practice, families are likely unaware of the exact returns of each activity, and it is

also clearly unrealistic to assume that they allocate time solely to maximize cognitive skills. Time uses could have returns along dimensions not well measured by achievement tests. For instance, "excessive" enrichment activities might be harmful for non-cognitive skills (our basic finding) but nonetheless have a high return in the form of improved college admission prospects. Alternatively (and not mutually exclusively), some activities may be undertaken simply because they are enjoyable. Nonetheless, we believe this stylized model offers an interesting intepretative benchmark for our empirical estimates.

6.2 Is the Income Gap in Enrichment Activities Contributing to Inequality?

The fact that children in the higher income terciles spend more time on enrichment than those in the lower tercile (see Figure 17 in p. 51 of Appendix C) naturally leads to concerns that differences in time use might be contributing to the intergenerational transmission of inequality. Our uncorrected estimates lend support to these concerns – the cognitive and non-cognitive effects of enrichment are positive and significant for everyone, and are larger for lower-income children. These estimates suggest that closing the "enrichment gap" between high- and low-income children would help close the sizeable cognitive and non-cognitive skill gaps between these groups.

Our corrected estimates tell a more nuanced story. For cognitive skills, we cannot reject that high-, middle- and low-income families are all investing efficiently (i.e., up to the point that $\beta = 0$). Because higher-income families devote more total time to enrichment, this suggests in turn that such families have access to superior skill production technology. However, because our estimates also suggest that lower-income families are already investing efficiently *given their skill-production technology*, we do not find evidence that encouraging lower income families to undertake more enrichment would lower the cognitive skills gap. On the margin, they can only decrease their cognitive skills by increasing the

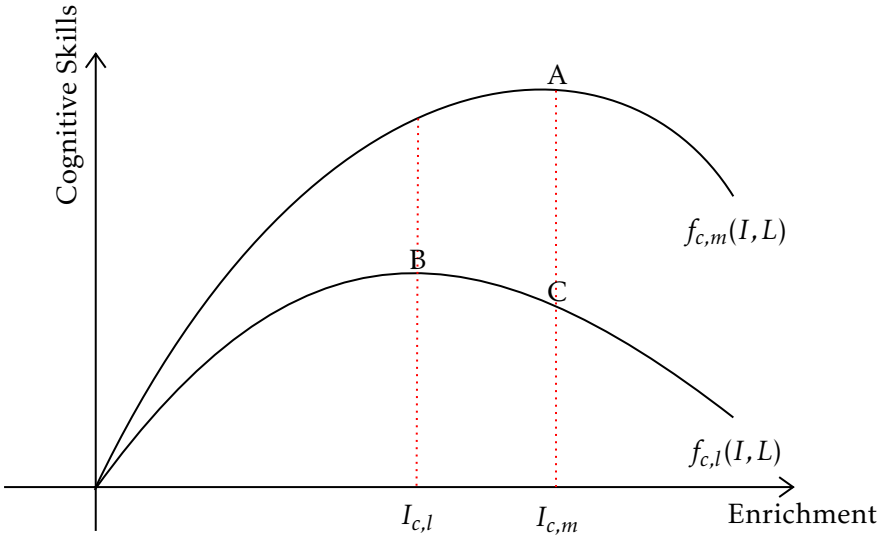
time they spend on enrichment.[5]

These points are illustrated graphically in Figure 11. The figure plots the cognitive production functions for middle-income (m) and low-income (l) families under the assumptions introduced in Section 6.1. Middle-income families are more efficient at producing cognitive skills with enrichment activities, perhaps because of complementarities between income and enrichment activities (e.g., richer parents can afford higher quality tutoring, drivers, shorter commutes to activities etc.), so their optimal level of enrichment ($I_{c,m}$) is larger than the optimal level for low-income families ($I_{c,l}$). The figure makes it clear that simply inducing low-income families to increase their enrichment choice towards $I_{c,m}$, without otherwise altering the productivity differences between the two groups, would only serve to increase the cognitive skills gap (the vertical difference between points A and C is wider than the vertical difference between points A and B). By contrast, a policy that altered the low-income productivity of enrichment activities (by offering help with complementary inputs, thus shifting the low-income curve towards the middle-income curve) could be successful at reducing inequality.

The non-cognitive skills plot, Figure 12, yields broadly similar conclusions, albeit with an important caveat. The figure shows the non-cognitive production functions for middle- and low-income families. Consistent with our empirical estimates, the cognitive-maximizing choice for low-income families ($I_{c,l}$) also maximizes non-cognitive skills approximately – these families are already investing efficiently. By contrast, the cognitive-maximizing choice for middle-income families ($I_{c,m}$) is well past the non-cognitive optimum – lowering the enrichment time of middle-income families towards the non-cognitive optimum would actually increase the non-cognitive gap. Whether the income gap in enrichment time contributes to or dampens the inequality overall depends on the relative importance of non-cognitive versus cognitive skills in generating economic outcomes.

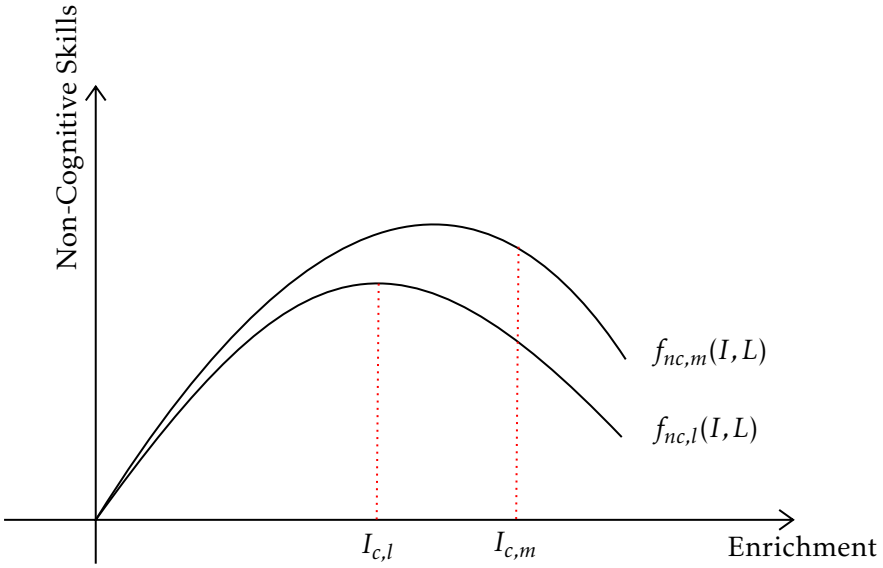
[5]Of course, the skills gap would narrow if higher income families did less enrichment, but this would only lower the gap by directly harming higher income children and would additionally introduce inefficiency.

Figure 11: Cognitive Skills Production by Parental Income



Note: This figure illustrates the cognitive skill production functions for low- and middle-income children. Under the assumption that enrichment is chosen to maximize cognitive skills, middle-income children choose to spend $I_{c,m}$ hours (point A) while low-income children choose to spend $I_{c,l}$ hours (point B). Eliminating the gap in enrichment by simply increasing enrichment of low-income children to $I_{c,m}$ (point C) reduces cognitive skills for low-income children, increasing the gap relative to middle-income children. To reduce this gap, it is necessary to elevate the low-income curve towards the middle-income curve. This would directly lower the skill gap, even holding enrichment time fixed. Moreover, it would lower the gap yet further as low-income children endogenously increase their choice of enrichment closer to $I_{c,m}$.

Figure 12: Non-cognitive Skills Production by Parental Income



Note: This figure illustrates the possible relationship between non-cognitive skill production for middle- ($f_{n,m}$) and low-income ($f_{n,l}$) children and the corresponding cognitive-maximizing choices of I. Assuming I is chosen to maximize cognitive skills, low-income children choose $I_{c,l}$ while middle-income children choose $I_{c,m}$. As drawn, $I_{c,l}$ is approximately the same as $I_{n,l}$, the non-cognitive maximizing level of I, while $I_{n,m}$ is lower than $I_{c,m}$. Decreasing I from $I_{c,m}$ towards $I_{n,m}$ would actually increase the gap in non-cognitive skills. To reduce the gap in non-cognitive skills, it is necessary to elevate the low-income curve towards the middle-income curve, perhaps by affecting complementary inputs.

However, the overall recommendation for non-cognitive skills is the same as for cognitive skills. Simply incentivizing low-income children to spend more time on enrichment without changing the productivity of that activity will not reduce inequality, as these children seem to be choosing optimally already. In contrast, doing so by changing the productivity of these activities (shifting the low-income curve upwards) would yield the desired effect.

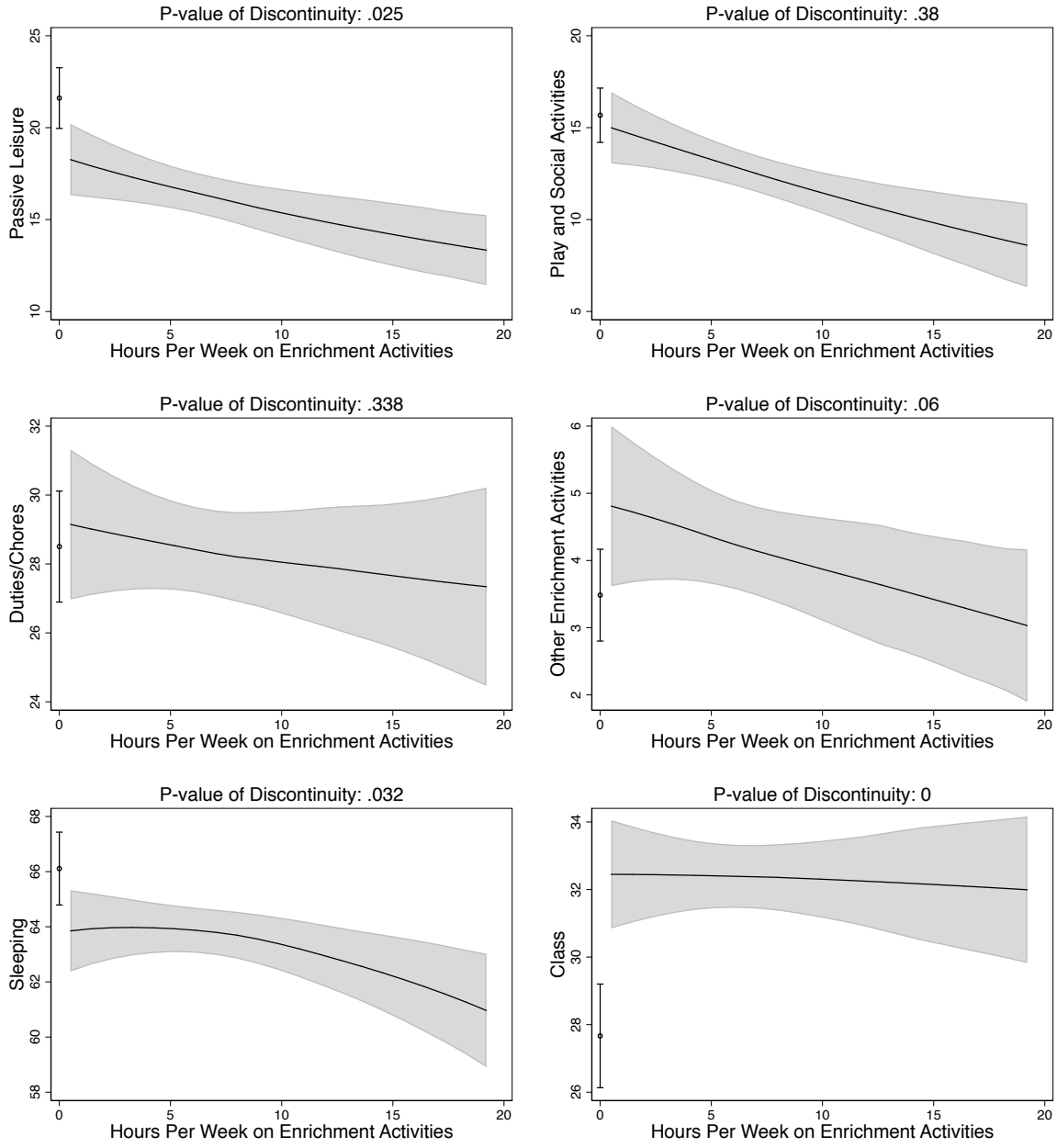
6.3 Which Activities are Crowded Out by Enrichment?

To interpret our findings, it is useful to gain some insight into the types of activities for which enrichment time substitutes. In this section, we focus on middle-income children attending grades 9-12. This is the group with the most pronounced negative effects of enrichment on non-cognitive skills (see Table **??**, p.**??**). Because our causal estimates are *relative* to alternative time uses, the large negative effects for this group implies that enrichment must be substituting other activities which are more productive for non-cognitive skills at the margin.

A detailed analysis that aims at obtaining the exact extent of substitution between enrichment and each alternative activity requires the identification of causal substitution effects and is beyond the scope of this paper. However, Figure 13 provides some suggestive evidence of the patterns of substitution. Each panel plots the average time spent in different activities (class time, sleep, passive leisure, etc.) for each level of enrichment time.[6] The figure suggests that as children spend more time on enrichment activities, they tend to spend less time on passive leisure, play and social activities, duties/chores, other enrichment activities (activities included in the broad enrichment category but not included in the enrichment category) and sleep. The relationship between enrichment and class time is less clear.

[6]See the data Section 2 and Figure 16 in p.49 of Appendix C for a description of each category.

Figure 13: 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.

Of course, these relationships are not necessarily causal, so they may not reflect actual substitution. However, we can use Caetano (2015)'s test of exogeneity (see Section 3 and Appendix A) to gain insight into whether endogeneity is likely to play a major role in

these observed correlations. For instance, the average amount of passive leisure for the 0-enrichment children is starkly different from the average passive leisure of those children who do just one hour or two of enrichment per week. This suggests that this correlation is not necessarily causal. Following this logic, we do not suggest that this negative correlation is evidence of causal substitution. A similar logic allows us to refrain from concluding that there is substitution between enrichment activities and "other enrichment activities" or sleep.

By contrast, the gradient between play/social activities and enrichment is very negative and yet there is no evidence of discontinuity at zero (p=0.38). This suggest that some of this negative correlation might actually reflect substitution.

Extending this logic to all other panels, we conclude that there is no suggestive evidence that children in this group substitute enrichment away from class time, sleep, passive leisure and other enrichment activities (though absence of evidence of substitution is not evidence of the absence of substitution). Conversely, there is some evidence that they do substitute enrichment away from play and social activities, as well as from duties/chores. The gradient between duties/chores and enrichment is less steep than the gradient between play/social activities and enrichment, suggesting that the evidence of enrichment substitution away from play and social activities is the strongest.

7 Conclusion

This paper estimates the total effect of time spent on enrichment activities on cognitive and noncognitive skills. To do this, it proposes a correction for the endogeneity bias stemming from the non-random choice of enrichment time. The correction is straightforward to implement and requires only that the confounding unobservables be symmetrically distributed in the tails. Moreover, the correction permits the distribution of unobservables to vary arbitrarily with the controls. A major advantage of our approach is that it permits

causal inference from cross-sectional data using relatively parsimonious, pre-determined controls.

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 causal effect of enrichment activities is negligible for cognitive skills and negative for non-cognitive skills. The negative effects for non-cognitive skills are particularly stark for middle-income families with children in high school.

We interpret our results through the lens of a stylized model of time allocation and skill production. We argue that if parents choose enrichment time to maximize cognitive skills, we would expect to see zero marginal returns to enrichment for cognitive skills and negative returns for non-cognitive skills. However, parents cannot maximize cognitive skills and non-cognitive skills at the same time, leading them to choose levels of enrichment that may be greater than the optimal level had they intended to maximize non-cognitive skills.

This stylized model may also explain why the effects are particularly negative for middle-income youth 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. This force may be particularly acute for middle-income youth, who may suffer similar pressure to attend college as high-income youth but may have comparatively more limited resources to purchase complementary inputs.

We provide some suggestive evidence that middle-income youth in high school are substituting enrichment away from play and social activities, which may be more productive for non-cognitive skills at the margin. Future research with data on a larger sample of children should aim at identifying causal substitution patterns among activities for different groups of children, as they may be very heterogeneous, particularly along

socioeconomic lines and across grades. Other areas for future research include identifying the effect of "screen time" on skills[7] and using bunching along multiple dimensions to understand and correct for multi-dimensional unobservables.

[7] Our paper Caetano et al. (2019) presents preliminary evidence of the effect of television time on skills using the same PSID/CDS data.

References

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

Avent, R. (2017). High-pressure parenting. *1843 Magazine*.

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, 75(299):493–517.

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 of Surprising Continuity? *Demography*, 37:401–414.

Bonhomme, S., Lamadon, T., and Manresa, E. (2017). Discretizing Uunobserved 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. (2019). A test and a correction of endogeneity bias in linear models with bunching. *Working Paper*.

Caetano, G., Kinsler, J., and Teng, H. (2018). Towards causal estimates of children's time allocation on skill development. Forthcoming at *Journal of Applied Econometrics*.

Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:697–812.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, 1(3):111–34.

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). *Figure 1.6: Enrichment Expenditures on Children, 1972-2006*, chapter 1, page 11. Russell Sage.

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.

Gray, P. (2010). The decline of play and rise in children's mental disorders. *Psychology Today (blog)*.

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. (2019). Evolutionary functions of play: Practice, resilience, innovation, and cooperation. *The Cambridge Handbook of Play: Developmental and Disciplinary Perspectives*.

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.

Khazan, O. (2016). Ending extracurricular privilege. *The Atlantic*.

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.

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.

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 2010.

Rosen, R. (2015). Why affluent parents put so much pressure on their kids. *The Atlantic*.

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.

Walker, M. (2017). *Why we sleep: Unlocking the power of sleep and dreams*. Simon and Schuster.

A The naive approach has a positive bias.

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

Caetano (2015)'s test of exogeneity suits this application perfectly. The test uses 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 $\mathbb{E}[S|I=0,X]$.

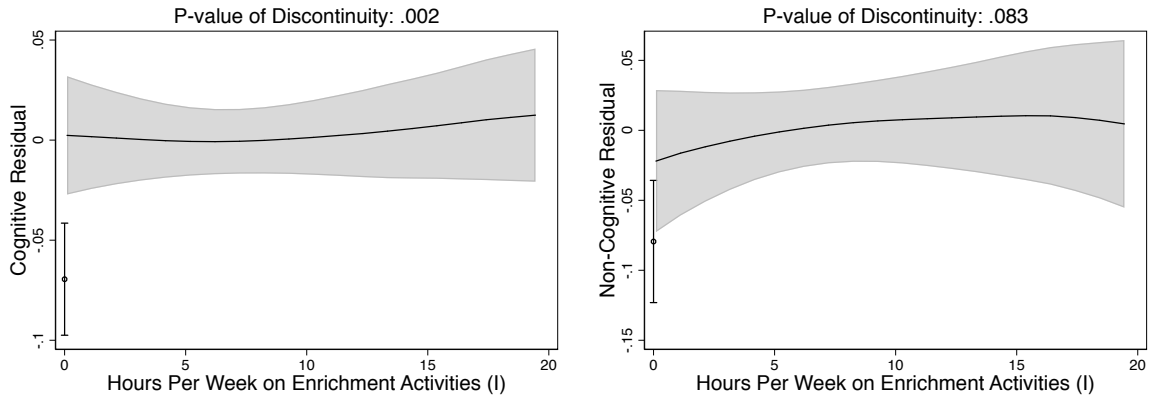
We perform this test using the full list of controls included in our main analysis in Section 5. We find strong evidence that $\mathbb{E}[S|I=0,X]$ is actually discontinuous at $I=0$. Thus, we conclude that I is endogenous and the naive estimator of β must be biased.[8]

In order to display this multivariate result in a two-dimensional plot, Figure 14 shows the residuals of the 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 discontinuously lower at zero enrichment. Thus, we conclude that the OLS estimator of β in equation (1) is biased upward for both types of skill – the unobservables that contribute positively to enrichment also directly contribute positively to both cognitive and non-cognitive skills.

[8] Precisely, this is a joint test of the exogeneity of I and the specification of equation (1). The fact that the test strongly rejects means that either I is endogenous or equation (1) is mispecified. Whatever the reason for the rejection, the estimate of β will be biased.

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 (bandwidth=15 hours). 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.

Note that this conclusion can be drawn without imposing any structure in the model.

B Clustering and the discretization of X

In Section 4, we discuss two strategies for the estimation of $\mathbb{E}[I^*|I^* \leq 0, X]$, heteroskedastic symmetric and heteroskedastic Tobit, which require the distribution of X to have a discrete support. In most applied settings, including ours, there will be some necessary controls with 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 different elements of the original, non-discretized controls.

Our idea is to classify observations with similar observed controls into discrete clusters and use 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 group – two children in the same cluster have by construction more similar controls than two children in different clusters. Note that we use cluster membership indicators in place of X in the construction of $\mathbb{E}[I^*|I^* \leq 0, X]$, but we allow the unclustered,

non-discretized X's to still enter the outcome equation.

Formally, let k_i be the cluster to which child i belongs. The underlining assumption in our method is that $\mathbb{E}[I^*|I^* \leq 0, X = X_i] = \mathbb{E}[I^*|I^* \leq 0, X \in k_i]$. In other words, we require that the heterogeneity in η operates only through the clusters, rather than through the (possibly continuous) unclustered controls. 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 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 for an increasing number of clusters, it increases 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).

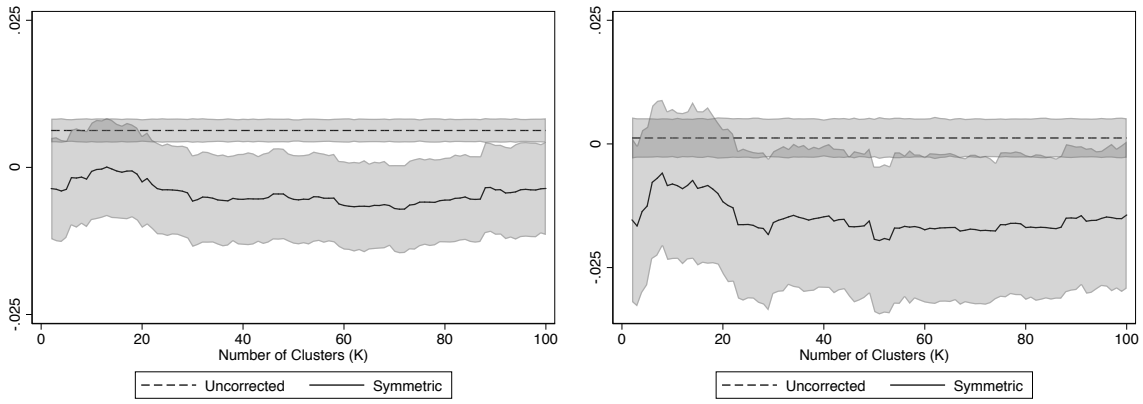
We have experimented with different linkage methods (average, complete, and Ward's) and different dissimilarity measures (Gower, L_1, L_2, and correlation), and the results are uniformly very similar across these different cases. We report results with Ward's linkage method, in which the criterion at 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 for both continuous and binary variables.

Our use of clustering differs from recent applications within economics. A growing literature 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). By contrast, 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.[9]

Figure 15 shows the main results of the paper in the symmetric case for different numbers of clusters K, when clusters are used for the discretization of X to estimate $\mathbb{E}[I^*|I^* \leq 0, X]$. Clearly, after a certain point, the results do not change meaningfully when more clusters are added. This suggests that our results are not an artifact of our discretization of X.

Figure 15: Uncorrected and Corrected Cognitive and Non-cognitive Estimates

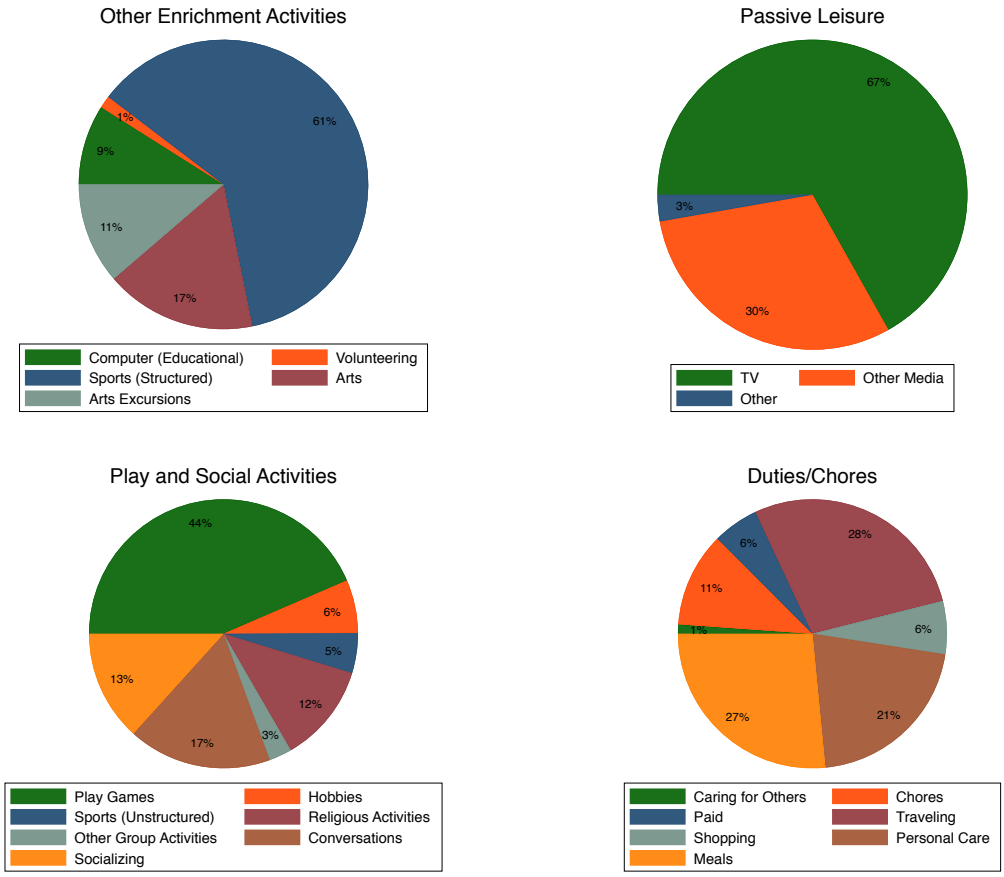


Note: Shaded areas depict 90% confidence intervals for the uncorrected and corrected estimates for the symmetric method. All standard errors bootstrapped using 100 iterations.

[9] In unreported estimates, we also allow cluster indicators to enter directly into the outcome equation as a means of allowing observables to affect skills in a more flexible way. These alternative specifications produce similar estimates as in our preferred specification, although the resulting point estimates are somewhat less precise.

C Additional Materials

Figure 16: Time Breakdowns - Other Time Inputs



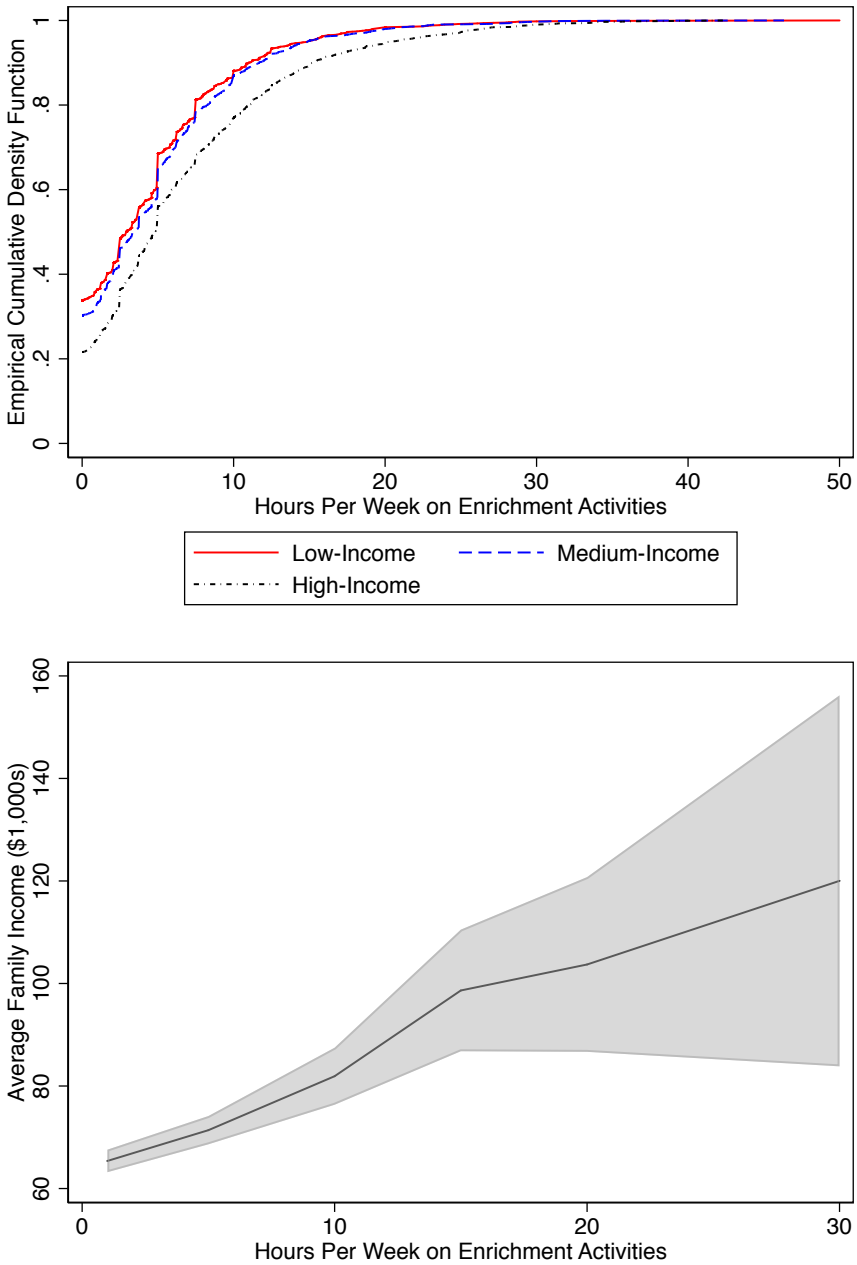
Note: Panels plot the average division of time into different categories over a typical week for our full CDS sample. The figure pools the 1997, 2002 and 2007 CDS waves.

Table 6: 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 Wave sample.

Figure 17: Family Income and Enrichment Time



Note: The top panel shows the empirical cumulative distribution functions of enrichment time for each family income tercile. The bottom panel plots the local linear polynomial estimator of the expected value of family income (in $1,000s) for a given amount of time spent on enrichment per week, along with its 90% confidence interval. The local polynomial is estimated only on non-bunched observations.

Table 7: Uncorrected and Corrected Results – Breaking Down Cognitive and Non-Cognitive Skills

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Cognitive						
Applied	β	0.013***	0.006***	-0.001	-0.002	-0.004
Problems		(0.003)	(0.001)	(0.005)	(0.005)	(0.006)
	δ			0.006	0.007	0.008*
				(0.004)	(0.004)	(0.005)
Letter Word	β	0.012***	0.005***	-0.003	-0.004	-0.006
		(0.003)	(0.001)	(0.006)	(0.005)	(0.005)
	δ			0.007	0.008**	0.009**
				(0.005)	(0.004)	(0.004)
Passage	β	0.014***	0.007***	-0.003	-0.005	-0.006
Comprehension		(0.003)	(0.001)	(0.005)	(0.005)	(0.005)
	δ			0.008*	0.010**	0.010***
				(0.004)	(0.004)	(0.004)
Non-Cognitive						
External	β	0.010***	0.004*	-0.012	-0.019**	-0.016*
		(0.002)	(0.002)	(0.010)	(0.009)	(0.008)
	δ			0.014*	0.019***	0.017**
				(0.008)	(0.007)	(0.007)
Internal	β	0.002	-0.002	-0.019**	-0.020**	-0.016*
		(0.002)	(0.002)	(0.010)	(0.010)	(0.010)
	δ			0.014*	0.014*	0.011
				(0.008)	(0.008)	(0.008)

Note: N=4,330. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications include 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Non-Cognitive Estimates by Grade - Low Income

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
PreK-5	β	0.003	0.011	0.016	0.013	0.013
		(0.009)	(0.009)	(0.037)	(0.043)	(0.044)
N=465	δ			-0.004	-0.002	-0.002
				(0.031)	(0.037)	(0.037)
6-8	β	0.006	0.007	0.073*	0.035	0.033
		(0.009)	(0.010)	(0.038)	(0.040)	(0.037)
N=501	δ			-0.053*	-0.022	-0.021
				(0.029)	(0.033)	(0.030)
9-12	β	0.020***	0.016**	-0.034	0.033	0.034
		(0.007)	(0.008)	(0.031)	(0.033)	(0.036)
N=468	δ			0.037	-0.014	-0.015
				(0.024)	(0.026)	(0.030)

Note: Number of observations (N) for each grade level is shown. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications include 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Non-Cognitive Estimates by Grade - High Income

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
PreK-5	β	0.002	-0.003	0.085**	0.064**	0.036*
		(0.006)	(0.007)	(0.039)	(0.031)	(0.020)
N=403	δ			-0.083**	-0.062**	-0.034**
				(0.036)	(0.028)	(0.017)
6-8	β	-0.006	-0.009*	-0.019	-0.025	-0.024
		(0.005)	(0.005)	(0.029)	(0.024)	(0.018)
N=451	δ			0.009	0.014	0.013
				(0.026)	(0.021)	(0.015)
9-12	β	0.002	0.001	-0.018	-0.023	-0.023
		(0.005)	(0.005)	(0.018)	(0.017)	(0.016)
N=580	δ			0.015	0.019	0.019
				(0.014)	(0.013)	(0.012)

Note: Number of observations (N) for each grade level is shown. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications include 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Cognitive Estimates by Grade - Low Income

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
PreK-5	β	-0.006	-0.001	-0.018	-0.007	-0.010
		(0.007)	(0.004)	(0.020)	(0.020)	(0.019)
N=465	δ			0.015	0.005	0.008
				(0.016)	(0.017)	(0.016)
6-8	β	0.028***	0.021***	0.005	0.003	-0.004
		(0.007)	(0.005)	(0.020)	(0.023)	(0.023)
N=501	δ			0.013	0.014	0.020
				(0.016)	(0.020)	(0.020)
9-12	β	0.024***	0.011**	-0.023	-0.001	0.001
		(0.006)	(0.005)	(0.018)	(0.019)	(0.020)
N=468	δ			0.025**	0.008	0.007
				(0.013)	(0.015)	(0.016)

Note: Number of observations (N) for each grade level is shown. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications include 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Cognitive Estimates by Grade - Middle Income

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
PreK-5	β	0.000	0.001	0.001	0.003	-0.003
		(0.008)	(0.005)	(0.022)	(0.022)	(0.018)
N=449	δ			0.000	-0.003	0.003
				(0.019)	(0.018)	(0.014)
6-8	β	0.007	-0.001	-0.001	-0.001	-0.011
		(0.005)	(0.004)	(0.018)	(0.017)	(0.017)
N=456	δ			0.000	0.000	0.008
				(0.016)	(0.015)	(0.014)
9-12	β	0.027***	0.010**	-0.001	-0.010	-0.015
		(0.005)	(0.005)	(0.012)	(0.014)	(0.016)
N=529	δ			0.008	0.014	0.018
				(0.008)	(0.010)	(0.012)

Note: Number of observations (N) for each grade level is shown. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications include 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Cognitive Estimates by Grade - High Income

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
PreK-5	β	0.022***	-0.003	0.018	0.014	0.005
		(0.008)	(0.004)	(0.027)	(0.020)	(0.013)
N=403	δ			-0.020	-0.017	-0.008
				(0.025)	(0.018)	(0.010)
6-8	β	0.009**	0.006*	-0.012	-0.022	-0.015
		(0.005)	(0.003)	(0.018)	(0.015)	(0.012)
N=451	δ			0.016	0.024*	0.018*
				(0.016)	(0.014)	(0.010)
9-12	β	0.015***	0.006**	0.001	0.002	0.006
		(0.003)	(0.003)	(0.010)	(0.009)	(0.009)
N=580	δ			0.004	0.003	0.000
				(0.008)	(0.007)	(0.007)

Note: Number of observations (N) for each grade level is shown. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications include 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.

Table 13: 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***	0.007***	-0.002	-0.001	0.000
		(0.002)	(0.001)	(0.006)	(0.005)	(0.004)
	δ			0.008	0.007	0.006*
				(0.006)	(0.005)	(0.003)
Non-Cognitive	β	0.009***	0.006***	-0.015	-0.012	-0.007
		(0.002)	(0.002)	(0.010)	(0.008)	(0.006)
	δ			0.019**	0.015**	0.011**
				(0.009)	(0.007)	(0.005)

Note: N=4,330. Bootstrapped standard errors in parentheses (1,000 iterations). The corrected specifications include 50 clusters. *** p<0.01, ** p<0.05, * p<0.1.