Economics of Education Review 98 (2024) 102503

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Are children spending too much time on enrichment activities?☆

Carolina Caetano [a], Gregorio Caetano [a,*], Eric Nielsen [b]

[a] University of Georgia, United States of America
[b] Federal Reserve Board, United States of America

ARTICLE INFO

JEL classification:
I21
I2
J01
C24

Keywords:
Enrichment
Cognitive
Non-cognitive
Bunching
Homework

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 the opportunity cost of spending time on enrichment, as it may replace activities such as sleep and socializing. Our study controls for selection on unobservables using a control function approach that leverages the fact that many children spend zero hours per week on enrichment activities. At zero enrichment, confounders vary but enrichment does not, giving us direct information about the effect of confounders on skills. Using time diary data available in the Panel Study of Income Dynamics (PSID), we find that the net effect of the last hour of enrichment is close to zero for cognitive skills and negative for non-cognitive skills. The negative effects for non-cognitive skills are concentrated in high school, consistent with elevated academic pressure related to college admissions.

1. Introduction

Families spend substantial resources on activities intended to increase their children's skills. These "enrichment" activities include homework, tutoring, reading, and extracurricular activities such as music and art lessons. The money and time committed to these activities are substantial, and the gap in these investments by socioeconomic status has become more intense in the past decades, possibly associated with the increasing competition for college admission (Bound, Hershbein, & Long, 2009; Ramey & Ramey, 2010). Many of these enrichment activities are directly aimed at increasing skills, leading to concerns that they may contribute to cross-sectional and intergenerational inequality (Aguiar & Hurst, 2007; Bianchi, 2000; Doepke & Zilibotti, 2017, 2019; Duncan & Murnane, 2011; Guryan, Hurst, & Kearney, 2008; Rønning, 2011).

However, 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 socializing and sleeping. Moreover, time spent on enrichment could have spillover effects into the remainder of the day. For example, a teenager who is overstimulated by an after-school activity may go to bed later than usual. Yet sleep is an activity with direct, positive impact on skills (Eide & Showalter, 2012; Groen & Pabilonia, 2019; Heissel & Norris, 2018; Lenard, Morrill, & Westall, 2020; Wolfson et al., 2003). The opportunity costs of enrichment activities might therefore be substantial depending on the activities replaced.

Indeed, the potential perils of spending too much time on enrichment activities is well documented in ethnographic studies in the child development literature (e.g. Galloway, Conner, & Pope, 2013; Ginsburg et al., 2007; Gray, 2011; Jarvis, Newman, & Swiniarski, 2014; Luthar, 2003; Luthar & Becker, 2002; Veiga, Neto, & Rieffe, 2016; Villaire, 2003), has been the subject of many books (e.g. Abeles, 2015; Anderegg, 2003; Gray, 2013; Lareau, 2003; Lukianoff & Haidt, 2018; Rosenfeld & Wise, 2000; Warner, 2005), and has been widely covered in the popular press (e.g. Avent, 2017; Gray, 2010; Khazan, 2016; Rosen, 2015; Rosin, 2015). In general, there is an acknowledgment in these writings that an additional hour of enrichment must be replacing some other activity, and that the net effect of enrichment may depend on the specific displaced activity. However, there is no systematic discussion about how these substitution patterns may change depending on the age of the child, among other dimensions. Moreover, each of these works tends to focus on a specific aspect of the issue. For instance,

☆ We thank Marinho Bertanha, Hao Fe, Bruno Ferman, Leonard Goff, Peter Hinrichs, Josh Kinsler, Matthew Larsen, Salvador Navarro, Rodrigo Pinto, Vitor Possebom, Pedro Sant'Anna, Meghan Skira, David Slichter, Christopher Taber and seminar participants at various conferences and institutions for helpful discussions and comments. Michael Ahn and Hannah Hall provided excellent research assistance. 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.

* Corresponding author.
E-mail addresses: carol.caetano@uga.edu (C. Caetano), gregorio.caetano@uga.edu (G. Caetano), Eric.R.Nielsen@frb.gov (E. Nielsen).

https://doi.org/10.1016/j.econedurev.2023.102503
Received 16 November 2022; Received in revised form 5 December 2023; Accepted 7 December 2023
Available online 4 January 2024

while some studies have discussed the importance of play and socialization (e.g., Gray, 2011), other studies have focused on specific populations, such as high school students in some college preparatory schools, where they show that many children have difficulty coping with the pressure of performing well academically (e.g., Galloway et al., 2013).

A related literature studies the impact of homework, which is typically the single largest component of enrichment in our data, on cognitive skills. As surveyed in Cooper, Robinson, and Patall (2006), early studies in this literature tend to find large positive effects but also rely mostly on correlational evidence. More recent studies that attempt to control for confounders find smaller effects that are indistinguishable from zero in most cases (e.g., Eren & Henderson, 2011). Notably, these studies do not estimate effects on non-cognitive skills, even though this is where most of the opportunity costs of enrichment seem to be concentrated according to the ethnographic studies discussed above.

To our knowledge, there are no prior studies estimating the effect on both cognitive and non-cognitive skills of spending more time on enrichment activities which account for the opportunity costs due to substitutions across activities and which also control for confounders. To be sure, there is an established literature studying the impact of different uses of time on the skills of children, especially in economics, and many of these papers systematically consider confounders (e.g., Caetano, Kinsler, & Teng, 2019; Del Boca, Monfardini, & Nicoletti, 2017; Fiorini & Keane, 2014; Hsin & Felfe, 2014; Jürges & Khanam, 2021 and Todd & Wolpin, 2007). However, as we discuss in detail in Section 2, these studies focus on the identification of parameters of the skill production function, which are quantities that do not speak directly to the debate over whether children are spending too much time on enrichment activities because they do not incorporate the consequences of substitution across activities.

This paper aims to contribute to this debate by using detailed time diary data to estimate the net average effect of enrichment activities on cognitive and non-cognitive skills, taking into account the effects of the substituted activities on skills. We allow for heterogeneity in the effects of enrichment depending on three developmental stages of the child: pre-kindergarten to 5th grade, 6th to 8th grade, and 9th grade to 12th grade.

In order to identify this net effect, we need to control for confounders — unobserved factors that influence both skills and the choice of enrichment. The economics literature identifying skill production functions has mostly handled confounders by controlling for a very detailed list of observables, including lagged skill measures. However, as we show in Section 2, this approach unfortunately cannot be used in our context, since our need to identify a different quantity limits the types of observables we can include as controls.

We thus adopt an alternative approach to control for confounders that exploits the fact that children in our data bunch at zero hours of enrichment activities per week (see Fig. 1). We argue that many of these children are at a corner solution: time spent on an activity cannot be negative, so children with low propensities to spend time on enrichment all choose the lowest feasible amount of enrichment, zero.[1] This yields useful variation in confounders: at zero enrichment, all children chose the same amount of the "treatment", namely zero. However, these children have different propensities towards enrichment: while some of them are nearly indifferent between their choice of zero enrichment and some other activity, others are far from indifferent; even a large shift in the costs/benefits of enrichment time would not induce them to move away from zero hours. Thus, at zero enrichment, confounders vary but treatment does not, so the variation in the outcome among observations at zero enrichment gives us direct information about the effect of confounders on skills. We use this idea to create a control function approach that corrects for the effect of confounders on skills. This approach is developed formally in Caetano, Caetano, and Nielsen (2023) and has been also used in other contexts, such as the study of the effect of maternal labor supply on the skills of the child (Caetano, Caetano, Nielsen, & Sanfelice, 2021). A related approach has also been used to test a selection-on-observables assumption in the time use literature discussed earlier, when other time inputs are included as controls (e.g. Caetano et al., 2019 and Jürges & Khanam, 2021).

Using time diary data from the Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID), we find that the net effect of enrichment on cognitive skills is small and indistinguishable from zero and that the net effect of enrichment on non-cognitive skills is quite negative and significant. This negative effect on non-cognitive skills is concentrated in high school, which is when enrichment activities become more oriented around homework and less oriented around social activities. Our measure of non-cognitive skills combines both externalizing behaviors related to outward aggression/antisocial behavior and internalizing behaviors related to anxiety/depression/self esteem. Despite the apparent differences in these behaviors, we nonetheless find similar negative effects considering internalizing and externalizing behaviors separately. We likewise find null effects for each of the three constituent achievement tests that combine to form our cognitive skills measure.

Our results are robust along many dimensions. We find very similar results using a broader definition of enrichment time, as well as using alternative constructions of cognitive and non-cognitive skills. We also conduct a systematic sensitivity analysis, where we show that our main empirical conclusions do not change when we consider violations of the main identifying assumptions of the paper.

Our findings highlight the pitfalls and trade-offs associated with intensive investment in enrichment activities, especially around high school, when enrichment activities become more oriented towards academic activities. Many youth seem to be spending so much time on enrichment that, on average, their last hour on these activities is actively harming their non-cognitive skills with no offsetting gain to their cognitive skills. In Section 8, we present a stylized model of optimal time allocation that can rationalize these results. In the model, enrichment time is chosen to maximize cognitive skills. The resulting first-order condition yields zero net effects for cognitive skills. Moreover, because enrichment activities have higher returns for cognitive relative to non-cognitive skills, the first-order condition simultaneously yields negative effects for non-cognitive skills. Although this simple model offers an explanation for the results, it is important to note that further research is needed to understand whether this model is a good approximation to the actual motivation of parents and children.

In addition to these substantive empirical results, our paper highlights the potential for using bunching in the treatment to correct for selection on unobservables. Bunching of a treatment variable at one extreme due to a constraint is very common in many applied settings, including settings of interest to education and child development researchers. For example, many other activities measured in terms of time use display bunching, and indeed we have examined in other work the effects of maternal working hours and television time on childhood skills (Caetano et al., 2023, 2021). The method could be applied to study the effects on childhood development of many other types of activities including social media usage, homework, active time with parents, time with friends, playing sports, etc. (Caetano et al., 2019; Jürges & Khanam, 2021). More broadly, consumption variables often display bunching, so for instance the effect on fetal and childhood development of maternal consumption of goods such as cigarettes (Almond, Chay, & Lee, 2005; Caetano, 2015; Oken, Levitan, & Gillman, 2008) and vitamins (Fawzi, Chalmers, Herrera, & Mosteller, 1993) could also be studied with the method. Finally, it is possible to study the effect of a school resource on student outcomes, as long as that school resource

[1] For brevity and simplicity, we will write throughout the paper that children "choose" enrichment. However, our analysis is agnostic about who (the child, the parent, or some combination) is actually making this choice.

is bunched at the extreme of the distribution. For instance, the distribution of specific school resources across schools might be bunched at zero, such as the amount of dollars allocated to a given subcategory of school spending (e.g. funding for school police as in Weisburst, 2019), or the proportion of students attending the school who are eligible for subsidized school meals (Millimet & Tchernis, 2013).

The rest of the paper is organized as follows. Section 2 discusses the need to depart from the prior literature in economics in order to inform the debate about whether children are spending too much time on enrichment activities. Section 3 presents the data, and Section 4 presents our identification strategy. Our results are shown in Section 5, and in Section 6 we test our identifying assumptions. Section 7 presents a sensitivity analysis. We interpret our key empirical results through the lens of a stylized model in Section 8 before we conclude in Section 9.

2. Motivating our departure from the prior literature

This section discusses our reasons for departing from the usual identification approaches in the economics literature that estimates the effects of time use. First, we discuss why the object of interest in our context is different from the one in these prior studies. Next, we argue that the methods to control for confounders that proved to be effective in these studies unfortunately do not work in our context. This motivates the need for our new approach, which we detail in Section 4.

2.1. Partial vs. Total effects of time inputs

Let S denote the child's skill, and suppose that it is determined by inputs I, O, and R through the production function f:

$$S = f(I, O, R), \tag{1}$$

where I is enrichment time, O is a vector of all other time uses, which together with I add up to 24 h per day, and R is a vector denoting any remaining inputs in the production function, such as market-purchased goods, parental engagement, developmental age, environmental characteristics and ability. R may also include the full history of prior skills and investments, as well as initial endowments, consistent with a dynamic process of skill formation, as in Cunha, Heckman, and Schennach (2010). Some or all elements of O and R may not be observed.

The partial effect of enrichment time on skills for a child with inputs $(I, O, R) = (i, o, r)$ is $\partial f(i, o, r)/\partial I$. This effect is often not empirically relevant on its own, however, as it is impossible to change I without changing at least one element of O — every child has 24 h per day. Moreover, R might also change with I. For instance, parental engagement might increase as the child replaces one hour of watching TV for one hour of homework.[2]

The total effect of I on S, which takes into account changes (O, R) as I changes, is therefore more empirically relevant. This effect is given by[3]

$$\frac{df(i,o,r)}{dI} = \frac{\partial f(i,o,r)}{\partial I} + \sum_j \frac{\partial f(i,o,r)}{\partial O_j}\frac{\partial O_j(i,o,r)}{\partial I} + \sum_l \frac{\partial f(i,o,r)}{\partial R_l}\frac{\partial R_l(i,o,r)}{\partial I} + \sum_l \sum_j \frac{\partial f(i,o,r)}{\partial R_l}\frac{\partial R_l(i,o,r)}{\partial O_j}\frac{\partial O_j(i,o,r)}{\partial I}. \tag{2}$$

For the jth element of O, $\partial O_j(i,o,r)/\partial I$ denotes the substitution/complementarity between enrichment I and the time input O_j. This term will be negative if I substitutes for O_j, and it is necessarily negative on average across all time uses, but it can also be positive for some particular j if I and O_j are complements (e.g. commuting to and attending a class). Analogously, $\partial R_l(i,o,r)/\partial I$ denotes the substitution/complementarity between enrichment and input R_l. In general, every term in Eq. (2) might depend on the values of (i,o,r), so $df(i,o,r)/dI$ will generally be heterogeneous across children.

A typical approach in the economics literature is to specify the production function (1) linearly with one time use in O excluded to avoid multicollinearity.[4] Under a suitable exogeneity assumption, one can identify a weighted average of the difference between the direct effect of enrichment and the direct effect of the excluded time input, O_1, i.e. $\partial f(i,o,r)/\partial I - \partial f(i,o,r)/\partial O_1$. This quantity is valuable, as it gives the marginal skill effect of substituting between I and the omitted category O_1, holding fixed other time uses and non-time skill inputs. However, it does not speak directly to the question of whether children are spending too much time on enrichment activities. As shown by Eq. (2), the answer to this question depends on many additional substitution terms that are not identified by this approach.[5]

To make progress on this debate, we pursue the identification of a weighted average of the total effect of a marginal increase in I on S, $df(i,o,r)/dI$. Our approach is necessarily a compromise, as we will not be able to separately identify each of the components inside Eq. (2). We also allow for as much heterogeneity in our estimates as feasible, although data restrictions limit our contribution in this direction. Specifically, motivated by the child development studies discussed in the introduction, we allow for unrestricted heterogeneity by type of skill (cognitive and non-cognitive).[6] We also allow for unrestricted heterogeneity across different grade ranges (PreK–5th grade, 6th–8th grade, 9th–12th grade) in order to capture heterogeneity across different developmental stages in the production function ($\partial f(i,o,r)/\partial I$, $\partial f(i,o,r)/\partial O_j$ and $\partial f(i,o,r)/\partial R_l$) as well as resources, restrictions, and household objectives that lead to different substitutions and complementarities ($\partial O_j(i,o,r)/\partial I$, $\partial R_l(i,o,r)/\partial I$ and $\partial R_l(i,o,r)/\partial O_j$).

[2] Although not necessary, an optimization framework can help build intuition about the relationship between I, O and R. For instance, following the choice framework from Becker (1965), assume households allocate time and goods in order to create other goods (such as children's skills), which in turn enter the utility function they seek to maximize. This optimization is constrained by both natural restrictions (e.g., each person in the household has 24 h a day, and no one can spend less than 0 min on any activity) and other restrictions which may relate to the household's resources. In general, the chosen elements of O and R in this optimization problem are, through the first order conditions, functions of the chosen I. Section 8 presents a stylized model in which I is chosen to maximize cognitive skills.

[3] For simplicity of exposition, this equation implicitly assumes that O_j does not affect any other element of O, and that R_l does not affect any other element of R. Moreover, R is not allowed to cause O, but the converse (O causing R) is allowed to happen. These assumptions do not change the take-away from this discussion; Eq. (2) would simply have additional terms.

[4] For examples, see Caetano et al. (2019), Del Boca et al. (2017), Fiorini and Keane (2014) and Jürges and Khanam (2021). Some studies explicitly omit a specific category, such as "sleep", while others implicitly omit a category reflecting the remaining time not accounted for by the other included time inputs.

[5] Generalizations of this "omitted time use" approach are still not sufficient. If O is observed, under further exogeneity assumptions, one can also identify a weighted average of $\partial f(i,o,r)/\partial O_j - \partial f(i,o,r)/\partial O_1$ for all $j > 1$, which allows one to uncover a weighted average of $\partial f(i,o,r)/\partial I - \partial f(i,o,r)/\partial O_j$ for any time use O_j. However, even with all of these average differences identified, we would still not identify the average of $\partial f(i,o,r)/\partial R_l$, nor would we know the patterns of substitution between all the time inputs and all other inputs in the production function, i.e. $\partial O_j/\partial I$, $\partial R_l/\partial I$ and $\partial R_l/\partial O_j$, for all j and l.

[6] We also consider heterogeneity within each type of skill (e.g., math vs. verbal cognitive skills, and internal vs. external non-cognitive skills).

To see how these two dimensions of heterogeneity – child age/grade and the type of skill – help us understand some of the relevant trade-offs involved, consider a hypothetical example comparing the total skill effects of an additional hour of enrichment for two otherwise-similar youth, a 6-year-old and a 16-year-old. Even in the (perhaps unrealistic) case that the partial effects $\partial f(i,o,r)/\partial I$ are equal for both children, the total effects $(df(i,o,r)/dI)$ might nonetheless be quite different. This could arise if the additional hour of enrichment replaces different activities for the two children, that is, if $\partial O_j(i,o,r)/\partial I$ differs, and if these alternative time uses have different partial effects $\partial f(i,o,r)/\partial O_j$. For example, enrichment might replace TV for the younger child and socializing for the older child. Additionally, the extra hour of enrichment might occur with a larger increase in parental engagement for the 6-year-old than for the 16-year-old (which refers to $\partial R_l(i,o,r)/\partial I$ and $\partial R_l(i,o,r)/\partial O_j$). Furthermore, parental engagement may be productive for both cognitive and non-cognitive skills for the 6-year-old, but it may only be productive for non-cognitive skills for the 16-year-old (which refers to $\partial f(i,o,r)/\partial R_l$). As this example illustrates, the scope for heterogeneity of $df(i,o,r)/dI$ across different age ranges and across different types of skills might be large, thus motivating our specification of heterogeneity along these two dimensions.

Turning to our empirical specification, we start with a simple equation relating skills to enrichment:

$$S = \beta I + U, \tag{3}$$

where β represents a weighted average of the total effect on skill across all children.[7] The error U is simply defined as $f(\cdot) - \beta I$ (where f is defined in Eq. (1)), so no assumption has been made up to this point. Section 4.2 will impose restrictions on U that are necessary for the identification of β.

Note that, differently from Eq. (1), Eq. (3) does not specify a production function. Rather, β represents the average of individual heterogeneous causal effects of I on S that are mediated by the true unknown production function $f(\cdot)$ and the associated substitution/complementarity patterns, as described in Eq. (2).[8] It can therefore be interpreted as the average quantity we would identify if we were to experimentally assign different people to different levels of enrichment I, since in that case, all other time uses, as well as other inputs of the production function, would be endogenous to I.[9]

2.2. Controlling for confounders

A key obstacle to the identification of β is that I and U are correlated. Prior research estimating time use parameters of the skill production function has made progress at solving similar endogeneity problems by assuming that I and U are uncorrelated conditional on an extensive list of controls.[10] This literature has found that two types of control variables are particularly important: *other time inputs* and *lagged skills*.

Unlike these papers, we cannot use these control variables in our setting, reducing our ability to account for endogeneity through observables. As explained in the previous section, using other time inputs would shut off parts of the effect of interest, as it would hold constant O_j for some j. Analogously, some elements of R are likely to be jointly- or post-determined relative to I (e.g., parental engagement) and thus should not be included as controls. Moreover, because the gap between successive survey waves in our data is 5 years, controlling for lagged skills would restrict the age range of our sample substantially, making it infeasible to study whether children in early grades are spending too much time on enrichment activities.[11] Moreover, controlling for lagged skills might not handle all of the confounders we are concerned about. In particular, some confounders that are not predictable from lagged skills might emerge in the 5 years after lagged skills are observed. There is a clear trade-off: as we control for lagged skills, which are helpful to deal with the endogeneity problem, we lose our ability to study our question for an important developmental stage. Other popular approaches exploiting the longitudinal aspect of the data (e.g., controlling for lagged inputs, controlling for within-child fixed effects) would lead to similar trade-offs (e.g., Fiorini & Keane, 2014; Todd & Wolpin, 2003). Losing the ability to study young children aged 5–9 is a steep price to pay in our context, particularly since we have an approach to deal with the confounders that lagged skills are most likely to help control for (see Section 4, in particular Remark 4.1).

3. Data

We use data from the Panel Study of Income Dynamics (PSID), making particular use of the 1997, 2002 and 2007 waves of the Child Development Supplement (CDS). The CDS data contain detailed time diary data as well as extensive measures of cognitive and non-cognitive skills. We link the CDS with the main PSID panel which allows us to build controls related to child, family and environmental characteristics.

Activities: The time diaries in each CDS wave collect data on the full 24-h breakdown of one random weekday and one random weekend day for each child. The child's activities during the selected days are coded into one of over 300 different categories reported by the child, or by the parent if the child is young, with subsequent editing and help from the PSID interviewer. We exclude cases where the day is described as non-typical, either the weekday or weekend day data is missing, or where the diary does not cover the full 24 h. 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 (Caetano et al., 2019; Fiorini & Keane, 2014). Finally, we aggregate the 300-plus primitive time-use categories into eight categories: enrichment activities, other enrichment activities, play and social activities, passive leisure, duties/chores, class time, sleep, and other.

Our definition of enrichment intends to capture the kinds of activities that are typically considered to be investments in children's skills. Therefore, our baseline measure includes only those activities that are unambiguously related to skill development over and above class time in school. In a typical week, children on average spend about 45 min per day on this type of enrichment. This is a nontrivial amount of time, especially considering that, after accounting for the time spent

[7] As discussed above, in practice we allow β to vary by grade range and type of skill. In an earlier version of this paper, we allowed for heterogeneous results for each skill by both grade range and family income terciles, but the estimates were not sufficiently precise to infer a clear pattern. These estimates are available upon request.

[8] Eq. (3) does not assume that f is linear in I; rather, it identifies the best linear approximation of the (potentially nonlinear) function f (see Theorem 3.1.6 in Angrist & Pischke, 2008). Therefore, β is the best approximation of the total effect of an increase of I of one unit incorporating both direct and indirect effects. A useful alternative interpretation can be obtained from Proposition 2 in Yitzhaki (1996): β identifies a weighted average of the df/dI with positive weights averaging one. Precisely, Proposition 2 in that paper implies that $\beta = \int \omega(i) \frac{d}{di} \mathbb{E}[f(I,O,R)|I=i]di$, where $\omega(i) > 0$, and $\int \omega(i)di = 1$. If df/dI is bounded, the Dominated Convergence Theorem then allows us to write $\beta = \int \omega(i)\mathbb{E}\left[df(i,o,r)/dI \middle| I=i\right] di$.

[9] This "experiment" may lead one to consider the possibility of the existence of an instrumental variable (IV) for I that would identify this average effect. However, an IV is very difficult to find in this setting, as it needs to affect S only directly through enrichment (I), and not directly through other activities (O_j) and other inputs (R_l), such as family resources.

[10] See, e.g., Caetano et al. (2019), Fiorini and Keane (2014) and Todd and Wolpin (2007).

[11] The sample of observations where lagged skills are observed is also selected beyond the age/grade of the child. See Footnote 21.

Table 1
Summary statistics.

Activities (hours per week)	All	Grades K-5	Grades 6-8	Grades 9-12
Enrichment	5.22 (6.00)	5.43 (5.35)	5.37 (5.76)	4.93 (6.68)
Other enrichment	4.03 (6.21)	3.01 (4.69)	4.34 (6.10)	4.62 (7.23)
Active leisure	12.30 (10.19)	12.54 (9.23)	10.80 (8.70)	13.42 (11.88)
Passive leisure	17.48 (11.94)	15.53 (9.89)	18.30 (11.57)	18.40 (13.55)
Duties/Chores	24.68 (11.29)	22.83 (8.45)	22.55 (9.19)	28.14 (13.94)
Class	30.96 (10.78)	31.17 (9.57)	32.08 (9.49)	29.77 (12.56)
Sleep	67.20 (9.19)	70.53 (7.09)	67.32 (8.06)	64.30 (10.62)
Other	6.12 (9.87)	6.96 (10.03)	7.24 (10.00)	4.42 (9.38)
Control variables				
Enrichment = 0	0.29 (0.45)	0.19 (0.39)	0.26 (0.44)	0.39 (0.49)
Broad enrichment = 0	0.15 (0.36)	0.10 (0.30)	0.11 (0.32)	0.23 (0.42)
Child is male	0.50 (0.50)	0.51 (0.50)	0.48 (0.50)	0.50 (0.50)
Child is white	0.48 (0.50)	0.50 (0.50)	0.47 (0.50)	0.46 (0.50)
Child is black	0.40 (0.49)	0.37 (0.48)	0.41 (0.49)	0.41 (0.49)
Child is hispanic	0.07 (0.26)	0.08 (0.26)	0.07 (0.26)	0.08 (0.26)
Child has siblings	0.88 (0.33)	0.85 (0.36)	0.90 (0.30)	0.89 (0.31)
Child is in grade PreK–5	0.31 (0.46)	1.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Child is in grade 6–8	0.33 (0.47)	0.00 (0.00)	1.00 (0.00)	0.00 (0.00)
Child is in grade 9–12	0.37 (0.48)	0.00 (0.00)	0.00 (0.00)	1.00 (0.00)
1997 wave	0.26 (0.44)	0.45 (0.50)	0.34 (0.48)	0.03 (0.16)
2002 wave	0.46 (0.50)	0.55 (0.50)	0.34 (0.47)	0.49 (0.50)
2007 wave	0.28 (0.45)	0.00 (0.00)	0.32 (0.47)	0.48 (0.50)
Child's father is alive	0.97 (0.16)	0.98 (0.14)	0.97 (0.16)	0.97 (0.18)
Child's mother is alive	0.99 (0.08)	1.00 (0.06)	0.99 (0.08)	0.99 (0.11)
Child is in Gifted Prog.	0.26 (0.44)	0.14 (0.35)	0.27 (0.44)	0.35 (0.48)
Child is in Spec. Educ. Prog.	0.08 (0.27)	0.07 (0.26)	0.11 (0.31)	0.06 (0.25)
Child is home schooled	0.01 (0.11)	0.01 (0.10)	0.01 (0.12)	0.01 (0.11)
Child is in private school	0.08 (0.27)	0.09 (0.29)	0.08 (0.26)	0.07 (0.25)
Household income (in $1000s)	73.27 (84.05)	67.97 (70.34)	70.52 (77.73)	80.14 (98.37)
Age (years)	11.86 (3.32)	7.92 (1.23)	11.54 (0.86)	15.46 (1.46)
Observations	4330	1331	1414	1585

Note: Activity categories are exhaustive. All control variables are indicators, with the exception of the last two, where the units are mentioned in parentheses. The 1997, 2002 and 2007 CDS waves are pooled. Standard deviations shown in parentheses.

on sleep, school, and duties/chores, children have on average less than 7 h remaining per day (see Table 1).

The primary component of enrichment is homework, at two-thirds of the total.[12] The next most important component of enrichment is reading a book, at 14% of the total. While 7% of enrichment time is spent on before- or after-school programs, relatively little is spent on each of the remaining categories: other reading (e.g., magazines and newspapers), being read to (e.g., by parents), other academic lessons (e.g., tutoring, academic courses and lectures), non-academic lessons (e.g., piano and soccer lessons), and other education (e.g., driving lessons, military training). Table 5 in Section 8 shows the average breakdown of enrichment activities into these sub-categories.

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 baseline notion of enrichment plus "other enrichment:" making art/music, visiting museums, organized (structured) sports, volunteer work, the educational use of computers, and so forth. Table B.6 in Appendix B presents the breakdown of "other enrichment" into its constituent pieces, demonstrating that on average about two-thirds of the category is organized sports.

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

Skills: Following the literature (e.g., Caetano et al., 2019), we create our primary cognitive skill measure by applying iterated principle factor analysis to the standardized letter-word (lw), applied problems (ap), and passage comprehension (pc) subtests of the Woodcock Johnson Revised Tests of Achievement, Form B, which are available in each CDS wave. These three measures are strongly positively correlated — in our pooled sample, $corr(\mathrm{lw}, \mathrm{ap}) = 0.83$, $corr(\mathrm{lw}, \mathrm{pc}) = 0.90$, and $corr(\mathrm{ap}, \mathrm{pc}) = 0.84$. We likewise construct our non-cognitive skill measure through iterated principle factor analysis applied to parental assessments captured in all 36 questions on the child's behavior available in the PSID CDS. The loading factors for these scales are shown in Table B.7 in Appendix B. Our cognitive and non-cognitive measures are all constructed so that a higher score of each component is better and are all normalized to have a mean of zero and a standard deviation of one.

[12] Although homework is assigned by teachers, the amount of time children spend on homework (and even whether they do their homework at all) is a choice. Moreover, parents (for children at younger ages) and the children themselves (at older ages) choose which schools/classes to enroll in, thereby choosing indirectly the amount of homework they are assigned. Thus, as with other forms of enrichment, there is potential endogeneity in the choice of homework time.

For robustness, we also use as alternative measures of non-cognitive skills the internalizing and externalizing subscales of the behavior problems index (BPI), a standardized scale included in each CDS wave. The internalizing scale captures the prevalence of withdrawn behaviors, while the externalizing scale captures outwardly aggressive behaviors (Peterson & Zill, 1986). We also use each component of our cognitive skill measure (applied problems, letter word, and passage comprehension) as separate measures of cognitive skill.

Controls: We use only controls that are pre-determined from the perspective of I, in order to preserve the meaning of our estimates as total effects of I, rather than direct (partial) effects, as discussed in Section 2.2. Our list of controls includes the child's age and squared age (in months), and indicators for: CDS wave (1997, 2002 and 2007), grade (thirteen variables, from kindergarten through grade 12), gender, ethnicity (black, Hispanic and other non-white ethnicity), whether the child has siblings, family income tercile, whether the mother is alive, and whether the father is alive.[13] As a robustness check, we drop the grade indicator variables from the control set, since they might be influenced by enrichment, particularly conditional on age and CDS wave. Dropping these controls barely changes our estimates. Importantly, we do not include lagged skills and time spent on other activities as controls, as we discuss in Section 2.

Summary Statistics: Table 1 presents summary statistics for our sample. We have a pooled sample of 4330 children ranging from 5 to 18 years of age, with an average age of just under 12.

While children in our data spend on average about 45 min per day 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, 1% are home schooled, and 8% attend a private school. The grade ranges pre-kindergarten–5, 6–8 and 9–12 are roughly equally represented in the sample, with about one-third of the observations each. The other columns of the table show the analogous summary statistics separately for each of these grade ranges. It is worth mentioning two striking patterns by grade. First, the time spent on duties/chores increases substantially on average in high school, which may reflect the additional responsibilities given to teenagers. Second, the standard deviation of the time spent on various activities is greater for older children. This makes sense: older children have a longer time to form different habits and interests, and thus some specialization in certain activities might be expected as they grow. Because different teenagers specialize in different lifestyles, the variance is higher for them. This increased variance leads to a disproportional amount of bunching of enrichment at zero in high school relative to earlier grades. Intuitively, more students in this age range might want to "borrow" more time from enrichment by making it negative (if they could) in order to intensify their investment in other endeavors, such as work, sports, or leisure. We further elaborate on this intuition in the next section.

4. Identification strategy

Recall that the model we wish to estimate is given by Eq. (3), restated below

$$S = \beta I + U.$$

[13] For some of these control variables, some observations have a missing value (less than 1% of the sample). In these cases, we include the missing observations in our sample by assigning them a unique value for the relevant control variable and creating an indicator variable for whether that observation had a missing value for that control. We then include these indicators as additional controls. The resulting estimates are very similar to the case where we simply drop all observations with any missing control variables.

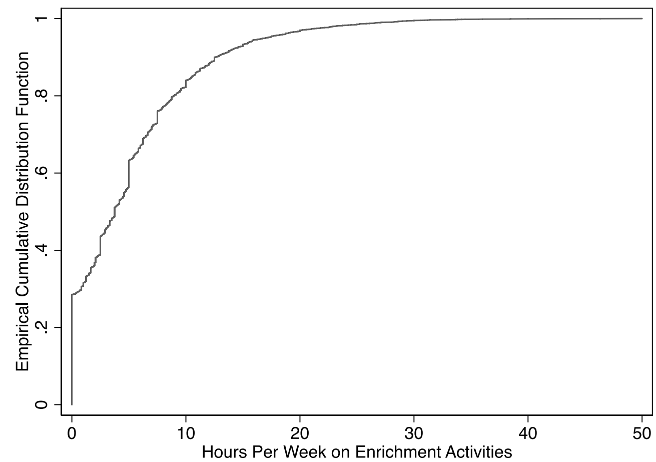

Fig. 1. Unconditional distribution of I. Note: This figure shows the cumulative distribution function of I.

The challenge in estimating β is that I and U are likely to be correlated. In this section, we show how our empirical approach tackles this endogeneity concern using bunching. First, we argue that I is the result of a constrained choice. Based on this observation, we build a control function that aims to solve the endogeneity problem.

4.1. Evidence that enrichment is a constrained choice

We start by showing direct evidence that the distribution of enrichment has bunching at zero. Fig. 1 shows that the cumulative distribution function (CDF) of enrichment is fairly smooth for all positive values of enrichment, but about 30% of the observations are bunched at zero.

Why would this bunching occur? We argue that bunching happens because some children are at a corner solution: they would like to choose a quantity of enrichment that is below zero, but they are constrained to choose only non-negative amounts. In this scenario, the group at zero enrichment is particularly heterogeneous, since they all choose the same amount of enrichment (zero), but the non-negativity constraint may be binding to different degrees for different children in the sense that their unconstrained choices would differ. Intuitively, different children at $I = 0$ want to "borrow against enrichment time" in different amounts in order to increase their time spent on other activities, but they cannot.

It is helpful conceptually to separate the actual, realized number of enrichment hours I from the number of enrichment hours the child would have chosen without the non-negativity constraint, which we denote I^*. The variable I^* represents a combination of factors, observed and unobserved, pertaining to the characteristics of the child and her environment (including the characteristics of the family), which lead the child to want to choose a given number of enrichment hours. Thus, while the treatment I is fully observed, I^*, which can be viewed as an index of the confounders that affect the choice of I, is only partially observed (when $I > 0$).

The separation between I and I^* helps us understand the bunching in Fig. 1. Children of "type" $I^* \geq 0$, choose exactly what their type leads them to choose: $I = I^*$. However, children of type $I^* < 0$ choose $I = 0$ because they cannot choose what their type would lead them to choose. Thus, there are two groups of children at $I = 0$: those of type $I^* = 0$ who are exactly indifferent between a marginally positive amount of I and $I = 0$ and those of types $I^* < 0$ who are away from exact indifference. While children of type $I^* = 0$ choose $I = 0$ as an

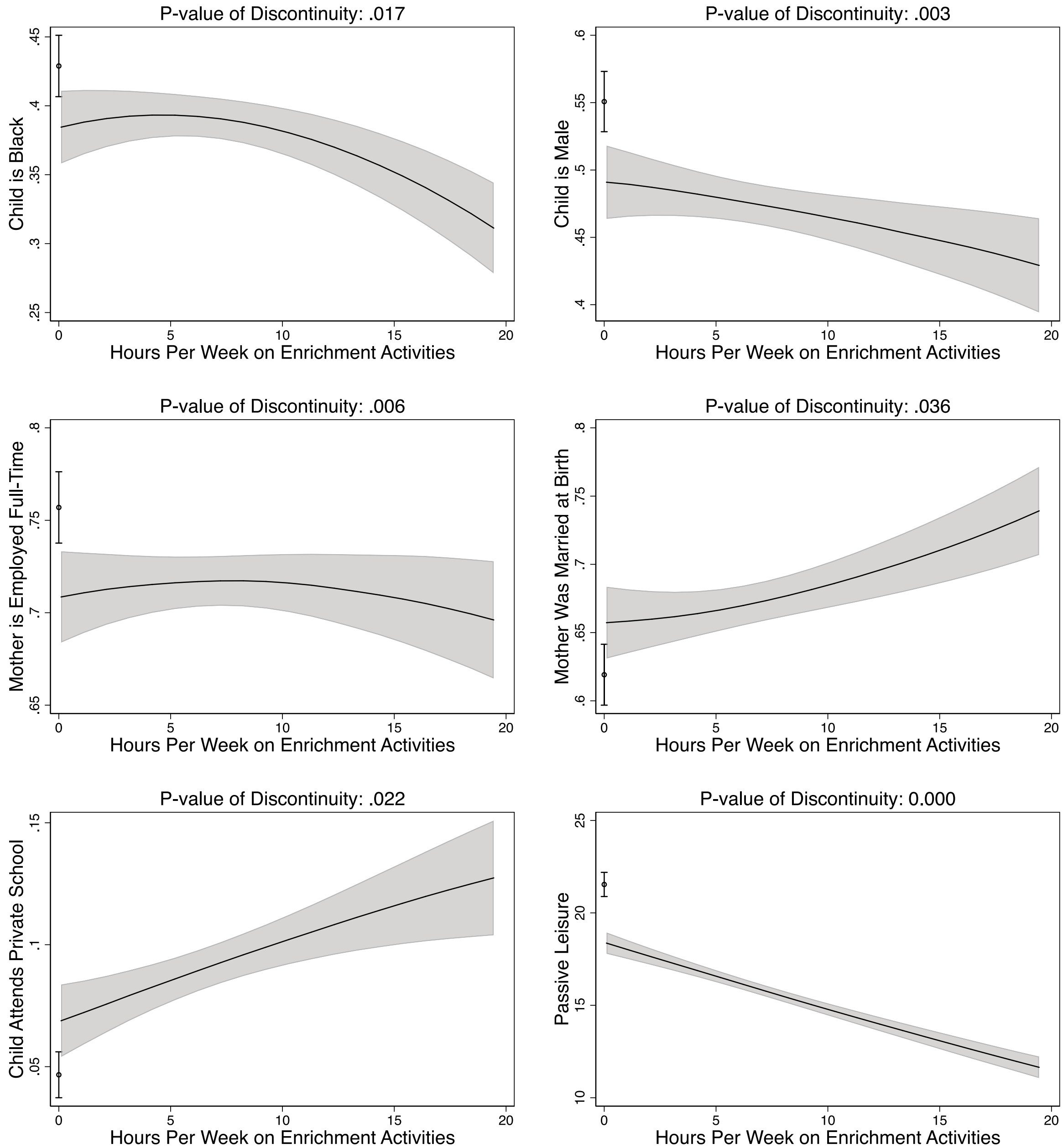

Fig. 2. Evidence of corner solution. Note: Each panel shows a plot of the local linear estimator (bandwidth = 10) of the expected value of a variable conditional on enrichment time, along with its 90% confidence interval. The expected value of the variable among the children who spent no time on enrichment is also shown, along with its 90% confidence interval. Finally, the *p*-value of a test for whether there is discontinuity at zero is shown in the header of each panel.

"interior solution", children of types $I^* < 0$ choose $I = 0$ as a "corner solution"– the restriction that I cannot be negative is binding for them. This implies that the average child who chooses a marginally positive amount of enrichment, $I > 0$, is likely very different from the average child at $I = 0$, since this latter group includes a discretely different group of children, namely those of type $I^* < 0$.

Fig. 2 substantiates this explanation. It shows that the average child who spends no time on enrichment is discontinuously different from the average child who spends any positive time at all on enrichment. The upper left panel of the figure shows a local linear fit of an indicator of whether the child is black conditional on the amount of time the child spends on enrichment, as well as the proportion of children who are black among the children who spend zero time on enrichment. The children at zero are discontinuously more likely to be black than the children who spend marginally positive amounts of time on enrichment. In the header of the panel, we show the *p*-value of a test of whether the share of black children is continuous at zero enrichment time, and it is clear that we can confidently reject this hypothesis ($p = 0.017$). The other panels of Fig. 2 show similar patterns. Children who spend no time on enrichment are discontinuously more likely to be male ($p = 0.003$), to have a mother who works full-time ($p = 0.006$), to have an unmarried mother at birth ($p = 0.036$), to not be enrolled in a private school ($p = 0.022$) and to spend more time on passive leisure activities ($p = 0.000$). That is, in each case, we find that the children at zero seem to be negatively selected on observables associated with higher expected achievement.[14]

The last panel in Fig. 2 reflects the stark differences in the lives of the children spending no time versus those spending a little time on

[14] Fig. 3 in Section 6 provides evidence that this is also true with respect to unobservables.

enrichment. Children at zero enrichment spend on average four more hours per week on passive leisure (mostly TV) than the children at one hour of enrichment. Since the total number of hours in a week is the same for everyone, this means that these two groups of children are on average spending at least three additional hours per week on different activities, beyond the one hour difference in enrichment and passive leisure.

In the next section, we show how we use the fact that confounders are discontinuous at $I = 0$ in order to control for them. But first, we provide some additional intuition for why a key unobserved confounder is likely to be discontinuous at $I = 0$.

Remark 4.1 (*"Ability" as a Confounder*)**.** A key potential confounder in this literature is the ability of the child to perform enrichment activities. For instance, children with higher reading ability might want to spend more time reading (because they might get more rewards from it), which would generate a positive correlation between skills and enrichment due to reverse causality.[15] Fortunately, ability is likely to be discontinuous at $I = 0$, so we should be able to control for it with our approach. To see this, note that there are likely two types of children choosing $I = 0$: (a) those with abilities that are very similar, on average, to those at small values of $I > 0$, and those who might have abilities that are very different from those at small values of $I > 0$. At $I = 0$, these two groups are lumped together, so the average ability at $I = 0$ should be very different from the average ability for any small value of $I > 0$.[16]

4.2. Control function approach

Our approach is motivated by our findings in Section 4.1: some children are at a corner solution when choosing the time they spend on enrichment activities. Let I^* denote the latent *desired* choice of enrichment hours per week. The treatment variable I is equal to this latent desired variable I^* only when it is non-negative:

$$I = \max\{0, I^*\}, \quad \text{with} \quad 0 < \mathbb{P}(I^* < 0) < 1. \tag{4}$$

The bunching condition $0 < \mathbb{P}(I^* < 0) < 1$ in Eq. (4) implies that a portion of the observations are at a corner solution, so that their desired choice in the unconstrained optimization would have been different from their actual choice in the constrained optimization. This condition is clearly met in our case based on the evidence of bunching presented in Fig. 1 and the discontinuities shown in Fig. 2.

We now consider the outcome equation, where we impose structure by specifying the relationship between I^* and U (from Eq. (3)). In particular, we assume that U is linear in I^*:

Assumption 1 (*Linearity in I^**)**.** $U = h(X) + \delta(X)I^* + \epsilon$ where $h(X)$ and $\delta(X)$ are nonparametric functions of controls X, and $\mathbb{E}[\epsilon|X, I^*] = 0$.

Note that I^* is by construction a sufficient index for all confounders in I because it tracks any variation in I. Thus, if we were to have written U as a general function of I^*, we would have made no assumption at all. The restriction we place on U then is that it varies with I^* linearly conditional on controls X. Because X is allowed to enter nonparametrically (on both $h(X)$ and $\delta(X)$), controls may play a useful role in relaxing Assumption 1, as only the component of I^* that varies across observations with the same value of X needs to be restricted to enter the outcome equation linearly. In Sections 6 and 7.1, we provide more context and supporting evidence for Assumption 1.

Eq. (3) and Assumption 1 together imply

$$S = \beta I + \underbrace{h(X) + \delta(X)I^* + \epsilon}_{U}. \tag{5}$$

This equation makes clear that if we could observe I^*, we would be able to identify β through Assumption 1 alone. Because I^* is not observed when $I^* < 0$, we instead seek a proxy of it. Noting that $I^* = I + I^*\mathbf{1}(I = 0)$ and taking expectations,[17]

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

Eq. (6) shows that $I + \mathbb{E}[I^*|I = 0, X]\mathbf{1}(I = 0)$ is a proxy for I^*. Thus, if we could somehow identify $\mathbb{E}[I^*|I = 0, X]$, then we could add $I + \mathbb{E}[I^*|I = 0, X]\mathbf{1}(I = 0)$ to the regression as another control, allowing us to identify β, $h(X)$ and $\delta(X)$.[18]

Intuitively, $\mathbb{E}[I^*|I = 0, X]$ is the average distance from indifference between spending and not spending time on enrichment among those children who bunch at zero. By Eq. (4), we know that $\mathbb{E}[I^*|I = 0, X] < 0$. However, we do not know its exact value because I^* is not observed when it is negative. In order to achieve point identification of β, we proceed by making distributional assumptions on I^* that allow us to point-identify $\mathbb{E}[I^*|I = 0, X]$. In particular, we consider three alternative distributional assumptions, which are nested and ordered from strongest to weakest:

Assumption 2 (*Distributional Assumption*)**.** The distribution $I^*|X$ follows one of three distributions:

1. (Homoskedastic Tobit): $I^*|X \sim \mathcal{N}(X'\theta, \sigma^2)$.
2. (Semiparametric Tobit): $I^*|X \sim \mathcal{N}(\mu(X), \sigma^2(X))$
3. (Nonparametric Tail Symmetry): Let $q(X)$ denote the $(1 - \mathbb{P}(I = 0|X))$th quantile of $I^*|X$. Then, for all $y \leq 0$,

   $$\mathbb{P}(I^* \leq y|X) = 1 - \mathbb{P}(I^* \geq q(X) - y|X).$$

The homoskedastic Tobit assumption states that $I^*|X$ is normally distributed with a mean that depends linearly on X and a variance that is constant. The semiparametric assumption relaxes both the linearity of the conditional mean in X and the homoskedasticity assumptions, keeping only the normality assumption. The nonparametric tail symmetry assumption relaxes the normality assumption, stating that the unobserved lower tail of $I^*|X$ is the mirror image of the corresponding upper tail. Thus, for example, if 30% of the observations are bunched at zero, which is the average bunching in our data, the nonparametric tail symmetry assumption states that the distribution of the bottom 30% of the data is the mirror image of the distribution of the top 30% of the data.

Two points are worth emphasizing regarding Assumption 2. First, because $I^*|X$ is fully observed above the bunching point, it is possible to assess the plausibility of any particular distributional assumption using the non-bunched portion of the data. Indeed, we carry out such an analysis in Section 6.2. Second, we adopt the nonparametric tail symmetry assumption as our preferred assumption because it is both fairly general and matches the data well. However, we view tail symmetry simply as an assumption consistent with our data that allows us to achieve point identification. In Section 7.2, we consider a wide variety of other distributional assumptions (symmetric or otherwise), showing that our main empirical conclusions are not an artifact of this particular distributional assumption.

Summing up, beyond the existence of bunching (Eq. (4)), our approach makes two identifying assumptions: (1) conditional on controls

[15] Conversely, children with lower reading ability might spend more time on tutoring lessons to compensate for it, which would generate a spurious negative correlation between skills and enrichment.

[16] This argument assumes that there are enough children in group (b), i.e., that there is excess heterogeneity in those children at $I = 0$ relative to those at small values of $I > 0$. Figs. 2, 3 and B.9 suggest that this must be the case, otherwise the discontinuities documented there would be difficult to explain.

[17] Note that $\mathbb{E}[\epsilon|I, X] = 0$ by Assumption 1 and the Law of Iterated Expectations ($\mathbb{E}[\epsilon|I, X] = \mathbb{E}[\mathbb{E}[\epsilon|I^*, X]|I, X] = 0$).

[18] Although equivalent, the model presented here uses a simpler notation from the one presented in Caetano et al. (2023). See Bertanha, Caetano, Jales, and Seegert (2022).

X, the component of skills that is not attributable to treatment effects is linear in I^* (Assumption 1) and (2) one of the three distributional assumption on $I^*|X$ (Assumption 2). Using any of the three distributional assumptions, we can estimate $\mathbb{E}[I^*|I=0,X]$ using the formulas presented in Appendix A. The estimates of $\mathbb{E}[I^*|I=0,X]$ can then be used to generate the new regressor $I+\hat{\mathbb{E}}[I^*|I=0,X]\mathbf{1}(I=0)$ that can be added to the model specified in Eq. (6).

Remark 4.2 (*What is Inside I^*?*)**.** Because it tracks all variation in I, the variable I^* is a composite index of every unobserved factor affecting our measure of enrichment. It therefore likely includes factors associated with preferences (e.g., the child's personality, family values regarding education), constraints (e.g., whether the grandparents live nearby, commuting distance to available enrichment activities) and any potential measurement error in our measure of enrichment (e.g., the interviewed week was not representative of their typical week[19]). Because this list is diverse, it is possible that (a) I^* represents a different mix of these factors for different children, and (b) I^* has a different effect on skills depending on the child. It is even possible that I^* might have a positive effect on skills for some children while having a negative effect on skills for other children. To allow for possibility (a), we allow the distribution of I^* to be different across different values of X (see Fig. 6). To allow for possibility (b), we further allow the effect of I^* on skills in Eq. (5), $\delta(X)$, to vary across different values of X (see Fig. 7). We also test for, and rule out, the possibility that the effect of I^* on skills varies with I (i.e. $\delta(I,X)$), which might happen, for instance, if some confounders exist only for some values of $I>0$ but not for $I=0$ (see Figs. 4 and 5).

4.3. Estimation details

Section 3 details our list of controls X, some elements of which are continuous. We therefore "discretize" X before estimating the expectation $\mathbb{E}[I^*|I=0,X]$ under the semiparametric Tobit and nonparametric tail symmetry assumptions. Let $\{\hat{C}_1,\dots,\hat{C}_K\}$ be a finite partition of the support of X into sets, which we call clusters, and let $\hat{C}_K=(\mathbf{1}(X\in\hat{C}_1),\dots,\mathbf{1}(X\in\hat{C}_K))'$ be the cluster indicators. In the estimation of the expectation, we substitute X with $\hat{C}_K$, which has finite support. The estimator $\hat{\mathbb{E}}[I^*|I=0,X]=\hat{\mathbb{E}}[I^*|I=0,\hat{C}_K]$ is thus constructed using a two-step procedure in which first X is discretized and then either semiparametric Tobit or nonparametric tail symmetry is applied separately for each cluster.

We select the clusters so that observations in the same cluster have similar Xs. The clustering method we adopt means that as K (the number of clusters) grows, the observations within the same cluster become more similar in terms of how close are the values of X.[20] In general, if $\mathbb{E}[I^*|I=0,X]$ is continuous, then $\hat{\mathbb{E}}[I^*|I=0,\hat{C}_K]$ will approximate $\mathbb{E}[I^*|I=0,X]$ as K grows.

We also use the same clusters to relax the specification of controls. We specify controls as $h(X)=X'\tau+\sum_{k=1}^{K}\alpha_k\mathbf{1}(X\in C_k)$ in Eq. (6), so the cluster indicators control nonparametrically for differences across clusters, while differences within cluster due to X are controlled linearly. As the number of clusters K increases, the nonparametric match improves, leaving less unexplained variation within cluster, which in part can be controlled for X linearly. Putting this all together, our homogeneous estimates are based on linear regressions of the form

$$S=\beta I+X'\tau+\sum_{k=1}^{K}\alpha_k\mathbf{1}(X\in C_k)+\delta(I+\hat{\mathbb{E}}[I^*|I=0,\hat{C}_K]\mathbf{1}(I=0))+\varepsilon.$$

Our baseline estimates use $K=50$ total clusters and assume further (as in the above equation) that δ does not vary with X. Section 7.1 contains robustness analyses which relax both of these choices. First, we change the value of K. Second, we allow for $\delta(X)$ to change with clusters of X by specifying $\delta(X)=\sum_{k=1}^{K_\delta}\gamma_k\mathbf{1}(X\in C_k)$ for various values of K_δ. We show that the estimates of β barely change when either K or K_δ increases.

We report everywhere bootstrapped standard errors in which we re-estimate the cluster-level expectations with each bootstrap iteration. Thus, our reported standard errors account for the first-stage estimation error associated with the generated regressor approach. See Caetano et al. (2023) for a formal justification of the bootstrap in this setting.

5. Empirical results

5.1. Full-sample estimates

Table 2 presents our main results estimated on our full sample not broken down by grade range. Column (i), which reports the results of simple regressions of skills on enrichment time without controls, shows that both cognitive and non-cognitive skills are strongly positively correlated with enrichment time. Column (ii), which adds our full set of controls (including cluster indicators) into the specification from column (i), shows that while observables seem to explain part of the correlation between enrichment time and skills, the residual relationships remain positive, particularly for cognitive skills.

The remaining columns in Table 2 show our corrected estimates of β (Eq. (6)) under the different assumptions on the distribution of $I^*|X$ discussed in Section 4 ranging from the strongest (homoskedastic Tobit) to the weakest (nonparametric tail symmetry).

For cognitive skills, all of the corrected estimates are quite similar — the β estimates fall from 0.011 standard deviations (s.d.) with controls but no correction to around −0.004 s.d. The large differences between column (ii), where the estimate is positive and highly significant, and columns (iii)-(v), where the estimates are negative and insignificant, show that our correction method is able to handle endogeneity which was not absorbed by the pre-determined controls. Our most general correction method (tail symmetry) yields a 90% bootstrapped confidence interval of $[-0.012,0.008]$.

Correcting for selection has even more dramatic consequences for the non-cognitive estimates — the corrected non-cognitive β's are all negative, with point estimates ranging between −0.024 and −0.015 s.d. The point estimate using the tail symmetry method (column (v)) is −0.019, significantly different from zero at 10%. This point estimate implies that an additional hour of enrichment time per week causally lowers non-cognitive skills by 0.019 sd.

For both cognitive and non-cognitive skills, the δ estimates are positive and highly significant, confirming the evidence from Fig. 2 that confounders tend to be positively correlated with S (see also Fig. 3). The fact that the β estimates in the "No Controls" column (i) are larger than in the "Uncorrected" column (ii) provides yet further evidence of positive bias.

Finally, note that the standard errors in column (ii) of Table 2 are much smaller than the standard errors from the corrected models in columns (iii)-(v). This is a feature of our approach, not a bug. The only difference between the corrected and uncorrected models is the presence of the term $I+\hat{\mathbb{E}}[I^*|I=0,X]\mathbf{1}(I=0)$. Adding one covariate

[19] Following the prior literature on time use, we try to avoid measurement error of this sort by restricting the sample to observations where the interviewed week was said to be representative of a typical week. Our approach still allows for measurement error, provided it is discontinuous at $I=0$. Caetano et al. (2019) shows this to be the case in the PSID-CDS data for observed variables related to measurement error (e.g., whether interview respondent was the primary caregiver instead of the child, whether the interview was completed with the help of an interviewer, and whether the interview was concluded face-to-face or by phone).

[20] We show results using hierarchical clustering, which is known for its stability and simplicity and for its ease of interpretation as we vary the number of clusters. Nevertheless, we obtained similar results with other clustering methods. Hierarchical clustering requires the choice of a dissimilarity measure and a linkage method. The reported results use the Gower measure and Ward's linkage, but we obtained similar results with other choices.

Table 2
Full-sample results: The effect of enrichment time on skills.

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

Note: N = 4330. Bootstrapped standard errors in parentheses (500 iterations). Columns (ii)–(v) use all control variables discussed at the end of Section 4. Columns (iii)–(v) makes two identifying assumptions: Eq. (5) and a distributional assumption on $I^*|X$. Column (iii) assumes homoskedastic Tobit, column (iv) assumes semiparametric Tobit, and column (v) assumes nonparametric tail symmetry. Section 6 tests these two assumptions. Specifications (ii)–(v) use $K = 50$ and specifications (iii)–(v) use $K_\delta = 1$, where K and K_δ refer to the total number of clusters defined in Section 4.3. See Figs. 6 and 7 in Section 7.1 for estimates of β for different values of K and K_δ. ** $p < 0.05$, * $p < 0.1$.

should not necessarily cause the standard errors in a regression to increase, particularly so dramatically. The fact that we see large increases in the standard errors in our application suggests greater underlying uncertainty surrounding the true causal effects of enrichment time on skills once endogeneity is accounted for. Not considering this correction term would lead to two inferential problems: estimates that are both biased and overly precise. This could lead to excessively optimistic and confident expectations of policymakers or families regarding the benefits of enrichment activities.

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

Our results are also robust to alternative definitions of enrichment time. First, we consider broad enrichment, which expands the notion of enrichment to include additional activities less directly oriented towards the development of cognitive skills such as organized sports, arts, and volunteering, as detailed in Section 3. Table B.9 in Appendix B shows that using this broader measure yields remarkably similar estimates to the baseline results presented in Table 2. The uncorrected estimates again show significant, positive associations between (broad) enrichment and skills, while the corrected estimates again indicate a null effect for cognitive skills and a significant negative effect for non-cognitive skills. Indeed, the corrected non-cognitive point estimate assuming symmetry is very similar to the baseline estimate and is significant at the 95% level. Conversely, when we restrict the definition of enrichment to consist only of homework, we find the same pattern of cognitive estimates near zero and non-cognitive estimates that are even more negative and even more significant (Table B.10 in Appendix B). The following remark discusses our results for homework and relate them to the literature on the topic.

Robustness to Using Lagged Skill Measures as Controls: As discussed in Section 2.2, lagged test scores have been used successfully in prior literature to handle endogeneity. We do not use lagged scores in our primary analysis because doing so reduces our sample size substantially, rendering an analysis of by-grade heterogeneity impossible. Nonetheless, as a robustness exercise, in Table B.12 (Appendix B) we report the estimates analogous to those presented in Table 2, but with the inclusion of lagged scores. The subsample used in Table B.12 is naturally much smaller, with about a third of the size of the full sample, and is notably selected.[21] The results with lagged skill measures are generally closer to zero and less precise, but the qualitative conclusions are unchanged. For cognitive skills, the uncorrected estimates are positive and significant while the corrected estimates are close to zero and insignificant. For noncognitive skills, the corrected estimates are negative, although they are not statistically significant. In particular, we cannot reject at standard levels equality between the estimates with and without lagged skills.

Remark 5.1 (*Homework*)**.** Homework is the single largest component of our measure of enrichment activities, so we provide here a more detailed discussion of our findings in relation to the literature on the effects of homework on childhood skills, beyond our broad discussion in the introduction. To our knowledge, no paper in this literature estimates the impact of homework on non-cognitive skills. Moreover, only recently have there been studies estimating the effect of homework on cognitive skills which systematically consider confounders. Specifically, Aksoy and Link (2000) uses child-fixed effects to estimate the effect of homework on math scores for 10th graders, and Eren and Henderson (2011) uses the within-child, between-subject identification strategy pioneered by Dee (2007) to absorb both child- and teacher-specific unobservables to estimate the effect of homework for 8th graders on math, science, English and history achievement. The main finding in this literature is that math homework is effective for increasing math skills, but the effect seems to be smaller and insignificant for lower socioeconomic status children. Eren and Henderson (2011) also finds that homework in other subjects (English, science and history) does not seem to be effective on average and, moreover, that there does not seem to be any cross-subject spillover (i.e., effect of homework of a given subject on the skill of another subject), even for math homework.

For comparison, we estimate the effect of homework time alone on different measures of skills and find results broadly consistent with this prior literature. The point estimates are twice as large for applied problems relative to letter-word, and even larger relative to passage comprehension. These results can be found in Table B.11 in Appendix B. Note that, although all of the point estimates are positive, they are mostly insignificant at standard levels. This is consistent with the findings in the homework literature: our measure of homework is a composite of all types of homework, while Eren and Henderson (2011) uses specific measures of homework for each subject, so naturally our estimates should be smaller, especially given that paper's evidence of the ineffectiveness of non-math homework and its null findings for cross-subject spillovers.

[21] This subsample contains no children in grade range K-5, 487 children in grade range 6–8 and 1085 children in grade range 9–12. The subsample is also selected conditional on the grade range: among other differences, the children from this subsample tend to have a higher household income and are more likely to be White than the children in the same grade range in the full sample.

Table 3
Cognitive estimates by grade levels.

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

Note: Number of observations (N) for each grade range is shown. Bootstrapped standard errors in parentheses (500 iterations). Specifications (ii)–(v) use $K = 50$ and specifications (iii)–(v) use $K_\delta = 1$, where K and K_δ refer to the total number of clusters defined in Section 4.3. ** $p < 0.05$, * $p < 0.1$.

Table 4
Non-cognitive estimates by grade levels.

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

Note: Number of observations (N) for each grade range is shown. Bootstrapped standard errors in parentheses (500 iterations). Specifications (ii)–(v) use $K = 50$ and specifications (iii)–(v) use $K_\delta = 1$, where K and K_δ refer to the total number of clusters defined in Section 4.3. ** $p < 0.05$, * $p < 0.1$.

5.2. Estimates by grade

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

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

Table 4 repeats the analysis for non-cognitive skills. The uncorrected estimates suggest a significant, positive association between enrichment and non-cognitive skills for high school only. Interestingly, this grade range happens to be exactly the one in which we find the most evidence of endogeneity, as seen by the estimates of δ in columns (iii)–(v).[22] This accounts for why, when the endogeneity is corrected, the treatment effect estimates are negative and significant. As discussed in Section 3, this makes sense: this is the age when enrichment decisions diverge across families, with some teenagers preparing for selective college admission and others focusing on other endeavors.

The corrected non-cognitive estimates for youth in high school are economically large. Our preferred point estimate assuming tail symmetry implies that an additional hour of enrichment per week lowers non-cognitive skills by −0.039 sd. To put the effect size of our estimate in context, Jackson (2018) estimates that a 1 sd increase in teacher non-cognitive value-added corresponds to an increase of about 0.08 sd in non-cognitive skills, roughly twice the size of our estimate. A large literature in economics finds that non-cognitive skills are strongly correlated with various later-life outcomes; we can thus provide yet more context for our estimates by translating our non-cognitive effects to effects on later-life outcomes using estimates from this literature. For instance, Heckman, Stixrud and Urzua (2006) estimate that a 1 sd increase in non-cognitive skills is associated with 0.1–0.14 log-point increases in wages at age 30. Carneiro, Crawford, and Goodman (2007) similarly find that a 1 sd increase in non-cognitive skills is associated with an 8 month increases in cumulative labor market experience by age 42 and a 20% reduction in the likelihood of being diagnosed with a mental health disorder by age 42. These and other estimates from the literature therefore suggest that increases in enrichment time on the order of a few hours per week should correspond to economically significant decreases in various life outcomes.

[22] To see stronger evidence in favor of the claim that endogeneity is more pronounced in high school than in other grades, compare the larger discontinuities in the unobservables for high school in Fig. B.9 in Appendix B with the discontinuities for the full sample in Fig. 3. This evidence is stronger because it makes neither distributional nor linearity assumptions, see Section 6.1.

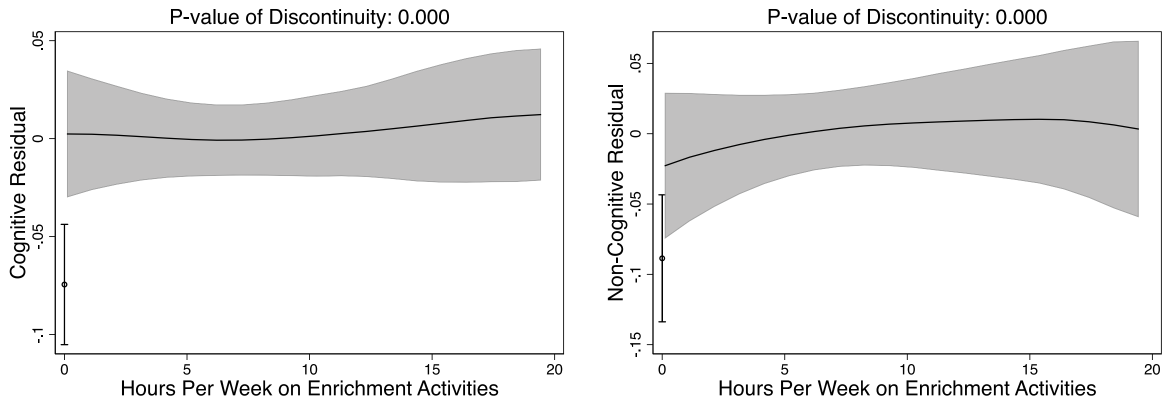

Fig. 3. Two tests under no distributional assumption. Note: Each panel shows two tests, one at $I > 0$ and another at $I = 0$, both based on the expected value of the residuals from Eq. (3) (estimated for $I > 0$ only) conditional on I. We show a local linear plot of these residuals (bandwidth = 10) along with the 90% confidence interval. For $I > 0$, it is clear that the linearity cannot be rejected, so we cannot reject Assumption 1. For $I = 0$, it is clear that the average of the residuals is negative, which is equivalent to $\delta > 0$ (p-value of the test of whether $\delta = 0$ is also shown at the top). These two tests make no distributional assumption. In these regressions, $K = 50$ as in the main results shown in Section 5.

6. Testing the identification assumptions

The correction methodology relies on two key assumptions: (1) that unobservables (indexed by I^*) affect skills linearly conditional on X (Assumption 1) and (2) the particular distributional assumption made on $I^*|X$ necessary to identify $\mathbb{E}[I^*|I = 0, X]$ (Assumption 2). In this section, we discuss the tests we implement to detect violations of either of the two identifying assumptions of the paper, separately or jointly.

6.1. *Testing Assumption 1 alone*

If Assumption 1 holds for $\delta(X) = \delta$,[23] then for $I > 0$, $\mathbb{E}[S|I, X] = (\beta + \delta)I + h(X)$. That is, Eq. (5) implies that the conditional expectation for $I > 0$ must be linear in I, and this does not depend on any distributional assumption. Therefore, Eq. (5) may be tested by applying any specification test to the regression of Y on I and $h(X)$ for $X > 0$.

We run such regressions, collect the residuals and show the local linear polynomial estimators of the expected value of these residuals conditional on I. The results of this exercise can be seen in Fig. 3. For both cognitive and non-cognitive skills, the average residual is close to zero for all values of $I > 0$. This suggests that the linearity assumption in Eq. (5) holds for the full sample. (Fig. B.9 in Appendix B shows the analogous figure for the high school subsample.) Standard specification tests (e.g. Ramsey, 1969's RESET test) support this interpretation, as they fail to reject our preferred specification of Eq. (5).

Fig. 3 also shows the results of a second test. As demonstrated in Caetano et al. (2023), the average residual at $I = 0$ is equal to $\delta\mathbb{E}[I^*|I = 0]$. Since $\mathbb{E}[I^*|I = 0] < 0$, the sign of the average residuals at zero is opposite to the sign of δ. Accordingly, the header of each panel shows the p-value of a test of whether $\delta = 0$.[24] The results suggest that $\delta > 0$ for both cognitive and non-cognitive skills. Indeed, our estimates of δ in Section 5 are positive, but note that $\delta > 0$ can already be inferred here with no distributional assumption.

6.2. *Testing Assumption 2 alone*

The correction method requires a distributional assumption on $I^*|X$ to permit the identification of $\mathbb{E}[I^*|I = 0, X]$. In this section, we test the distributional assumptions we make in our empirical work directly, without any assumption on Eq. (5), using Kolmogorov-Smirnoff-type tests. We can do this because $I^* = I$ when $I > 0$, so we can compare the observed distribution of I above 0 to that implied by any particular distributional assumption.

We first test the Homoskedastic Tobit assumption by comparing the enrichment distributions for white versus Hispanic high school students.[25] Fig. B.10 in the Appendix shows the homoskedastic fit, and the evidence of heteroskedasticity is clear, as the variance for white high school students is greater than the variance for Hispanic high school students. Indeed, we can reject this distributional assumption at standard levels of significance using a Kolmogorov–Smirnov test.

Next, we consider the semiparametric Tobit assumption, which relaxes the linear mean and homoskedasticity requirements while maintaining the assumption that $I^*|X$ follows a normal distribution. The fitted distributions for white and Hispanic high school students in this case (Fig. B.11 in the Appendix) clearly show that allowing for different variances by groups noticeably improves the fit relative to the homoskedastic case (Fig. B.10). Nonetheless, it is apparent that the upper tail of the data is heavier than the upper tail implied by a normal distribution for both groups.[26] We are not able to reject the semiparametric Tobit assumption with a Kolmogorov–Smirnov test, but we are able to reject it at 10% with the test developed in Goldman and Kaplan (2018) which is designed to have more power to detect deviations from the null at the extremities of distributions.

Finally, we drop the normality assumption entirely and assume only tail symmetry. Tail symmetry is not directly testable by observing the empirical distribution of I for $I > 0$. However, provided the bunching fraction is below 50%, (full) symmetry is testable because we can compare the fitted and raw distributions for part of the support.[27] We are unable to reject symmetry using two distribution tests (Kolmogorov–Smirnov and Goldman & Kaplan, 2018) for all values of X, irrespective of the total number of clusters of the observables or the grade range we consider. Because symmetry implies tail symmetry, we consider our estimates under nonparametric tail symmetry as our main results throughout the paper. For completeness, we show in Fig. B.12 the white and Hispanic high school distributional fits in this case.

[23] For simplicity, we denote $\delta(X)$ as δ when we do not allow it to vary with controls. See Fig. 7 in Section 7.1 as evidence that this restriction is not consequential.

[24] Similarly, t-tests of the average residuals at $I = 0$ versus the corresponding average for $I \in (0, 5]$ reject equality for both cognitive and non-cognitive skills.

[25] For concreteness, we maintain these comparison groups throughout this section. However, we find broadly similar results – rejections of normality but not of symmetry – using other comparison groups.

[26] We observe a similar pattern for most values of X.

[27] To see how symmetry is testable, consider the case where the bunching rate is 20%. Then we can compare the empirical distribution between percentiles 20 and 50 with the mirror image of the empirical distribution between percentiles 50 and 80. Under symmetry, the two should match.

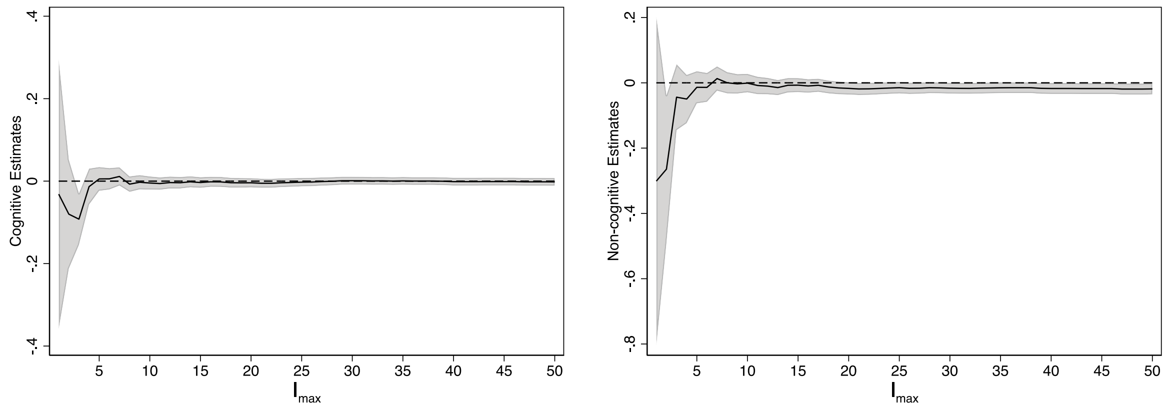

Fig. 4. Estimates for different truncations of the data by $I_{\max}$ — Full sample. Note: Each panel shows the estimate of β for cognitive (left panel) or non-cognitive skills (right panel) restricting the sample to only children whose enrichment hours are lower than or equal to $I_{\max}$ for values of $I_{\max}$ ranging from 1 (only those who chose $I = 0$ or $I = 1$) to 50 (everyone). We cannot falsify our two identifying assumptions with this test.

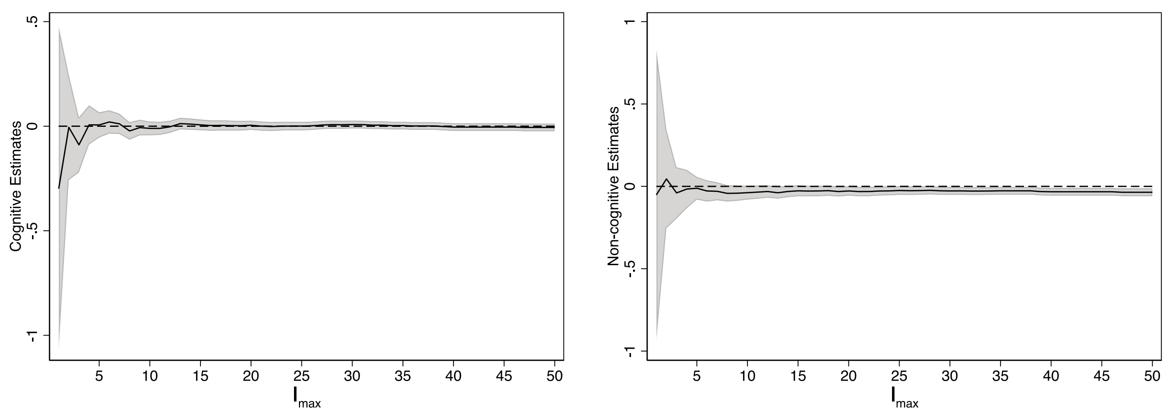

Fig. 5. Estimates for different truncations of the data by $I_{\max}$ — Grades 9–12. Note: Each panel shows the analogous estimates to Fig. 4 but for the sub-sample of children in high school.

It is reassuring that we are able to detect violations of the homoskedastic and semiparametric Tobit assumptions, even though their resulting estimates of β are similar to our preferred estimates of β under nonparametric tail symmetry. This is consistent with what Caetano et al. (2023) finds in a Monte Carlo study of the method: the distributional tests we use are very sensitive in the sense that they can detect violations of the distributional assumption even when they are small enough to have only modest consequences for the estimate of β.

6.3. Testing Assumptions 1 and 2 jointly

We also jointly test both identifying assumptions by restricting the sample to $I \in [0, I_{\max}]$ for increasingly large values of $I_{\max}$ and applying the control function correction for models estimated on each of the restricted samples. If Eq. (6) holds and $\mathbb{E}[I^*|I = 0, X]$ is correctly identified (i.e., the distributional assumption is valid), then there can be no sample selection bias from restricting the sample to only observations such that $I \leq I_{\max}$ when running regression (6). Thus, the $\hat{\beta}$ estimates should be stable for different values of $I_{\max}$. One can then test whether the estimated coefficients are the same as $I_{\max}$ increases.[28]

Intuitively, this test relies on the idea that linearity is a weaker assumption locally than globally. Thus, nonlinearities in Eq. (5) or nonlinear errors in the identification of $\mathbb{E}[I^*|I^* \leq 0, X]$ will cause the estimates of β to not be stable as we increase $I_{\max}$. In fact, this test is particularly sensitive to deviations from the identifying assumptions that occur due to the presence of confounders that only affect the outcome for values of I away from $I = 0$. This follows because this type of confounder will tend to cause nonlinearities in $\mathbb{E}[Y|I, X]$ that show up for high enough values of $I_{\max}$, where the confounder operates. Among other potential problems, this test should be sensitive to essential heterogeneity in which higher-return youth spend more time on enrichment (Heckman, Urzua and Vytlacil, 2006).

In Fig. 4, we show how our main estimates of $\hat{\beta}$ for cognitive (left panel) and non-cognitive skills (right panel) change for different truncations of our sample ranging from $I_{\max} = 1$ to $I_{\max} = 50$. Irrespective of our choice of $I_{\max}$, we maintain the same estimate of $\mathbb{E}[I^*|I = 0, X]$ using the nonparametric tail symmetry assumption. As the maximum hours per week spent on enrichment in our full sample is 50, the estimates in the far right of each panel are the estimates reported in Table 2. We find that the sample-truncated estimates of β are mostly quite similar to the main estimates from Table 2, except for very small values of $I_{\max}$ where the estimates are more negative (albeit with substantially wider confidence intervals).

For completeness, Fig. 5 shows the analogous plots for high school age youth. The findings are similar. We conclude that our two identifying assumptions – Eq. (5) and the nonparametric tail symmetry assumption – cannot be rejected with this test.

7. Sensitivity analysis

In this section, we provide sensitivity analyses of our preferred estimates that either relax or consider violations of our identifying assumptions.

[28] A limitation of this test is that under heterogeneous treatment effects, it is possible that estimates change with $I_{\max}$ even if Assumptions 1 and 2 were valid, since the target estimand would change with the sample.

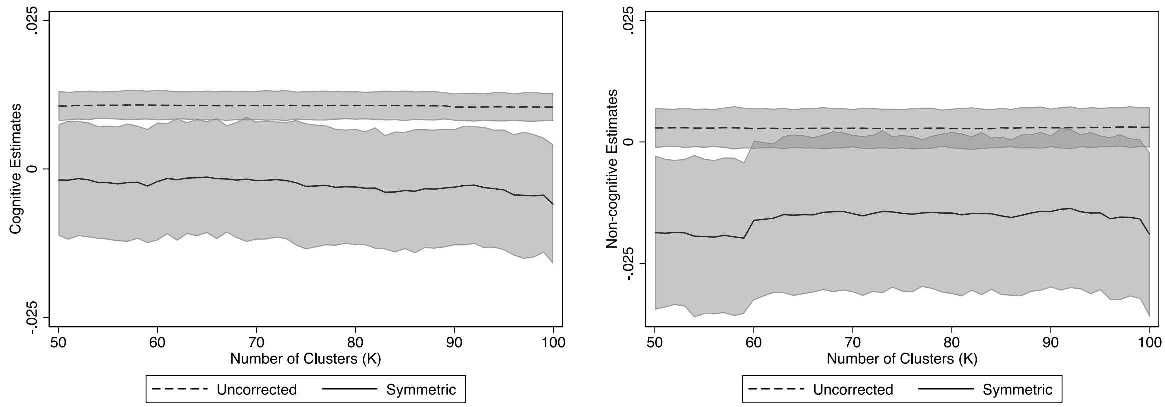

Fig. 6. Estimates by K. Note: This figure shows the estimates from columns (ii) and (v) for different number of clusters of X in the estimation of both $\mathbb{E}[I^*|I = 0, X]$ and $h(X)$. The shaded areas depict the 90% confidence intervals. All standard errors are bootstrapped using 500 iterations.

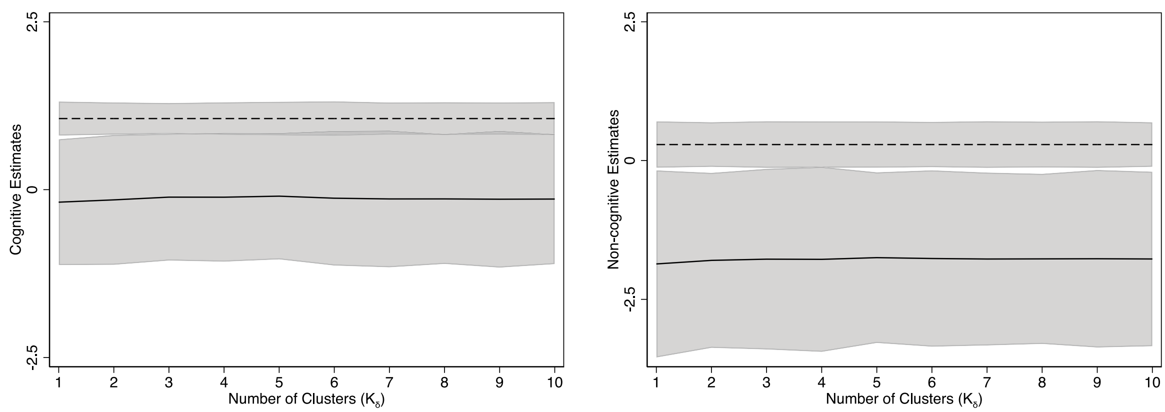

Fig. 7. Estimates by K_δ. Note: This figure shows the estimates from columns (ii) and (v) for different number of clusters of X in the estimation of $\delta(X)$. The left figure shows cognitive estimates, and the right figure shows non-cognitive estimates. The shaded areas depict the 90% confidence intervals. All standard errors are bootstrapped using 500 iterations.

7.1. Using clusters to relax the identifying assumptions

In Section 4.3, we discuss the use of indicators of clusters of X to approximate a nonparametric function of X. Because $h(X) = X'\tau + \sum_{k=1}^{K} \alpha_k \mathbf{1}(X \in C_k)$, the indicators $\mathbf{1}(X \in C_k)$ absorb all variation across clusters, while the linear component $X'\tau$ absorbs part of the within-cluster variation. As the total number of clusters grows, our specification becomes a better approximation of a nonparametric function of X because the observations within the same cluster become more similar to each other in terms of Xs, leaving less variation to be absorbed by the linear component. Thus, increasing the total number of clusters offers a simple sensitivity analysis to gauge whether the main estimates are sensitive to the choice of the number of clusters $K = 50$.

Fig. 6 shows the β estimates under tail symmetry for different numbers of clusters K. Recall that clusters are used in two ways here: to estimate the expectation $\mathbb{E}[I^*|I = 0, \hat{C}_K]$ as an approximation to $\mathbb{E}[I^*|I = 0, X]$ and to relax the specification of controls $h(X)$ by further controlling for differences in controls across clusters nonparametrically. The point at $K = 50$ in the figure is identical to the estimate of β shown in column (v) of Table 2, so increasing the number of clusters all the way to $K = 100$ does not change the estimates meaningfully. This gives us confidence that $K = 50$ seems to be sufficient to control for all confounders related to X.

Finally, we relax Eq. (5) by allowing $\delta(X)$ to vary by different clusters of X, as we did with $h(X)$ in Fig. 6 (see Section 4.3). By allowing I^* to have different effects on skills for different types of children belonging to different clusters, this generalization provides a lot of flexibility in the way that our method controls for confounders. For instance, it allows I^* to have a negative effect on skills for some values of X while having some positive effect on skills for other values of X. We keep $K = 50$ fixed and change the value of K_δ, the total number of clusters in $\delta(X)$. Fig. 7 shows that the estimates of β are nearly identical for both cognitive and non-cognitive skills as K_δ goes from $K_\delta = 1$ (as in the main results in Table 2) to $K_\delta = 10$. Because $\delta(\hat{C}_{K_\delta})$ should be approximating $\delta(X)$ better for higher values of K_δ, we are effectively weakening Assumption 1 as we increase K_δ, yet the main results do not change. These findings thus suggest that there is no further confounding variation even at $K_\delta = 1$.[29]

7.2. Assessing robustness to violations of Assumption 2

Here we assess the robustness of our key empirical results to violations of the distributional assumption. We do this using the insight that the distributional assumption only matters for our estimates insofar as it affects the identification of $\mathbb{E}[I^*|I = 0, X]$. Thus, we assess the robustness of our estimates to alternative distributional assumptions by systematically plugging in a wide range of values of $\hat{\mathbb{E}}[I^*|I = 0, X]$ to the generated regressor in Eq. (6). More concretely, for a range of values of the factor $f > 0$, we replace our preferred tail symmetry

[29] The uncorrected estimate from column (ii) in Table 2 is also shown in the figure, for comparison. Of course, because this estimator does not depend on the corrected term, the estimate does not change for different values of K_δ.

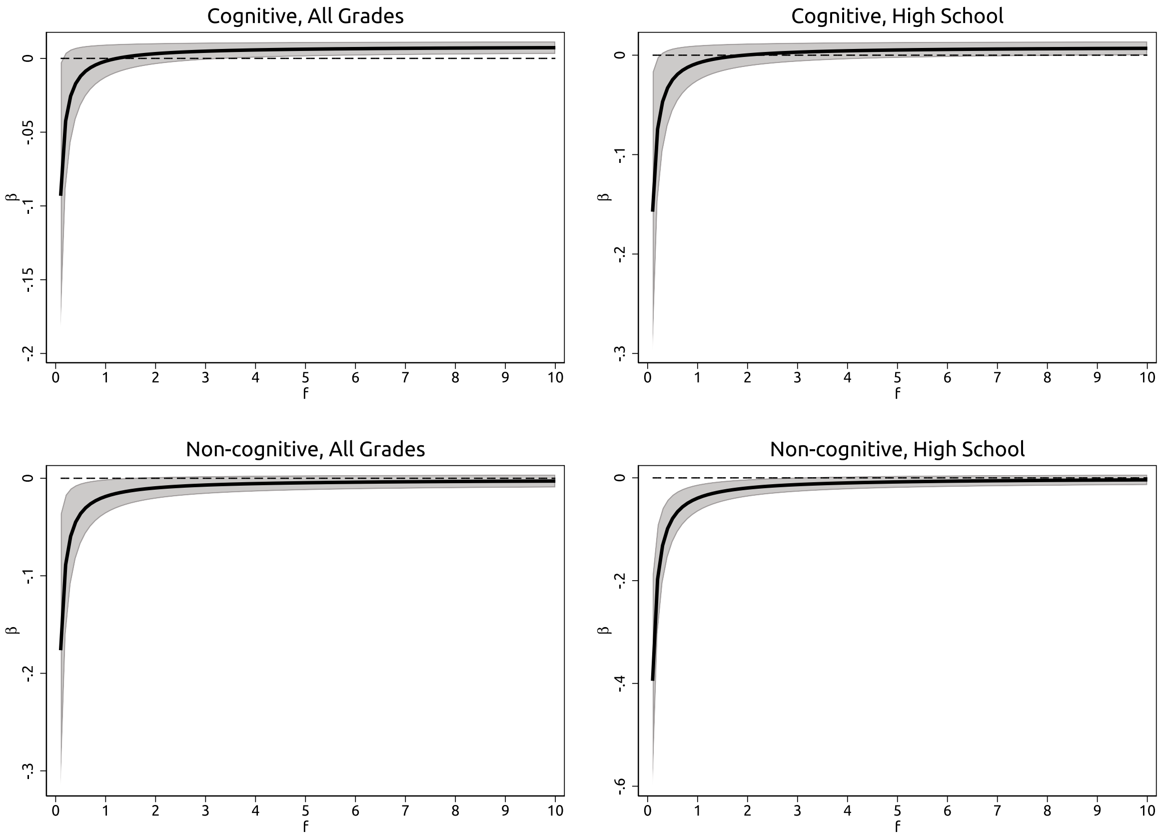

Fig. 8. Robustness to different estimated expectations. Note: For each panel, the thick black curve shows what would be the $\hat{\beta}$ from the regression in Eq. (6) were we to use $f * \hat{E}[I^*|I=0,X]$ in constructing the generated regressor instead of $\hat{E}[I^*|I=0,X]$ obtained assuming tail symmetry. The left two panels correspond to the sample and estimates presented in Table 2, while the right two panels correspond to the sample and estimates presented in Tables 3 and 4. The tail symmetry estimates from these tables are the exact ones represented in these plots when $f=1$.

estimates of the conditional expectations $\hat{\mathbb{E}}[I^*|I=0,X]$ by $f \cdot \hat{\mathbb{E}}[I^*|I=0,X]$, and we then carry out the rest of the estimation as before.

Fig. 8 shows the estimates for β corresponding to our headline estimates assuming tail symmetry from Tables 2, 3, and 4 for values of f ranging between 0.1 and 10 (i.e., from 10% to 1000% of the size of the tail symmetric estimated expectation). These plots make clear that our headline conclusions are robust to fairly severe mistakes in the estimation of $\hat{\mathbb{E}}[I^*|I=0,X]$. For example, the high-school cognitive estimates are indistinguishable from zero for $f > 0.2$ (above 20% of the tail symmetry estimates of $\hat{\mathbb{E}}[I^*|I=0,X]$) and negative for $f \leq 0.2$, while the high-school non-cognitive estimates are negative and statistically significant for $f < 2.7$ (below 270% of the tail symmetry estimates). For $f \geq 2.7$, the high school non-cognitive point estimates are still negative, although they are not statistically significant.

This analysis sheds light on why our empirical results are very similar across different distributional assumptions, including the two normality assumptions that are rejected by our data (see Section 6.2).

8. Understanding our empirical results

At first blush, our empirical results may seem surprising. Enrichment activities are oriented towards skill development, so why does the last hour of enrichment have a negligible effect on cognitive skills? Moreover, why does this last hour have a significant negative effect on non-cognitive skills in high school, but not in earlier grades? The framework developed in Section 2 suggests that a complete answer to these questions may depend on many factors: different skill production functions at different ages, different substitution patterns between enrichment and other time uses, and different usages of complementary resources such as parental time and market goods. Exploring and disentangling all of these potential factors is well beyond the scope of this paper. Instead, we offer here a speculative, yet parsimonious, explanation, based both on our data and a stylized model of optimal time allocation, that we believe can help understand our key findings.

Suppose for simplicity that there are only two uses of time: enrichment I and leisure L, with the total time budget normalized to 1. Define the cognitive (S_c) and non-cognitive (S_n) skills of each individual as smooth functions of these two inputs,

$$S_c = f_c(I,L) = f_c(I,1-I)$$
$$S_n = f_n(I,L) = f_n(I,1-I).$$

For notational ease we drop the subscript for each individual, with the understanding that not only the input I, but also the production functions f_c and f_n may vary with each individual depending on their observed and unobserved characteristics.

Next, we consider a stylized assumption that helps explain the main findings of the paper.

Assumption 3 (*Stylized Assumption on I, f_c, and f_n*).

a. I is chosen to maximize $f_c(I,1-I)$.
b. $\frac{df_n(I,1-I)}{dI} < \frac{df_c(I,1-I)}{dI}$ during high school.

Part (a) of this stylized assumption states that individuals choose their level of enrichment to maximize cognitive skills. Part (b) states that enrichment has a higher marginal return for cognitive than for non-cognitive skills at any level of enrichment during high school. Assumption 3 is not meant to be taken as valid for all individuals;

Table 5
Enrichment time breakdown.

Enrichment activity	All grades	Grades PreK–5	Grades 6–8	Grades 9–12
Homework	0.66	0.52	0.66	0.79
Reading a book	0.14	0.16	0.16	0.10
Before and after school programs	0.07	0.18	0.05	0.00
Other reading	0.04	0.05	0.05	0.03
Non-academic lessons	0.03	0.04	0.04	0.02
Other academic lessons	0.02	0.02	0.02	0.02
Other education	0.02	0.01	0.02	0.03
Being read To	0.01	0.03	0.01	0.00

Note: Each panel shows the average proportion of each type of enrichment activity over a typical week for different samples. The table pools the 1997, 2002 and 2007 CDS waves.

rather it serves as a helpful simplifying device that leads to a useful benchmark.

To gather some intuition for why Assumption 3a might be a reasonable benchmark, note that "enrichment activities", are, by their very definition, geared towards the improvement of skills somehow. Moreover, note from Table 5 that the vast majority of time spent on enrichment in the data is on activities that are more directed towards enhancing *cognitive* skills, such as homework.

Table 5 also helps motivate Assumption 3b. It shows that there are important differences in the composition of enrichment activities for high-school age children compared to younger ages. Older children, particularly in high school, spend less time on activities that tend to be self-directed or social, such as "reading a book" and "before/after school programs", which might be expected to have direct effects on both cognitive and non-cognitive skills. Instead, their enrichment time is tilted heavily towards homework, which might be expected to provide a direct effect on cognitive skills and a smaller (if any) direct effect on non-cognitive skills. We speculate that these composition differences may contribute to a divergence around high school between the marginal return of enrichment with respect to cognitive versus non-cognitive skills.[30]

If enrichment time is chosen to maximize cognitive skills, then at an interior solution I_c, the marginal skill return to optimally chosen enrichment, I_c, and leisure, $L_c = 1 - I_c$, will be equal:

$$\frac{\partial f_c(I_c, L_c)}{\partial I} = \frac{\partial f_c(I_c, L_c)}{\partial L}. \tag{7}$$

As shown in Eq. (2) (in Section 2), our cognitive estimates correspond in this model to the total derivative df_c/dI, which, by the above first-order condition, is equal to 0. In particular, noting that $\frac{\partial L}{\partial I} = -1$ because $L + I$ is constant,

$$\begin{aligned}\frac{df_c(I_c, L_c)}{dI} &= \frac{\partial f_c(I_c, L_c)}{\partial I} + \frac{\partial L}{\partial I}\frac{\partial f_c(I_c, L_c)}{\partial L} \\ &= \frac{\partial f_c(I_c, L_c)}{\partial I} - \frac{\partial f_c(I_c, L_c)}{\partial L} = 0.\end{aligned}$$

Thus, this simple model predicts zero total cognitive effects, approximately consistent with what we find across all age groups.[31]

Turning to non-cognitive skills, it is immediate that the corresponding first-order condition equating the marginal returns to I and L for non-cognitive skills will not generally hold at (I_c, L_c), the optimal levels for *cognitive skills*. This happens because one generally cannot maximize cognitive and non-cognitive skills at the same time when choosing the allocation of time I. Thus, under Assumption 3a alone, we can conclude that we should not expect to find near-zero non-cognitive estimates. Moreover, under Assumption 3b, it follows that, in high school,

$$\frac{df_n(I_c, L_c)}{dI} < \frac{df_c(I_c, L_c)}{dI} = 0.$$

This inequality implies that at the level of enrichment that maximizes cognitive skills, the total derivative $df_n(I_c, L_c)/dI < 0$, which is again consistent with what we find empirically.[32] In contrast, enrichment activities in earlier grades might not be geared so heavily towards cognitive skills, so $f_c(\cdot)$ and $f_n(\cdot)$ might be sufficiently similar to each other that $f_n(I_c, L_c)$ would approximate zero.[33]

This stylized model thus predicts that in cases where enrichment I is chosen to maximize cognitive skills, we should expect to see near-zero cognitive estimates and, depending on the extent to which I targets cognitive vs. non-cognitive skills (which may change across grades), non-cognitive estimates might be different from zero, and likely negative in high school.

As discussed in Section 2, the specific activity replaced by enrichment also matters. If the last hour of enrichment is obtained by spending one less hour on an activity that would have been beneficial for non-cognitive skills, such as sleeping or socializing with friends, then the negative non-cognitive effect would likely be greater. In this regard, we note that teenagers in high school seem to be spending substantially more time on duties and chores than younger children, as shown in Table 1. We conjecture that duties/chores is a special type of activity, in the sense that the time a child or teenager spends on it is more likely to be imposed by others, rather than chosen by them (e.g., perhaps their parents or their employer impose restrictions on the amount of duties/chores they need to accomplish on a day). If that is the case, then the fact that teenagers spend substantially more time on duties/chores than younger children may imply that they would have a higher opportunity cost for the last hour of enrichment, since they would have effectively less time available, after performing obligatory duties and chores, to allocate among the remaining activities.

These explanations are speculative, and the motivating model is very simple and omits many factors likely to be important. Making further progress in understanding the possibly heterogeneous opportunity costs of enrichment would involve estimating the additional causal effects detailed in Eq. (2).

9. Conclusion

Our results suggest that the sizable, positive correlations observed between enrichment time and childhood skills are mostly driven by unobservables. Using our control function approach to correct for the bias introduced by these unobservables, we find that the net causal effect of enrichment activities is around zero for cognitive skills. For non-cognitive skills, the corrected estimates are quite negative and very significant in high school, while closer to zero for earlier grades.

Our results suggest a number of potential avenues for future research. For example, prior literature has emphasized the importance of children's time with parents/adults for skill development, particularly in early childhood (Becker, 1991; Caetano et al., 2019; Hsin & Felfe, 2014; Linver, Brooks-Gunn, & Kohen, 2002; Yeung, Linver, & Brooks-Gunn, 2002). As demonstrated in Section 8, enrichment time for older

[30] This comparison is likely to understate the true disparity, since the nature of the homework across grades is different as well, with high school homework likely being less associated with non-cognitive skills than homework in earlier grades.

[31] For observations at a "corner solution" $I_c = 0$, $\partial f_c(0,1)/\partial I < \partial f_c(0,1)/\partial L$. This implies that the total derivative df_c/dI at the optimum would actually be negative for such individuals. Aggregating all individual effects df_c/dI across everyone at both interior and corner solutions, the average total effect should be negative but small, with magnitude depending on the probability of bunching.

[32] Although not necessary, for clarity one may make the standard assumptions that f_c and f_n are increasing and concave in I, so that $I_n < I_c$, where I_n is the level of I that maximizes non-cognitive skills.

[33] During grades K-5, enrichment activities might be geared predominantly towards non-cognitive skills, which would explain why non-cognitive point estimates, although noisy, are positive in those grades.

children is more oriented towards homework and other activities that are likely to be solitary. Thus, exploring heterogeneity in the effects of enrichment time by who the time is spent with (i.e. parents, other caregivers, friends, or alone) might prove fruitful. It would also be interesting to estimate the effects of enrichment time, particularly in high school, on alternative, longer-run outcomes.

As discussed in Sections 2 and 8, a comprehensive assessment of the mechanisms driving our results would require the estimation of the (likely heterogeneous) time use substitution patterns. Currently, these questions can only be studied with two datasets, both with limited sample sizes (the CDS-PSID and the Longitudinal Study of Australian Children—LSAC), which restrict the researcher's ability to uncover important sources of heterogeneity. We hope larger datasets on time use linked to measures of cognitive and non-cognitive skills become available, which will enable researchers to study these important questions.

Data availability

Data will be made available on request.

Acknowledgment

All authors approved the final version of the manuscript.

Appendix A. Details of the estimation of $\mathbb{E}[I^*|I = 0, X]$

This appendix provides the formulas of the three strategies to estimate $\mathbb{E}[I^*|I = 0, X]$ presented in Section 4.

Homoskedastic tobit

In this case, the formula of the expectation is

$$\mathbb{E}[I^*|I = 0, X] = X'\kappa - \sigma\lambda(-X'\kappa/\sigma) \quad (8)$$

where $\kappa = \pi + \theta$ and $\lambda(\cdot)$ is the inverse Mill's ratio. The parameters κ and σ are estimated straightforwardly via a Tobit regression of I on X.

Semiparametric Tobit

In this more general case, the formula of the expectation is

$$\mathbb{E}[I^*|I = 0, X] = \psi(X) - \sigma(X)\lambda\left(-\psi(X)/\sigma(X)\right), \quad (9)$$

where $\psi(X) = g(X) + l(X)$. We estimate $\psi(X)$ and $\sigma(X)$ separately for each cluster of X by running a Tobit regression of I on a constant using only the observations in that cluster. See Section 4.3 for details about clustering.

Nonparametric tail symmetry

Under tail symmetry, the formula is

$$\mathbb{E}[I^*|I = 0, X] = F^{-1}_{I|X}(1 - F_{I|X}(0)) - \mathbb{E}[I|I \geq F^{-1}_{I|X}(1 - F_{I|X}(0)), X], \quad (10)$$

where $F_{I|X}(\cdot)$ is the cumulative distribution function of I conditional on X. We carry out the estimation of $\mathbb{E}[I^*|I = 0, X]$ using Eq. (10) in three steps. For each cluster of X, we first estimate the probability of bunching at zero enrichment, $F_{I|X}(0)$. Then we estimate the quantile of I in the upper tail that corresponds to the mirror image of $I = 0$, $F^{-1}_{I|X}(1 - F_{I|X}(0))$. Finally we estimate the mean of $I|X$ at the upper tail, $\mathbb{E}[I|I \geq F^{-1}_{I|X}(1 - F_{I|X}(0)), X]$. See Section 4.3 for details about clustering.

Appendix B. Supporting tables and figures

Table B.6
The composition of various time use aggregates in the PSID-CDS.

Other enrichment	Share	Passive leisure	Share
Sports (Structured)	0.61	TV	0.67
Arts	0.17	Other media	0.30
Arts excursions	0.11	Other	0.03
Computer (Educational)	0.09		
Volunteering	0.01		
Play and social activities	**Share**	**Duties/Chores**	**Share**
Play and interactive games	0.44	Traveling	0.28
Conversations	0.17	Meals	0.27
Socializing	0.13	Personal care	0.21
Religious activities	0.12	Chores	0.11
Hobbies	0.06	Shopping	0.06
Sports (Unstructured)	0.05	Paid work	0.06
Other group activities	0.03	Caring for others	0.01

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

Table B.7
Cognitive and non-cognitive factor loadings.

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

Note: Cognitive and non-cognitive factor loadings are shown for each CDS wave. Each behavioral variable has been recoded so that a higher value corresponds to "better" behavior.

Table B.8
Uncorrected and corrected results — Alternative skill measures.

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Homosk. Tobit	(iv) Semip. Tobit	(v) Nonp. Tail symmetry
Cognitive						
Applied problems	β	0.013**	0.008**	0.000	−0.004	0.000
		(0.003)	(0.002)	(0.006)	(0.005)	(0.006)
	δ			0.007	0.010**	0.007
				(0.005)	(0.004)	(0.005)
Letter word	β	0.012**	0.008**	−0.001	−0.003	−0.001
		(0.003)	(0.001)	(0.006)	(0.005)	(0.006)
	δ			0.007	0.009**	0.007
				(0.005)	(0.004)	(0.005)
Passage comprehension	β	0.014**	0.009**	−0.001	−0.003	0.000
		(0.003)	(0.001)	(0.006)	(0.005)	(0.006)
	δ			0.009*	0.010**	0.008*
				(0.005)	(0.004)	(0.005)
Non-cognitive						
External	β	0.010**	0.006**	−0.014	−0.023**	−0.017*
		(0.002)	(0.002)	(0.009)	(0.008)	(0.009)
	δ			0.016**	0.023**	0.018**
				(0.008)	(0.007)	(0.008)
Internal	β	0.002	−0.001	−0.021**	−0.026**	−0.017*
		(0.003)	(0.003)	(0.010)	(0.009)	(0.010)
	δ			0.016**	0.020**	0.013
				(0.008)	(0.007)	(0.008)

Note: N = 4330. Bootstrapped standard errors in parentheses (500 iterations). Specifications (ii)–(v) use $K = 50$ and specifications (iii)–(v) use $K_\delta = 1$, where K and K_δ refer to the total number of clusters defined in Section 4.3. Both the external and internal behavioral indexes are scaled so that a higher score corresponds to better behavior. ** $p < 0.05$, * $p < 0.1$.

Table B.9
Uncorrected and corrected results — Broad enrichment.

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

Note: N = 4330. Bootstrapped standard errors in parentheses (500 iterations). Specifications (ii)–(v) use $K = 50$ and specifications (iii)–(v) use $K_\delta = 1$, where K and K_δ refer to the total number of clusters defined in Section 4.3. ** $p < 0.05$, * $p < 0.1$.

Table B.10
Uncorrected and corrected results — Homework only.

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Homosk. Tobit	(iv) Semip. Tobit	(v) Nonp. Tail symmetry
Cognitive	β	0.032**	0.010**	0.007	0.002	0.003
		(0.003)	(0.002)	(0.007)	(0.007)	(0.008)
	δ			0.003	0.006	0.006
				(0.005)	(0.005)	(0.006)
Non-cognitive	β	0.011**	0.006**	−0.020*	−0.032**	−0.029**
		(0.003)	(0.003)	(0.011)	(0.010)	(0.013)
	δ			0.020**	0.028**	0.028**
				(0.008)	(0.007)	(0.010)

Note: N = 4330. Bootstrapped standard errors in parentheses (500 iterations). Specifications (ii)–(v) use $K = 50$ and specifications (iii)–(v) use $K_\delta = 1$, where K and K_δ refer to the total number of clusters defined in Section 4.3. ** $p < 0.05$, * $p < 0.1$.

Table B.11
Alternative cognitive skill measures — Homework only.

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Tobit	(iv) Het. Tobit	(v) Het. Symmetric
Applied problems	β	0.032**	0.010**	0.016**	0.008	0.008
		(0.003)	(0.002)	(0.006)	(0.006)	(0.007)
	δ			−0.004	0.001	0.001
				(0.005)	(0.004)	(0.006)
Letter word	β	0.030**	0.007**	0.010	0.005	0.004
		(0.003)	(0.002)	(0.006)	(0.006)	(0.008)
	δ			−0.002	0.002	0.003
				(0.005)	(0.004)	(0.006)
Passage comprehension	β	0.031**	0.009**	0.005	0.002	0.001
		(0.004)	(0.002)	(0.006)	(0.005)	(0.006)
	δ			0.003	0.005	0.006
				(0.005)	(0.004)	(0.005)

Note: N = 4330. Bootstrapped standard errors in parentheses (500 iterations). Specifications (ii)–(v) use $K = 50$ and specifications (iii)–(v) use $K_\delta = 1$, where K and K_δ refer to the total number of clusters defined in Section 4.3. ** $p < 0.05$, * $p < 0.1$.

Table B.12
Adding lagged scores as controls: The effect of enrichment time on skills.

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Homosk. Tobit	(iv) Semip. Tobit	(v) Nonp. Tail symmetry
Cognitive	β	0.018**	0.006**	−0.003	−0.001	0.000
		(0.003)	(0.002)	(0.007)	(0.006)	(0.007)
	δ			0.007	0.006	0.005
				(0.006)	(0.005)	(0.006)
Non-cognitive	β	0.006**	0.001	−0.012	−0.013	−0.012
		(0.002)	(0.003)	(0.012)	(0.010)	(0.012)
	δ			0.010	0.011	0.010
				(0.009)	(0.008)	(0.010)

Note: N = 1572. Bootstrapped standard errors in parentheses (500 iterations). Columns (ii)–(v) use all control variables discussed at the end of Section 4, plus lagged cognitive and non-cognitive scores. Columns (iii)–(v) makes two identifying assumptions: Eq. (5) and a distributional assumption on $I^*|X$. Column (iii) assumes homoskedastic Tobit, column (iv) assumes semiparametric Tobit, and column (v) assumes nonparametric tail symmetry. Section 6 tests these two assumptions. Specifications (ii)–(v) use $K = 50$ and specifications (iii)–(v) use $K_\delta = 1$, where K and K_δ refer to the total number of clusters defined in Section 4.3. ** $p < 0.05$, * $p < 0.1$.

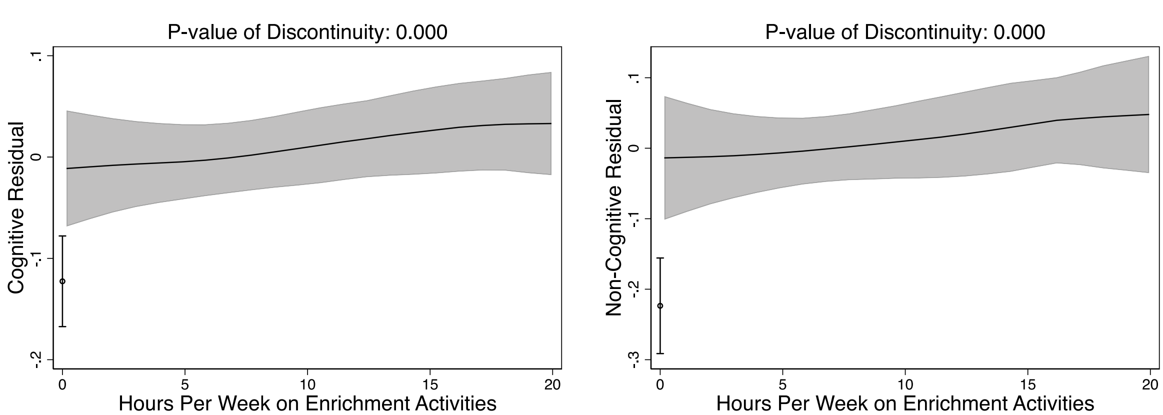

Fig. B.9. Two tests under no distributional assumption: Grades 9–12. Note: Each panel shows two tests, one at $I > 0$ and another at $I = 0$, both based on the expected value of the residuals from Eq. (3) (estimated for $I > 0$ only) conditional on I. We show a local linear plot of these residuals along with the 90% confidence interval. For $I > 0$, it is clear that the linearity cannot be rejected, so we cannot reject Assumption 1. For $I = 0$, it is clear that the average of the residuals is negative, which is equivalent to $\delta > 0$ (p-value of the test of whether $\delta = 0$ is also shown at the top). These two tests make no distributional assumption.

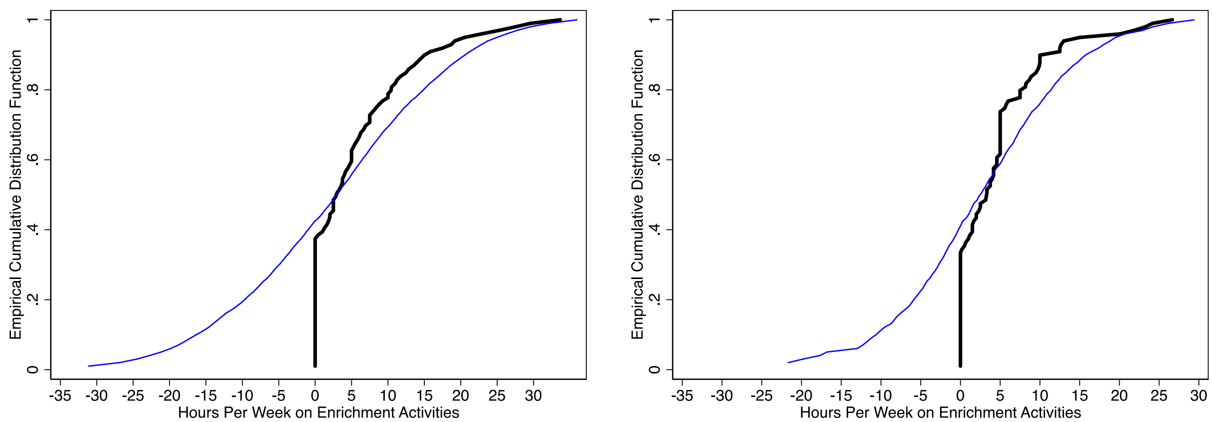

Fig. B.10. Homoskedastic tobit fit. Note: Each panel depicts the raw CDF of enrichment (I) for white (left panel) and Hispanic (right panel) high school students (thick curve) along with the corresponding homoskedastic Tobit fit (thin curve). The plots show evidence that the homoskedastic normal fit for positive values of enrichment is not satisfactory. We are able to reject this distributional assumption with a Kolmogorov–Smirnov test.

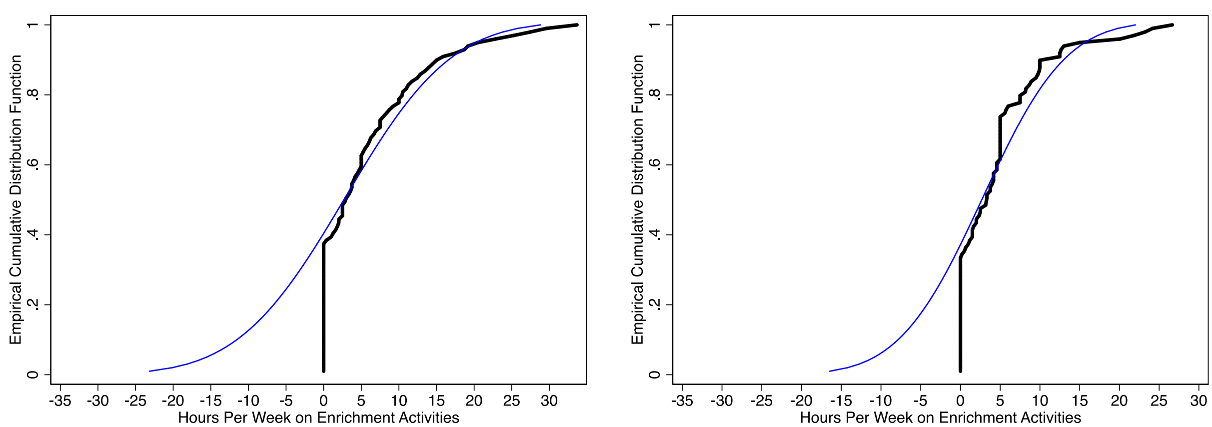

Fig. B.11. Semiparametric tobit fit. Each panel depicts the raw CDF of enrichment (I) for white (left panel) and Hispanic (right panel) high school students (thick curve) along with the corresponding semiparametric Tobit fit (thin curve). The fit improves relative to the homoskedastic case (Fig. B.10), but the upper tail of the raw distribution seems heavier than the fit. We are unable to reject this distributional assumption with a Kolmogorov–Smirnov test, but we can reject this assumption with a more powerful test (Goldman & Kaplan, 2018).

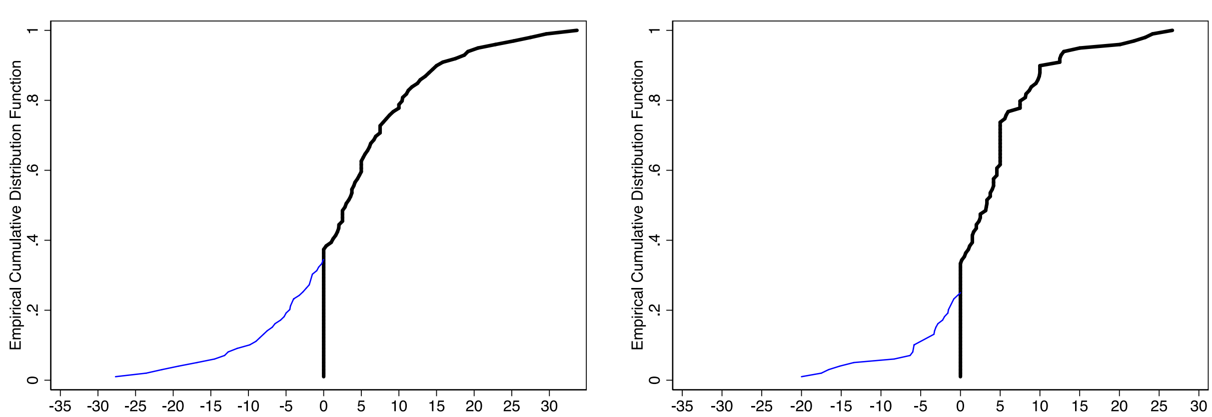

Fig. B.12. Nonparametric tail symmetry fit. Each panel depicts the raw CDF of enrichment (I) for white (left panel) and Hispanic (right panel) high school students (thick curve) along with the corresponding nonparametric tail symmetry fit (thin curve). We are unable to reject full symmetry with both distributional tests (Kolmogorov–Smirnov and Goldman & Kaplan, 2018).

References

Abeles, V. (2015). *Beyond measure, rescuing an overscheduled, overtested, underestimated generation*. New York, NY: Simon and Schuster.

Aguiar, M., & Hurst, E. (2007). Measuring trends in leisure: The allocation of time over five decades. *Quarterly Journal of Economics*, *122*, 969–1006.

Aksoy, T., & Link, C. R. (2000). A panel analysis of student mathematics achievement in the US in the 1990s: does increasing the amount of time in learning activities affect math achievement? *Economics of Education Review*, *19*(3), 261–277.

Almond, D., Chay, K. Y., & Lee, D. S. (2005). The costs of low birth weight. *Quarterly Journal of Economics*, *120*(3), 1031–1083.

Anderegg, D. (2003). *Worried all the time: rediscovering the joy in parenthood in an age of anxiety*. New York, NY: Free Press.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics*. Princeton University Press.

Avent, R. (2017). High-pressure parenting. *The Economist*, Feb/Mar 2017.

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

Becker, G. S. (1991). *A treatise on the family: enlarged edition*. Harvard University Press.

Bertanha, M., Caetano, C., Jales, H., & Seegert, N. (2022). Bunching designs: A guide to practice.

Bianchi, S. M. (2000). Maternal employment and time with children: Dramatic change or surprising continuity? *Demography*, *37*, 401–414.

Bound, J., Hershbein, B., & Long, B. T. (2009). Playing the admissions game: Student reactions to increasing college competition. *Journal of Economic Perspectives*, *23*(4), 119–146.

Caetano, C. (2015). A test of exogeneity without instrumental variables in models with bunching. *Econometrica*, *83*(4), 1581–1600.

Caetano, C., Caetano, G., & Nielsen, E. (2023). Correcting for endogeneity in models with bunching. *Journal of Business & Economic Statistics*, 1–13.

Caetano, C., Caetano, G., Nielsen, E., & Sanfelice, V. (2021). *The effect of maternal labor supply on children's skills*: *Working paper*, Available here.

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

Carneiro, P., Crawford, C., & Goodman, A. (2007). The impact of early cognitive and non-cognitive skills on later outcomes.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research*, *76*(1), 1–62.

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

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, *42*(3), 528–554.

Del Boca, D., Monfardini, C., & Nicoletti, C. (2017). Parental and child time investments and the cognitive development of adolescents. *Journal of Labor Economics*, *35*(2), 565–608.

Doepke, M., & Zilibotti, F. (2017). Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, *85*, 1331–1371.

Doepke, M., & Zilibotti, F. (2019). *Love, money, and parenting: how economics explains the way we raise our kids*. Princeton University Press.

Duncan, G., & Murnane, R. (2011). *Restoring opportunity: the crisis of inequality and the challenge for american education*. New York, NY: Russell Sage.

Eide, E. R., & Showalter, M. H. (2012). Sleep and student achievement. *Eastern Economic Journal*, *38*, 512–524.

Eren, O., & Henderson, D. J. (2011). Are we wasting our children's time by giving them more homework? *Economics of Education Review*, *30*(5), 950–961.

Fawzi, W. W., Chalmers, T. C., Herrera, M. G., & Mosteller, F. (1993). Vitamin A supplementation and child mortality: A meta-analysis. *JAMA*, *269*(7), 898–903.

Fiorini, M., & Keane, M. P. (2014). How the allocation of children's time affects cognitive and noncognitive development. *Journal of Labor Economics*, *32*(4), 787–836.

Galloway, M., Conner, J., & Pope, D. (2013). Nonacademic effects of homework in privileged, high-performing high schools. *The Journal of Experimental Education*, *81*(4), 490–510.

Ginsburg, K. R., et al. (2007). The importance of play in promoting healthy child development and maintaining strong parent-child bonds. *Pediatrics*, *119*(1), 182–191.

Goldman, M., & Kaplan, D. M. (2018). Comparing distributions by multiple testing across quantiles or CDF values. *Journal of Econometrics*, *206*(1), 143–166.

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

Gray, P. (2011). The decline of play and the rise of psychopathology in children and adolescents. *American Journal of Play*, *3*(4), 443–463.

Gray, P. (2013). *Free to learn: Why unleashing the instinct to play will make our children happier, more self reliant and better students for life*. New York, NY: Basic Books.

Groen, J. A., & Pabilonia, S. W. (2019). Snooze or lose: High school start times and academic achievement. *Economics of Education Review*, *72*, 204–218.

Guryan, J., Hurst, E., & Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic Perspectives*, *22*(3), 23–46.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482.

Heckman, J. J., Urzua, S., & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, *88*(3), 389–432.

Heissel, J. A., & Norris, S. (2018). Rise and shine. *Journal of Human Resources*, *53*(4), 957–992.

Hsin, A., & Felfe, C. (2014). When does time matter? Maternal employment, children's time with parents, and child development. *Demography*, *51*(5), 1867–1894.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072–2107.

Jarvis, P., Newman, S., & Swiniarski, L. (2014). On 'becoming social': The importance of collaborative free play in childhood. *International Journal of Play*, *3*(1), 53–68.

Jürges, H., & Khanam, R. (2021). Adolescents' time allocation and skill production. *Economics of Education Review*, *85*, Article 102178.

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

Lareau, A. (2003). *Unequal childhoods: class, race, and family life*. Berkeley, CA: University of California Press.

Lenard, M., Morrill, M. S., & Westall, J. (2020). High school start times and student achievement: Looking beyond test scores. *Economics of Education Review*, *76*, Article 101975.

Linver, M. R., Brooks-Gunn, J., & Kohen, D. E. (2002). Family processes as pathways from income to young children's development.. *Developmental Psychology*, *38*(5), 719.

Lukianoff, G., & Haidt, J. (2018). *The coddling of the american mind: how good intentions and bad ideas are setting up a generation for failure*. Penguin.

Luthar, S. S. (2003). The culture of affluence: Psychological costs of material wealth. *Child Development*, *74*(6), 1581–1593.

Luthar, S. S., & Becker, B. E. (2002). Privileged but pressured? A study of affluent youth. *Child Development*, *73*(5), 1593–1610.

Millimet, D. L., & Tchernis, R. (2013). Estimation of treatment effects without an exclusion restriction: With an application to the analysis of the school breakfast program. *Journal of Applied Econometrics*, *28*(6), 982–1017.

Oken, E., Levitan, E., & Gillman, M. (2008). Maternal smoking during pregnancy and child overweight: Systematic review and meta-analysis. *International Journal of Obesity*, *32*(2), 201–210.

Peterson, J., & Zill, N. (1986). Marital disruption, parent-child relationships, and behavioral problems in children. *Journal of Marriage and the Family*, *48*, 295–307.

Ramey, G., & Ramey, V. A. (2010). The rug rat race. *Brookings Papers on Economic Activity*, *2010*(1), 129–176.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *31*(2), 350–371.

Rønning, M. (2011). Who benefits from homework assignments? *Economics of Education Review*, *30*(1), 55–64.

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

Rosenfeld, A. A., & Wise, N. (2000). *The over-scheduled child: avoiding the hyper-parenting trap*. New York, NY: St Martin's Griffin.

Rosin, H. (2015). The silicon valley suicides. *The Atlantic*.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*(485), F3–F33.

Todd, P., & Wolpin, K. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*, *1*(1), 91–136.

Veiga, G., Neto, C., & Rieffe, C. (2016). Preschoolers' free play: Connections with emotional and social functioning. *International Journal of Emotional Education*, *8*(1), 48–62.

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

Warner, J. (2005). *Perfect madness: motherhood in the age of anxiety*. New York, NY: Riverhead Books.

Weisburst, E. K. (2019). Patrolling public schools: The impact of funding for school police on student discipline and long-term education outcomes. *Journal of Policy Analysis and Management*, *38*(2), 338–365.

Wolfson, A. R., Carskadon, M. A., Acebo, C., Seifer, R., Fallone, G., Labyak, S. E., et al. (2003). Evidence for the validity of a sleep habits survey for adolescents. *Sleep*, *26*(2), 213–216.

Yeung, W. J., Linver, M. R., & Brooks-Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child Development*, *73*(6), 1861–1879.

Yitzhaki, S. (1996). On using linear regressions in welfare economics. *Journal of Business & Economic Statistics*, *14*(4), 478–486.