Is Video Watching Bad for Kids? The Effect of Video Watching on Children's Skills*

Carolina Caetano† Gregorio Caetano‡ Débora Mazetto§
Meghan M. Skira¶

August 2026
Latest Version

Abstract

This paper documents video consumption among school-aged children in the U.S. and explores its impact on human capital development. Video watching is common across all segments of society, yet surprisingly little is known about its developmental consequences. With a bunching identification strategy, we find that an additional hour of daily video consumption has a negative impact on children's noncognitive skills, with harmful effects on both internalizing behaviors (e.g., depression) and externalizing behaviors (e.g., social difficulties). We find a positive effect on quantitative reasoning, though the effect on an aggregate measure of cognitive skills is smaller and not statistically significant. These findings are robust and largely stable across most demographics and different ways of measuring skills and video watching. We find evidence that for Hispanic children, video watching has positive effects on both cognitive and noncognitive skills—potentially reflecting its role in supporting cultural assimilation. Interestingly, the marginal effects of video watching remain relatively stable regardless of how much time children spend on the activity, with similar incremental impacts observed among those who watch very little and those who watch for many hours.

Keywords: video watching, TV watching, time use, cognitive skills, noncognitive skills, bunching.
JEL Classification Codes: D12, I20, I21, J13, C24.

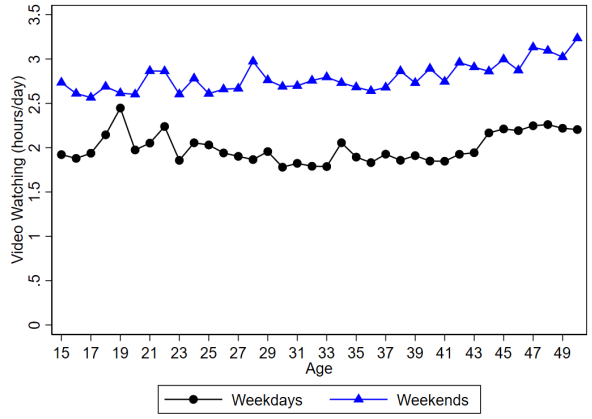
*We would like to thank Eric Nielsen, Viviane Sanfelice, Kelly Santos and seminar participants for their valuable comments and suggestions. All remaining errors are ours.
†University of Georgia; email: carol.caetano@uga.edu.
‡University of Georgia; email: gregorio.caetano@uga.edu.
§Tennessee Tech University; email: dmazetto@tntech.edu.
¶University of Georgia and NBER; email: skira@uga.edu.

1 Introduction

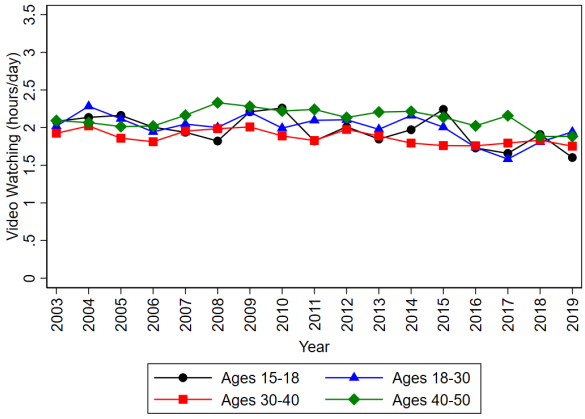
Video watching, whether through TV or more modern devices, is a common activity in nearly every household. As shown in the left panel of Figure 1, according to the American Time Use Survey (ATUS), Americans aged 15–50 tend to spend at least 2 hours a day watching some form of video content. Although they watch more on weekends, they still watch a significant amount during the week, when people tend to have busier schedules. What is particularly striking is the stability of this pattern across time and age cohorts, as shown in the right panel of Figure 1 for weekday video watching.

Figure 1: Trends in Video Watching by Age and Year (ATUS)

(a) By Age



(b) By Year, Weekday



Notes: Panel (a) shows average daily weekday and weekend video watching by age. Panel (b) shows average daily weekday video watching by year across age cohorts. Sample weights applied. Source: American Time Use Survey (ATUS) from 2003–2019 ($N = 131,575$).

Despite its prevalence and universality, there is little understanding of how early this video watching habit takes shape and the developmental consequences it may engender. We fill this gap by using the Child Development Supplement of the Panel Study of Income Dynamics (CDS-PSID) to first document the video watching behavior of school-aged children (ages 6–18) in the United States and then estimate the causal effects of video watching on children's skills.[1] In terms of descriptive patterns, our most striking finding is that video watching is generally stable across demographics such as gender, age, race/ethnicity, household income, and number of siblings, with the average child across these groups watching around 1.5 hours daily on weekdays and 2.7 hours daily on weekends.

After documenting these patterns, we estimate the causal effect of video watching on chil-

[1]By video watching, we mean watching a television broadcast, show, video, or movie via a television set, VHS or DVD player, computer, tablet, or phone. We define video watching in more detail in Section 2.

dren's development, focusing on a range of cognitive and noncognitive skill measures. These early-life skills are well-established predictors of long-term outcomes, including educational attainment, labor market success, and broader indicators of social well-being (Heckman, Stixrud, and Urzua 2006). Moreover, cognitive and noncognitive skills are known to accumulate and reinforce one another over time, amplifying the effects of early differences. As such, video watching behavior during childhood may influence the process of human capital formation directly—beyond any role it may play in shaping video consumption habits later in life.

The effect of video watching on skill development is unclear, given the many potential sources of heterogeneity across children. The impact may depend not only on the specific skill in question, but also on the nature of the video content and the activity displaced by video watching—factors likely to vary with each child's circumstances. While the American Academy of Pediatrics (AAP) has historically recommended screen time limits for school-aged children, more recently they have suggested a decentralized approach, emphasizing parental involvement and prioritizing content quality without issuing concrete suggestions (CCM 2016 and AAP 2025). This approach implicitly recognizes that the optimal "video diet" may vary by family. Yet, to our knowledge, there is a lack of causal evidence about the effects of video watching, especially evidence offering heterogeneous guidance that might help parents navigate this decision. Given the remarkable stability of video watching patterns over the past several decades, the absence of nuanced recommendations for this widespread behavior is striking, and it reflects the broader challenge of obtaining credible causal estimates of the effects of behavioral activities like video consumption.

Using a bunching identification strategy, we find that an additional hour of video consumption has a negative impact on children's noncognitive skills. For every additional daily hour spent watching video, the child experiences a reduction in their noncognitive skill of 0.091 standard deviations on average—with harmful effects on both internalizing behaviors (e.g., depression) and externalizing behaviors (e.g., social difficulties). We find a positive, though small and insignificant, effect on an aggregate measure of cognitive skill, and a statistically significant beneficial effect on quantitative reasoning (i.e., applied problems) in particular.

We examine effect heterogeneity across many dimensions, including children's grade in school, gender, race/ethnicity, family income, number of siblings, and the number of hours watching video. Our findings are robust and stable across most of the dimensions of heterogeneity we consider and across different ways of measuring skills. Although detecting statistically significant effect heterogeneity is challenging given the limited sample sizes, some patterns stand out. In particular, we find positive effects on both cognitive and noncogni-

tive skills for Hispanic children—consistent with the idea that video watching may aid in cultural assimilation (Dalisay 2012 and Li and Tsai 2015). We also find suggestive evidence that the adverse effects of video watching on noncognitive skills are more pronounced among middle-income children. This non-monotonic pattern may reflect heterogeneous trade-offs shaped by opportunities and constraints: for middle-income children, video watching may be more likely to crowd out activities that would have fostered socioemotional development compared to low-income children, but they may lack the resources relative to high-income children (e.g., parental time, guidance, or professional support) to offset any resulting behavioral costs (Lareau 2003 and Mollborn, Limburg, Pace, and Fomby 2022).

Another key finding is the absence of nonlinearities in the estimated effects of video watching. Historically, many pediatric guidelines and public health narratives caution against excessive screen time, implicitly suggesting that its marginal effect becomes more harmful as exposure increases. However, in its current guidance, the AAP avoids recommending a specific cap on screen time for school-aged children, noting that there is insufficient evidence that strict limits provide benefits (AAP 2025). Our finding that the marginal harm of video watching does not increase with longer viewing hours aligns with theories of rational inattention (Sims 2003), endogenous time allocation and quality adjustment (Becker 1965), as well as self-regulation (Carver and Scheier 1998). That is, children might endogenously adjust behavioral margins such as attention allocation, content selection, and cognitive engagement when watching video for longer hours, and these behavioral adjustments may reduce the potentially harmful effects of intense video watching.

In estimating the effects of video watching on skill development, we focus primarily on weekday viewing, as it entails a higher opportunity cost by directly competing with structured activities such as schoolwork and sleep. The findings are robust to detailed sensitivity analyses along several dimensions, such as whether we include weekend video watching.

We isolate these causal effects by exploiting the concentration (bunching) of children spending exactly zero time on video watching. The key idea of this identification strategy is that the selection (confounding) effect is discontinuous at the bunching point, while the treatment effect is continuous. Therefore, discontinuities observed in the outcome at the bunching point can only be attributed to the selection effect, allowing us to isolate the bias due to the confounding effect and thus indirectly identify the treatment effect of interest. As this discussion suggests, the key untestable identification assumption is the continuity of the treatment effect. Essentially, we assume that spending a few minutes per day watching video does not cause an abrupt change in the child's skills, relative to a counterfactual of not spending any time on video watching. The qualitative conclusions reported above rely

only on this continuity assumption and the assumption that the *average direction* of the confounding effect is the same across two groups—those watching video and those not watching video. Although our key findings are nonparametrically identified, to obtain our preferred quantitative estimates, we additionally impose testable assumptions on the confounding effect. We provide diagnostic evidence supporting these assumptions and a detailed sensitivity analysis to assess how deviations from these assumptions affect the reported estimates. We conclude that there is little scope for the negative (positive) effects of video watching on noncognitive (cognitive) skills to be less intense than reported, and a lot more scope for them to be more intense than reported.

Beyond documenting the patterns of video watching in the U.S., this paper contributes to the literature in three important ways. First, we estimate average treatment effects of time spent watching video, as opposed to intention-to-treat effects. We observe the treatment variable—the actual hours children spend watching video—whereas most of this literature does not. Instead, prior studies typically estimate intention-to-treat effects of video watching by exploiting variation in TV signal access or reception (Gentzkow and Shapiro, 2008; Hernæs, Markussen, and Røed, 2019; Nieto and Suhrcke, 2021) or in the broadcasting of specific shows (Kearney and Levine, 2015; Trudeau, 2016; Kearney and Levine, 2019; Jaeger, Joyce, and Kaestner, 2020). An exception is Huang and Lee (2010), which investigates how daily time spent watching TV at ages 6–7 and 8–9 (as reported by mothers) affects cognitive development at ages 8–9 by estimating dynamic panel data models using the 1979 National Longitudinal Survey of Youth. This distinction is important because average treatment effects of time spent watching video are objects of interest for parents navigating decisions around their children's screen time as well as policymakers designing guidance around this activity. In contrast, intention-to-treat effects likely encapsulate effects due to other changes in video watching behavior, beyond video watching time *per se*. Moreover, intention-to-treat analyses cannot assess the presence of nonlinearities in the estimated effects of video watching—often implicitly assumed in video watching guidelines—because identifying nonlinear effects requires observing time spent watching video.

Second, we estimate these average treatment effects on a variety of measures of cognitive and noncognitive skills. Most of the literature has focused on children's cognitive development (Gentzkow and Shapiro, 2008; Huang and Lee, 2010; Kearney and Levine, 2019; Hernæs et al., 2019) and the findings are mixed. Some studies explore other outcomes, such as teenage pregnancy (Kearney and Levine, 2015; Trudeau, 2016; Jaeger et al., 2020), but rarely socioemotional development.[2] An exception is Nieto and Suhrcke (2021), which lever-

[2]We focus our discussion of the literature on the most closely related studies that estimate effects of video watching on children's outcomes. There is a large and growing related literature examining the causal

ages the digital television transition to examine the intention-to-treat effect of gaining access to more TV channels on children's obesity and mental well-being in the United Kingdom. They find that access to digital television adversely affected children's mental health. Using United Kingdom data as well, Meroni, Piazzalunga, and Pronzato (2022) investigate the effect of video watching frequency (which is likely correlated with video watching time) on the noncognitive skill of children aged 7–11. With value-added models, they find that frequent video watching, as measured by a binary indicator for watching at least one hour per day, has harmful effects on socioemotional outcomes. The relative lack of focus on noncognitive skills stands out in light of the fact that most of the recommendations about video watching relate to potential detrimental effects on noncognitive skills (see Hill et al. 2016 and CCM 2016). To the best of our knowledge this is the first paper estimating average treatment effects of video watching on noncognitive skills using U.S. data.

Third, we exploit a more representative source of variation in video watching than prior studies, which often focus on events affecting specific subpopulations (e.g., particular demographic groups, time periods, or locations) or on the broadcasting of particular shows. Our approach captures variation that is broader and more generalizable, allowing us to estimate heterogeneous effects of video watching along many dimensions that, to our knowledge, have not been explored in causal studies.

The rest of the paper is organized as follows. Section 2 discusses the data and documents the patterns of video watching time among school-aged children. Section 3 presents our identification strategy to estimate the causal effect of video watching on skills. Our results are presented in Section 4, along with robustness analyses that relax our key assumptions. Section 5 reports heterogeneous results across different measures of skills, characteristics of the children and their families, and video watching hours. Conclusions and discussions are presented in Section 6.

2 Data

We use the Child Development Supplement (CDS) from the Panel Study of Income Dynamics (PSID) (Survey Research Center, 2019). The PSID is a nationally representative longitudinal household survey of families and individuals in the United States. The CDS-PSID provides detailed time diary data on children's daily activities and measures of their cognitive and noncognitive skills. It also contains rich information on child and family characteristics.

effects of internet access and media use more broadly on children's cognitive and noncognitive skill. See for example, Fiorini and Keane (2014); Vigdor, Ladd, and Martinez (2014); Colombo and Failache (2023); Colombo, Failache, and Querejeta (2025) among many others.

In 1997, the CDS-PSID surveyed about 3,600 children aged 0 to 13 in PSID families, and followed those same children up to age 18 in 2002 and 2007. In 2014 and 2019, the CDS-PSID included all eligible children in the household aged 0 to 17. We pool the data from the 1997, 2002, 2007, 2014, and 2019 waves, focusing on children aged 6 to 18 years old in kindergarten through 12th grade.[3] The CDS-PSID requires families to fill out a time diary detailing children's activities on a random weekday and a random weekend day.[4] With the help of the interviewer, the child (or the parent, if the child is too young) records each activity in the diary, selecting from a list of over 300 types of activities, and notes the start and end times, participants, and location. If a child does more than one activity at the same time, they indicate which activity was primary (i.e., the main activity) and which was secondary. We exclude diaries where the weekday or weekend diary day was considered atypical,[5] those with missing weekday or weekend data, and diaries not covering the full 24 hours, in line with previous research.[6] Our noncognitive skill analysis sample consists of 5,706 child-wave observations (which include 4,223 unique children) and our cognitive skill analysis sample consists of 5,274 child-wave observations (which include 3,891 unique children).

Our main focus is the impact of time spent watching video on a typical weekday, where video watching is the primary activity.[7] In the 1997, 2002, and 2007 waves of the CDS-PSID, the pertaining activity code is described as "TV," and it includes watching television programs (live or time-shifted), video tapes, and DVDs. In the 2014 and 2019 waves, the pertaining activity code is described as "watching a show, video, or movie." It still includes watching television programs, video tapes, and DVDs, but additionally includes time spent watching shows, videos, or movies on devices such as computers, phones, and tablets. We focus on video watching during weekdays because it accounts for a large share of the time available for discretionary activities, as the majority of children's daily weekday time is allocated to activities with limited choice, such as sleep and schoolwork. In robustness exercises, we examine effects of daily video watching during the weekend and the whole

[3] We focus on this age range because the cognitive and noncognitive skill measures are consistently available for children aged 6 to 18.

[4] In 2014, only 50% of eligible CDS children were assigned time diaries.

[5] Respondents are asked how typical each diary day was, with responses ranging from 1 ("very typical") to 5 ("not at all typical"). We define a diary day as atypical if it is described as "not at all typical."

[6] We also follow the standard practice in this literature of recording missing diary data from 10 p.m. to 6 a.m. as "sleeping" (Fiorini and Keane, 2014; Caetano, Kinsler, and Teng, 2019).

[7] We do not use secondary activity information because when no secondary activity is reported in the data, this can either mean that there was no secondary activity or that the child (or primary caregiver) filling out the time diary did not know whether a secondary activity took place. The data do not allow us to distinguish these two cases.

week.[8,9]

In our descriptive analyses, we document a variety of video watching patterns using the rich detail available in the time diaries. Given the time diary includes information about with whom the child did the activity, we report whether children watched video alone or with others.[10] Moreover, sometimes when children report watching video as an activity, they also record the title(s) of the television shows, movies, and videos they watched. The CDS-PSID reviews and codes the titles to describe each program's content, nature, and target audience and releases this information via their Time Diary Media Codes Files. We use information from these files to categorize the genre and age appropriateness of the content watched.[11] We also report the average hours of video watched with missing data on these attributes, as the missingness is non-trivial.

We start by reporting the patterns of video watching in the sample used for estimating the effects of video watching on noncognitive skill,[12] and then discuss the variables we use in our analyses. We use sample weights when calculating the summary statistics below in order to provide representative patterns among the U.S. population.

2.1 Time Watching Video

By Characteristics of Video Watching Behavior

Table 1 presents key summary statistics related to video watching, measured in hours per day. The first row reports average daily hours of video watching during a typical week,

[8]To construct average daily video watching across the whole week, we calculate weekday video watching on a typical day multiplied by 5 plus weekend video watching during a typical weekend day multiplied by 2. We then divide that sum by 7.

[9]It is for this reason that we exclude those with atypical weekend diaries from our main analysis samples. In the robustness exercises that examine the effects of daily video watching during a typical weekend and typical week, we want to maintain the same sample as in our baseline analyses.

[10]When someone else was present but did not participate in the activity, we categorized that activity as spent "alone."

[11]The Time Diary Media Codes Files report about 20 genre categories, which we aggregate into the following groups: informative/educational, fictional entertainment, reality/variety entertainment, and missing. Informative/educational includes news, documentaries, children's education, do-it-yourself/hobbies, and nature/environmental content. Fictional entertainment includes daily relationships, case solving, horror/thriller, action/adventure content, and soap operas. Reality/variety entertainment includes entertainment news, reality shows, game shows, talk shows, variety, music videos/concerts, and sports. To determine age appropriateness, we compare the child's grade in school with information from the Time Diary Media Code Files about whether the intended audience was children (from pre-K through elementary school age), adolescents (middle and high school age), adults, general (i.e., designed for all audiences), or indeterminate. For children in elementary school, age-appropriate content is targeted at children and general audiences. For children in middle and high school, age-appropriate content is targeted at children, adolescents, and general audiences. Content targeted at adults is considered not age appropriate.

[12]The quantitative patterns discussed below are similar if we instead use the cognitive skill analysis sample.

weekday, and weekend, including those who watch zero hours during those periods. Most of the video watching happens during weekends, but there is a substantial amount of video watching during weekdays. On average, children spend about 1.5 hours per day on weekdays and 2.7 hours per day on weekends watching video.

During weekdays, video tends to be watched in the 3 p.m. to 8 p.m. window, while during the weekend, video watching is spread out more evenly across the day, with a large portion during the 6 a.m. to 3 p.m. window. Moreover, on average, about one-third of the hours spent watching video were done alone. The vast majority of content is watched using a television set, VHS player, or DVD player (versus a computer, tablet, phone, or other device). Almost no time is spent watching shows characterized by informative/educational content (e.g., news, child education, documentaries), and most of it is spent on entertainment, especially fictional entertainment (e.g., action or adventure shows or movies). Finally, about two-fifths of the content is deemed age appropriate.[13]

[13]Conditional on watching video on a weekday, children spend 1.9 hours per day doing so, on average. For these children, 2.6% of hours have missing information about the device used, 28.6% of hours have missing information about the genre of the content, and 22.4% of hours have missing information about the targeted audience of the content.

Table 1: Time Watching Video by Characteristics of Video Watching Behavior

Hours/day	Whole Week	Weekdays	Weekends
Video Watching	1.82 (1.44)	1.47 (1.52)	2.69 (2.24)
Period of the Day			
6 a.m. to 3 p.m.	0.47 (0.65)	0.21 (0.65)	1.12 (1.40)
3 p.m. to 8 p.m.	0.91 (0.90)	0.85 (1.06)	1.05 (1.28)
8 p.m. to 6 a.m.	0.44 (0.62)	0.41 (0.70)	0.52 (0.87)
Company			
Alone	0.65 (1.00)	0.56 (1.08)	0.88 (1.55)
Not alone	1.17 (1.15)	0.92 (1.21)	1.81 (1.89)
Missing	0.00 (0.04)	0.00 (0.06)	0.00 (0.05)
Device			
TV, VHS, DVD player	1.66 (1.42)	1.35 (1.49)	2.43 (2.20)
Computer, tablet, phone, and others	0.11 (0.45)	0.08 (0.46)	0.17 (0.78)
Missing	0.05 (0.24)	0.04 (0.25)	0.09 (0.46)
Show Genre			
Informative/educational	0.09 (0.27)	0.09 (0.31)	0.11 (0.42)
Fictional entertainment	0.92 (1.04)	0.76 (1.12)	1.32 (1.67)
Reality/variety entertainment	0.28 (0.57)	0.20 (0.57)	0.46 (1.10)
Missing	0.53 (0.88)	0.42 (0.90)	0.80 (1.43)
Targeted Audience			
Age appropriate	0.74 (0.88)	0.58 (0.93)	1.13 (1.51)
Not age appropriate	0.67 (0.98)	0.56 (1.05)	0.95 (1.50)
Missing	0.41 (0.78)	0.33 (0.81)	0.62 (1.25)

Notes: The sample includes 5,706 child-wave observations from the 1997, 2002, 2007, 2014, and 2019 CDS-PSID waves. Sample weights are applied and standard deviations are shown in parentheses.

By Characteristics of Children

Table 2 breaks down the video watching patterns by the characteristics of the child. Children in middle school watch video a bit more than those in high school, who tend to watch a bit more video than those in elementary school. The patterns of video watching are relatively stable across gender, with boys spending a bit more time watching video on weekends. Non-Hispanic Black and Hispanic children tend to watch more video than Non-Hispanic White children and children from other racial and ethnic groups.

Table 2: Time Watching Video by Characteristics of Children

Hours/day	Observations	Whole Week	Weekdays	Weekends
Total	5,706	1.82 (1.44)	1.47 (1.52)	2.69 (2.24)
Grade				
Elementary school	2,644	1.72 (1.26)	1.35 (1.31)	2.65 (2.01)
Middle school	1,508	1.97 (1.49)	1.62 (1.57)	2.83 (2.43)
High school	1,554	1.84 (1.64)	1.53 (1.76)	2.63 (2.40)
Gender				
Boys	2,847	1.85 (1.44)	1.47 (1.52)	2.79 (2.29)
Girls	2,859	1.80 (1.45)	1.48 (1.53)	2.59 (2.19)
Race/Ethnicity				
Non-Hispanic White	2,774	1.71 (1.38)	1.36 (1.47)	2.60 (2.18)
Non-Hispanic Black	2,252	2.14 (1.49)	1.74 (1.55)	3.16 (2.50)
Hispanic	426	2.06 (1.47)	1.77 (1.57)	2.80 (2.18)
Non-Hispanic other	227	1.79 (1.75)	1.53 (1.84)	2.43 (2.21)
Missing	27	1.09 (1.10)	0.79 (1.06)	1.84 (1.89)

Notes: The sample includes 5,706 child-wave observations from the 1997, 2002, 2007, 2014, and 2019 CDS-PSID waves. Sample weights are applied and standard deviations are shown in parentheses.

By Characteristics of Families

Table 3 breaks down the video watching patterns by the characteristics of the family. Children from lower-income families tend to spend more time watching video. Children with no biological siblings in the household tend to spend a bit less time on video. Moreover, children in households where the main PSID respondent completed the interview in a language other than English tend to watch more video than children from households where the respondent completed the survey in English.

Table 3: Time Watching Video by Characteristics of Families

Hours/day	Observations	Whole Week	Weekdays	Weekends
Total	5,706	1.82 (1.44)	1.47 (1.52)	2.69 (2.24)
Income Tercile				
Lower tercile	1,848	2.05 (1.50)	1.68 (1.60)	2.98 (2.39)
Middle tercile	1,853	1.90 (1.43)	1.55 (1.55)	2.78 (2.19)
Higher tercile	1,864	1.61 (1.38)	1.28 (1.44)	2.42 (2.13)
Missing	141	2.08 (1.54)	1.65 (1.47)	3.13 (2.54)
Number of Biological Siblings in Household				
No siblings	1,071	1.75 (1.39)	1.41 (1.47)	2.59 (2.23)
One sibling	2,434	1.83 (1.44)	1.46 (1.53)	2.74 (2.31)
Two siblings	1,489	1.82 (1.43)	1.49 (1.48)	2.64 (2.15)
Three or more siblings	712	1.94 (1.57)	1.60 (1.70)	2.78 (2.18)
Language of the Interview				
English	5,308	1.78 (1.43)	1.43 (1.51)	2.66 (2.23)
Other language	265	2.18 (1.51)	1.89 (1.60)	2.88 (2.22)
Missing	133	2.11 (1.56)	1.69 (1.48)	3.17 (2.59)

Notes: The sample includes 5,706 child-wave observations from the 1997, 2002, 2007, 2014, and 2019 CDS-PSID waves. Sample weights are applied and standard deviations are shown in parentheses.

By Characteristics of Child-Wave Observation

We also consider characteristics of the child-wave observation. Table 4 shows that video watching has declined in more recent waves, which may be partially explained by the fact that the age distribution of children in each wave is different and the availability of entertainment other than video watching increased. Because we pool observations from several waves, some children are observed more than once. About 63% of the child-wave observations are obtained from those observed more than once in the sample, although these children make up only 49% of the unique children in the sample (2,088 out of 4,223 unique children).

Table 4: Time Watching Video by Characteristics of Child-Wave Observation

Hours/day	Observations	Whole Week	Weekdays	Weekends
Total	5,706	1.82 (1.44)	1.47 (1.52)	2.69 (2.24)
Wave				
1997 (age range: 6 to 14)	1,070	1.82 (1.29)	1.45 (1.35)	2.73 (2.06)
2002 (age range: 6 to 18)	1,907	1.99 (1.49)	1.66 (1.62)	2.84 (2.23)
2007 (age range: 10 to 18)	1,111	1.90 (1.53)	1.54 (1.58)	2.80 (2.44)
2014 (age range: 6 to 17)	841	1.54 (1.30)	1.20 (1.37)	2.39 (2.14)
2019 (age range: 6 to 18)	777	1.54 (1.46)	1.21 (1.51)	2.38 (2.22)
Child-Wave Observations				
Child is observed once	2,135	1.73 (1.46)	1.40 (1.53)	2.55 (2.24)
Child is observed more than once	3,571	1.87 (1.43)	1.51 (1.52)	2.77 (2.24)

Notes: The sample includes 5,706 child-wave observations from the 1997, 2002, 2007, 2014, and 2019 CDS-PSID waves. Sample weights are applied and standard deviations are shown in parentheses.

Overall, the main takeaway from these patterns of video watching behavior is that children, on average, watch a non-trivial amount of video per day, irrespective of the way we break down the data. Furthermore, video watching is generally stable across demographics such as gender, age, race/ethnicity, household income, number of siblings, and interview language, with the average child watching around 1.5 hours daily on weekdays and 2.7 hours daily on weekends.

Video on Weekdays and Other Activities

We now examine weekday video watching in greater detail, since this is the treatment variable in our main analysis. Table 5 shows how the distribution of other activities differs depending on how many hours the child watches video during weekdays. We follow Caetano, Caetano, and Nielsen (2024a) in defining these other activities. Those who watch more video tend to spend less time in school-related activities (e.g., class time, homework), extra-curricular activities (e.g., structured sports, volunteering), active leisure activities (e.g., reading, hobbies), and daily routine activities (e.g., travel, personal care, meals).

Note that the standard deviation of the distribution of time spent on other activities (in parentheses) tends to be lower in the middle of the distribution of video watching hours than at the extremes, specifically at 0 hours of video watching (first column) and at 4 or more hours of video watching (last column). While the last column aggregates all instances of 4 hours or more into the same category, this does not apply to the first column, as watching negative hours of video is impossible. We will explore this idea in our identification strategy, discussed in Section 3.

Table 5: Distribution of Daily Activities

	Video Watching on Weekdays				
Activities (hours/day)	0 hours	0-1 hours	1-2 hours	2-4 hours	4+ hours
Extra-Curriculars	1.12 (1.40)	0.86 (1.26)	0.60 (1.00)	0.40 (0.76)	0.16 (0.44)
Lessons	0.40 (0.74)	0.30 (0.58)	0.22 (0.51)	0.18 (0.45)	0.10 (0.27)
Structured sports	0.38 (0.91)	0.26 (0.71)	0.19 (0.61)	0.14 (0.50)	0.05 (0.31)
Before or after school time	0.11 (0.56)	0.14 (0.59)	0.10 (0.49)	0.02 (0.21)	0.00 (0.05)
Volunteering or associations	0.22 (0.69)	0.16 (0.59)	0.08 (0.39)	0.06 (0.31)	0.01 (0.12)
Sleep	9.14 (1.58)	9.41 (1.46)	9.35 (1.38)	9.10 (1.38)	9.01 (2.05)
Class Time	6.49 (1.79)	6.62 (1.64)	6.44 (1.91)	6.14 (2.24)	4.06 (3.29)
Homework	0.85 (1.12)	0.75 (0.94)	0.66 (0.86)	0.57 (0.72)	0.29 (0.57)
Active Leisure	1.77 (1.82)	1.57 (1.54)	1.41 (1.50)	1.14 (1.43)	1.25 (1.63)
Reading	0.25 (0.61)	0.17 (0.38)	0.16 (0.39)	0.10 (0.34)	0.07 (0.28)
Hobbies	0.12 (0.46)	0.14 (0.62)	0.10 (0.36)	0.07 (0.29)	0.07 (0.31)
Arts and excursions	0.05 (0.39)	0.03 (0.28)	0.03 (0.27)	0.02 (0.24)	0.02 (0.18)
Sports and physical activities	0.18 (0.55)	0.28 (0.73)	0.22 (0.59)	0.24 (0.66)	0.32 (0.78)
Religious activities	0.06 (0.31)	0.03 (0.21)	0.02 (0.18)	0.02 (0.17)	0.01 (0.15)
Conversations	0.26 (0.72)	0.16 (0.42)	0.18 (0.46)	0.19 (0.49)	0.23 (0.84)
Socialization	0.20 (0.88)	0.16 (0.68)	0.12 (0.52)	0.09 (0.47)	0.14 (0.72)
Affection time	0.05 (0.45)	0.05 (0.40)	0.04 (0.35)	0.02 (0.12)	0.01 (0.06)
Board and family games	0.40 (0.93)	0.43 (0.81)	0.44 (0.93)	0.32 (0.81)	0.34 (0.80)
Other active leisure activities	0.20 (0.68)	0.12 (0.49)	0.09 (0.40)	0.07 (0.32)	0.05 (0.30)
Passive Leisure	0.85 (1.43)	1.34 (1.09)	2.22 (1.11)	3.51 (1.21)	6.16 (1.89)
Video watching	0.00 (0.00)	0.68 (0.27)	1.57 (0.29)	2.85 (0.56)	5.43 (1.40)
Computer (personal use)	0.27 (0.78)	0.20 (0.61)	0.16 (0.55)	0.21 (0.66)	0.18 (0.51)
Video games	0.47 (1.19)	0.38 (0.81)	0.39 (0.83)	0.35 (0.79)	0.46 (1.02)
Other passive leisure activities	0.12 (0.45)	0.09 (0.35)	0.09 (0.34)	0.11 (0.41)	0.09 (0.32)
Duties	0.56 (1.30)	0.38 (0.85)	0.36 (0.82)	0.32 (0.78)	0.51 (1.01)
Care for others	0.04 (0.26)	0.02 (0.15)	0.04 (0.34)	0.02 (0.19)	0.04 (0.32)
Chores	0.27 (0.56)	0.25 (0.46)	0.25 (0.47)	0.24 (0.55)	0.42 (0.80)
Paid work	0.25 (1.17)	0.11 (0.72)	0.07 (0.58)	0.06 (0.51)	0.05 (0.54)
Routines	3.08 (1.27)	2.98 (1.21)	2.87 (1.15)	2.74 (1.16)	2.44 (1.24)
Travel time	0.99 (0.81)	0.93 (0.73)	0.87 (0.66)	0.85 (0.65)	0.70 (0.58)
Shopping	0.11 (0.39)	0.12 (0.46)	0.10 (0.34)	0.10 (0.36)	0.19 (0.54)
Personal care	1.04 (0.61)	1.01 (0.55)	0.96 (0.58)	0.92 (0.53)	0.79 (0.53)
Meals	0.94 (0.62)	0.92 (0.55)	0.94 (0.58)	0.88 (0.60)	0.76 (0.54)
Other Activities	0.14 (0.82)	0.09 (0.44)	0.10 (0.65)	0.09 (0.38)	0.11 (0.43)
Observations	1,286	1,485	1,366	1,165	404

Notes: The sample includes 5,706 child-wave observations from the 1997, 2002, 2007, 2014, and 2019 CDS-PSID waves. Sample weights are applied and standard deviations are shown in parentheses.

2.2 Other Variables

Below, we describe the remaining variables we use when estimating the effect of video watching time on skills.

Skills

We consider several measures of cognitive and noncognitive skills. The measures of cognitive skills are obtained via test scores on letter word identification, applied problems, and passage comprehension. The measures of noncognitive skills are based on the 32 questions answered by parents about their child's behavior that compose the Behavior Problems Index (BPI).[14] The questionnaire inquires whether various behavioral statements are not true, sometimes true, or often true, such as whether the child cheats or lies, struggles to get along with others, has concentration difficulties, or tends to cry excessively. Appendix Table A.1 shows the wording of these 32 questions as they were asked in the survey. We rescale the answers so that a higher score indicates better development, and we consider different aggregations of these questions. To create our main aggregated measures of cognitive and noncognitive skills, we follow the previous literature (e.g., Caetano et al. 2019) and use iterated principal factor analysis separately per wave, standardizing the predicted factor scores inside the sample to have mean zero and standard deviation equal to one. Appendix Table A.2 shows the factor loadings used to create the aggregated cognitive and noncognitive scores.

As an alternative measure of noncognitive skills, we use the BPI provided directly by the CDS-PSID, which is a count of how many behavioral statements were sometimes or often true. We also use the internalizing and externalizing subscales of the BPI. While the BPI internalizing subscale assesses the child's withdrawn behaviors (e.g., depression), the externalizing subscale assesses outward behaviors (e.g., aggression).[15] In all cases, scores have been standardized to have mean zero and standard deviation equal to one within the analysis sample in each wave, and a higher score indicates better development.

Last, we group the individual BPI items into six domains following Zill (1990): not antisocial, not anxious or depressed, not headstrong, not hyperactive, not immaturely dependent, and no peer conflict. For each domain, we use iterated principal factor analysis separately per wave and standardize the estimated factor scores within the sample to have mean zero and standard deviation equal to one. As with all our skill measures, a higher score indicates better development.

[14] To maintain sample composition stability in the analyses of cognitive skill measures, we require that all three subcomponents for that child-wave observation are not missing. Likewise, for analyses of noncognitive skill measures, we require that responses to all 32 BPI questions for that child-wave observation are not missing.

[15] The internalizing (externalizing) subscale is a count of how many internalizing (externalizing) statements were sometimes or often true. In each wave, the CDS-PSID performs a confirmatory factor analysis on the two subscales, and only items that load are included in the subscales and BPI score provided by the CDS-PSID. About 30 items load on one or both subscales in each wave, and the set of items that load is largely stable across waves. As our sample only includes children aged 6 to 18 enrolled in kindergarten through 12th grade, the externalizing subscale and BPI score include the two school-behavior items.

Controls

We use as control variables a set of predetermined characteristics (with respect to hours of video watched), which include characteristics specific to the child, to the family, and to the observation. All controls come from the CDS or the closest PSID wave prior to the CDS wave. The controls include the child's age and squared age (in months) and indicators for: grade (from kindergarten through grade 12), gender, race/ethnicity (Non-Hispanic White, Non-Hispanic Black, Hispanic, and other non-Hispanic ethnicities), household income tercile,[16] whether the mother is alive, whether the father is alive, number of biological siblings in the household, whether the mother was married at the time of the child's birth, whether the primary PSID respondent in the household answered the survey in English, whether the child is homeschooled, private school enrollment, gifted program enrollment, and special education enrollment. We also include CDS-PSID wave (1997, 2002, 2007, 2014, and 2019) fixed effects as controls.[17]

[16] We use total household income in the closest available PSID wave prior to the CDS wave and convert it to 2019 dollars using the CPI-U inflation index from the Bureau of Labor Statistics.

[17] For some control variables, some observations have missing values. We retain those observations by assigning a unique value to each missing control variable and including an indicator variable for whether that value was originally missing.

Table 6: Summary Statistics

Variable	Mean	Std.Dev.
Video Watching During Weekdays	1.52	1.53
Child is Male	0.50	0.50
Child is Non-Hispanic White	0.49	0.50
Child is Non-Hispanic Black	0.40	0.49
Child is Hispanic	0.08	0.26
Child is of Another Race/Ethnicity	0.04	0.20
Child's Age (years)	11.79	3.38
Mother was Married at Birth	0.62	0.48
Father is Alive	0.97	0.16
Mother is Alive	0.99	0.08
Child Has No Siblings	0.19	0.39
Child Has One Sibling	0.43	0.49
Child Has Two Siblings	0.26	0.44
Child Has Three or More Siblings	0.12	0.33
Household Income ($,000)	89.68	133.93
Child is Homeschooled	0.02	0.13
Child Attends Private School	0.08	0.27
Child is in Gifted Program	0.23	0.42
Child is in Special Education	0.13	0.34
Language of the Interview is English	0.95	0.21
Wave 1997	0.19	0.39
Wave 2002	0.33	0.47
Wave 2007	0.19	0.40
Wave 2014	0.15	0.35
Wave 2019	0.14	0.34
Observations	5,706	

Notes: Final sample pools data from the 1997, 2002, 2007, 2014, and 2019 CDS-PSID waves. Sample weights not applied.

Table 6 shows the summary statistics for our treatment and control variables. Unlike all the statistics shown so far, this table does not use sample weights, so as to reflect the actual sample used in our causal analysis. Half of the sampled children are non-Hispanic White, with the rest being mostly non-Hispanic Black. The average age is around 12 years. About 62% of the sample have mothers who were married when the child was born. Almost all children have both mothers and fathers alive.[18] The average household income is approximately $90,000 per year in 2019 dollars. A very small proportion of the children are homeschooled, with about 13% of children attending special education and about 23% of the children attending

[18]These variables are used to proxy for the presence of the parents in the household. An alternative would be to use variables regarding the presence of the father and the mother in the household. However, these variables have many missing observations.

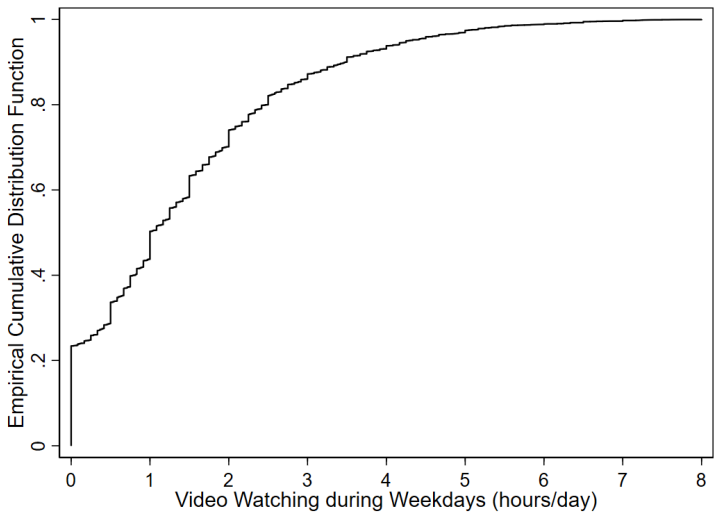
a gifted program in school. Finally, the sample is somewhat skewed towards earlier waves, although all waves are well represented.

3 Empirical Strategy

3.1 Intuition

We first describe our identification strategy intuitively and with graphical evidence, and in Section 3.2 we describe it more formally. We begin by plotting the empirical cumulative distribution function (CDF) of our treatment variable, the number of hours per day that children spend watching video during weekdays, in Figure 2. This variable is distributed fairly continuously on the positive side of the distribution, but then there is a large concentration of observations bunched at zero minutes. Specifically, over 20% of children spend no time watching video on weekdays.

Figure 2: Evidence of Bunching at Zero



Notes: This figure shows the cumulative density function (CDF) of daily hours of weekday video watching. The sample includes data from the 1997, 2002, 2007, 2014, and 2019 CDS-PSID waves ($N = 5,706$).

To explain this bunching, we separate conceptually the *treatment* variable H_i from the *selection* variable H_i^*. The treatment variable H_i is the actual, typical number of hours of video watching per day, our causal variable of interest for child i. In contrast, H_i^* is the potential confounder for that child. It indexes the "type" of each child, incorporating all factors that correlate with H_i. H_i^* includes factors like the child's preferences and constraints that determine H_i, (e.g., preference for video watching, preference for activities that might compete with video watching). It also includes restrictions that might prevent the child

from watching video, or from watching as many hours as the child would otherwise like (e.g., a disability, parental pressure to limit screen time, whether the family owns a TV). For simplicity, below we describe children as "choosing" H_i given their preferences and constraints dictated by their type H_i^*, even though the decision may in practice be made largely by parents and the constraints imposed on children might be such that it does not feel like a choice from their perspective.

To illustrate the value of this conceptual separation between H_i and H_i^*, consider a simple comparison of the outcomes of children who spend, say, 2 daily hours watching video with those who spend 1 daily hour watching video:

$$\underbrace{\mathbb{E}[S_i|H_i=2]-\mathbb{E}[S_i|H_i=1]}_{\text{observed difference}}=\mathbb{E}[S_i(2)|H_i=2]-\mathbb{E}[S_i(1)|H_i=1]$$
$$=\underbrace{\mathbb{E}[S_i(2)|H_i=2]-\mathbb{E}[S_i(1)|H_i=2]}_{\text{causal effect}}+\underbrace{\mathbb{E}[S_i(1)|H_i=2]-\mathbb{E}[S_i(1)|H_i=1]}_{\text{selection effect}}$$

where the second line is obtained by adding and subtracting $\mathbb{E}[S_i(1)|H_i=2]$. In the equation above, S_i denotes the outcome (cognitive or noncognitive skill), and we use the standard notation for potential outcomes: $S_i(h)$ is the potential outcome of child i if they experienced h hours of daily video watching. This decomposition shows the fundamental problem of causality: what is observed equals the causal effect plus the selection (confounding) effect. Note that H_i in the equation above plays two different roles: (a) it indexes the value of h in "$S_i(h)$," denoting the number of hours spent on video watching in that counterfactual, and (b) it indexes the selected sample of the population we are considering, or the "type" (when we condition the expectation, i.e., "$|H_i=h$" in the equation above). Throughout the paper, we will use H_i to denote only the treatment variable when it plays role (a), and we will use H_i^*, the selection variable, when we refer to role (b).

One might think that H_i and H_i^* always take the same value in the data, and therefore this conceptual separation is unnecessary. Indeed, the person who *actually chooses* h hours of video watching is typically the person *of type* h. However, this is not always true in the case of bunching. While each child choosing $H_i=h>0$ must have had preferences and constraints that made them choose h hours of video watching, so their type is indeed $H_i^*=h$, something unique happens at $H_i=0$. The value 0 is a natural lower bound on the value of H_i (because the actual hours spent watching video cannot be negative), but there is no analogous lower bound on the selection variable H_i^*. Therefore, children who spend $H_i=0$ hours watching video may be of different types $H_i^*\leq 0$. Among the children at $H_i=0$, there may be children of type $H_i^*=0$, which are those who choose $H_i=0$ as an interior solution (i.e., they are exactly indifferent between spending zero and a marginal

amount of time watching video), and there may be children of types $H_i^* < 0$, which are those who choose $H_i = 0$ as a corner solution (i.e., the restriction that H_i cannot be negative is binding for them). For instance, consider a child who would prefer to spend more time on an enrichment activity and is strictly willing to exchange an hour of video watching for an extra hour of enrichment activity, if possible, but this exchange is not possible because that child already spends no time watching video ($H_i = 0$). Alternatively, consider the nudge necessary to make those at $H_i = 0$ choose instead a positive amount of video watching, $H_i > 0$. For children of type $H_i^* = 0$, even a small nudge would make them do so, while for children of type $H_i^* < 0$, that nudge might not be enough.

We formalize the discussion above with the following equation relating the treatment variable H_i and the selection variable H_i^*:

$$H_i = H_i^* \cdot \mathbf{1}\left(H_i^* \geq 0\right), \text{ with } \mathbb{P}(H_i^* < 0) > 0, \tag{1}$$

where $\mathbf{1}(\cdot)$ is the indicator function. The bunching condition in equation (1) assumes that there are at least some children of type $H_i^* < 0$. Specifically, we assume that there are children at $H_i = 0$ whose preferences and constraints led them to be some distance below the exact indifference level $H_i^* = 0$, which explains the concentration of children at $H_i = 0$ in Figure 2.

Because of the existence of children at the corner solution, the average type H_i^* must be discontinuous at $H_i = 0$. This discontinuity is given by:

$$\Delta_{H^*} := \lim_{h \downarrow 0} \mathbb{E}[H_i^* | H_i = h] - \mathbb{E}[H_i^* | H_i = 0] = 0 - \mathbb{E}[H_i^* | H_i = 0] = -\mathbb{E}[H_i^* | H_i = 0] \neq 0. \tag{2}$$

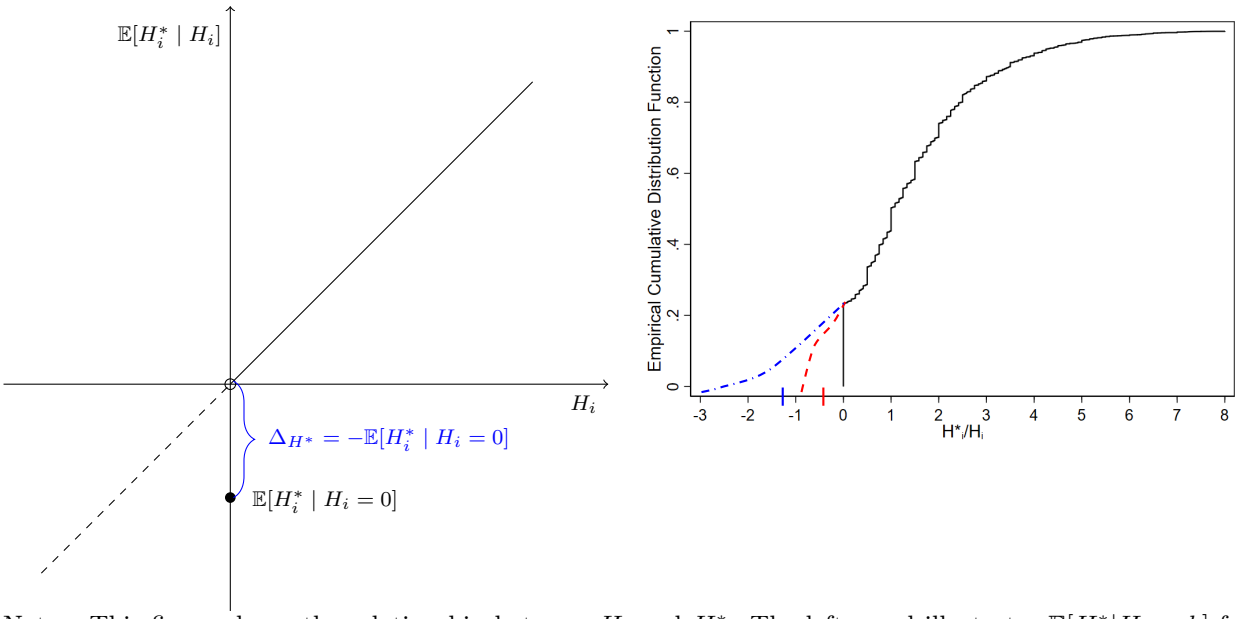
The first term of the equation shows that as we consider children who spend fewer and fewer hours watching video, their average type converges to zero.[19] The discontinuity $\Delta_{H^*} = -\mathbb{E}[H_i^* | H_i = 0]$ represents the average distance from exact indifference between spending no time watching video and spending a marginal amount of time watching video among the children at $H_i = 0$. This discontinuity arises because we are averaging different types at $H_i = 0$, and some types are at a corner solution, as can be seen by the dashed line in the left panel in Figure 3. Relative to the children at the interior solution $H_i^* = 0$, children at the corner solution may live in households that differ in parenting style, household rules, access to devices, parental beliefs about screen time, and many other dimensions affecting the desirability of watching video, so much so that their value of H_i^* is lower. Such differences

[19] One might be concerned about whether the treatment variable is distributed sufficiently continuously around zero for this approximation argument to work in practice. The slope of the CDF plot as H_i approximates $H_i = 0$ in Figure 2 implies that indeed there are some children spending positive marginal hours (e.g., a few minutes or less per day) watching video.

are not a threat to our identification strategy; rather, they are precisely what generate heterogeneity in H_i^* among children at $H_i = 0$, which will help us isolate the selection effect.

Figure 3: Relationship between H_i and H_i^*

(a) Discontinuity of $\mathbb{E}[H_i^* \,|\, H_i = h]$ at $h = 0$ (b) Distribution of H_i and H_i^*



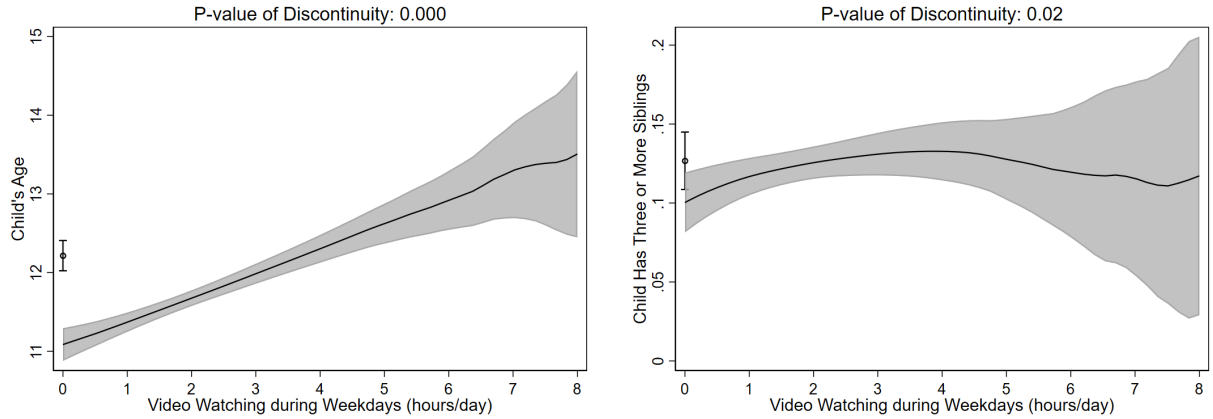
Notes: This figure shows the relationship between H_i and H_i^*. The left panel illustrates $\mathbb{E}[H_i^*|H_i = h]$ for different values of h, with $\mathbb{E}[H_i^*|H_i = h] = h$ for $h > 0$ and $\mathbb{E}[H_i^*|H_i = h] < 0$ for $h = 0$ due to the presence of observations with $H_i^* < 0$ (see equation (1)). The right panel shows as a solid line the distribution of H_i (and also of $H_i^*|H_i > 0$), which is identical to the plot in Figure 2. Moreover, the dashed lines show two hypothetical lower tails of the distribution of H_i^*, along with their corresponding values of $\mathbb{E}[H_i^*|H_i = 0]$ as tick markers on the horizontal axis.

We cannot observe this discontinuity directly because we do not observe H_i^* for those at $H_i = 0$. The right panel of Figure 3 illustrates what is and is not observed in the data. The full distribution of the treatment variable H_i in the solid black line is observed, and is identical to Figure 2. However, only the positive part of the distribution of the selection variable H_i^* is observed (i.e., the part where $H_i^* = H_i$). Nothing in the data tells us directly how exactly H_i^* is distributed among observations at $H_i = 0$; the only thing we know is that $H_i^* \leq 0$ (otherwise these children would have chosen $H_i > 0$ instead). The dashed lines in the right panel of Figure 3 represent two hypothetical ways in which H_i^* could be distributed in the lower tail, one where types are more concentrated near 0 (red dashed line) and another where they are more spread out (blue dash-dotted line). These hypotheticals lead to different values of $\mathbb{E}[H_i^*|H_i = 0]$, which are shown as tick markers on the horizontal axis. There is an infinite number of possible values for $\mathbb{E}[H_i^*|H_i = 0]$, depending on the assumed lower tail. We therefore introduce an assumption below that allows us to pin down

the value of $\mathbb{E}[H_i^*|H_i = 0]$ by inferring the lower tail as the mirror image of the upper tail of the distribution, which we observe in the data.

Even without directly observing the discontinuity $\Delta_{H^*} = -\mathbb{E}[H_i^*|H_i = 0]$, it is easy to see that the type of the child changes discontinuously at $H_i = 0$. Indeed, Figure 4 establishes that key observed variables that are likely associated with H_i^* are themselves discontinuous at $H_i = 0$. Specifically, each panel shows the average value of a key observed variable for the $H_i = 0$ subsample and a local linear regression fit of that variable on H_i, estimated on the $H_i > 0$ subsample. The left panel shows a negative discontinuity in the age of the child at zero, while the right panel shows a negative discontinuity in the proportion of children who have three or more siblings.

Figure 4: Evidence of Children at the Corner Solution



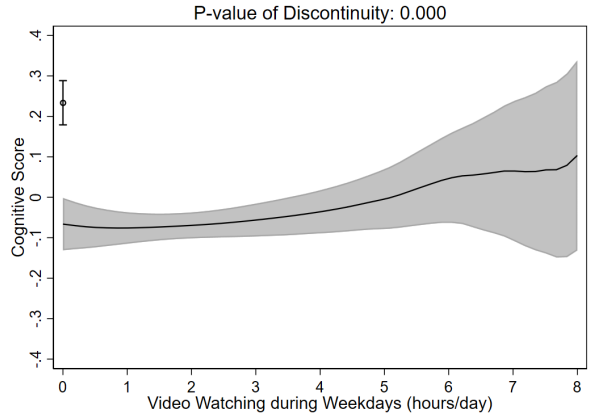
Notes: For each panel, we report $\mathbb{E}[X_i|H_i = h]$ estimated as a simple average for $h = 0$ and as a local linear regression of X_i on H_i (with a bandwidth of 2 hours) for $h > 0$, along with the 95% confidence intervals. The p-value of a test for whether there is a discontinuity at $h = 0$ is shown on top of each panel. The definition of the covariate X_i is shown in the vertical axis.

If the selection variable H_i^* varies discontinuously at $H_i = 0$, and H_i^* is a confounder in our causal analysis (i.e., it is associated with the outcome), then the outcome should vary discontinuously as well. As expected, we also find discontinuities in our outcome variables at $H_i = 0$. In Figure 5, we present analogous discontinuity plots for children's cognitive (left panel) and noncognitive skills (right panel). There is strong evidence of negative selection for cognitive skills, reflecting that children who spend a marginal amount of hours watching video have discontinuously lower cognitive skill than those who spend zero hours watching video. In contrast, for noncognitive skills, there is evidence of positive selection—children who spend a marginal amount of hours watching video have discontinuously higher noncognitive skill than those who spend zero hours watching video. These discontinuities are plausibly due to the discontinuity in $\mathbb{E}[H_i^*|H_i = h]$ shown in equation (2) (left panel of Figure 3). The only alternative explanation is to implausibly assume discontinuous treatment effects, which

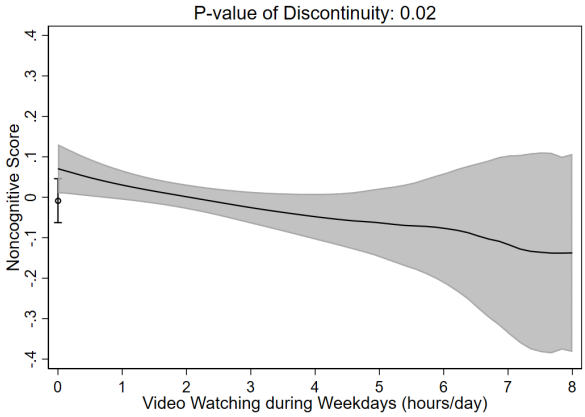
would mean that children who spend just a few minutes watching video per day would see a discontinuous change in their skills relative to the counterfactual of spending zero minutes watching video.

Figure 5: Evidence of Selection

(a) Cognitive Skills



(b) Noncognitive Skills



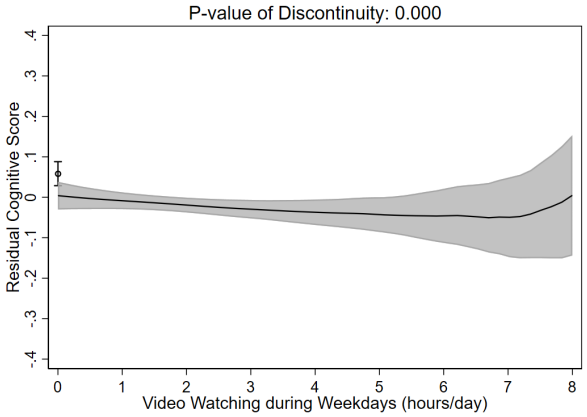
Notes: For each outcome variable S_i, we report $\mathbb{E}[S_i|H_i = h]$ estimated as a simple average for $h = 0$ and as a local linear regression of S_i on H_i (with a bandwidth of 2 hours) for $h > 0$, along with the 95% confidence intervals. The p-value of a test for whether there is a discontinuity at $H_i = 0$ is shown on top of each panel.

So far, we have shown evidence of selection due to H_i^*, but this selection may in principle be entirely explained by the discontinuities in observable characteristics, such as the ones we show in Figure 4. Next, we rule out this possibility by showing direct evidence of selection on unobservables in Figure 6 (Caetano, 2015). Each panel is analogous to the corresponding panel in Figure 5, but the plots are now constructed using residualized outcomes after accounting for the set of control variables used in our analysis, so observed characteristics can no longer account for any portion of the discontinuity.[20] As the plots show, we find strong evidence of negative selection on unobservables for cognitive skills and positive selection on unobservables for noncognitive skills. As the next remark shows, these residual plots are informative about the likely signs of the causal effects of H_i on S_i.

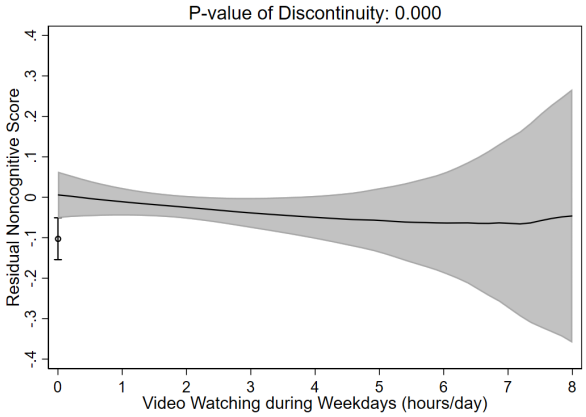
[20] In this figure we report the discontinuity in the residual of S_i after controls are accounted for, $S_i - \hat{\mathbb{E}}[S_i|X_i]$, where X_i is the vector of covariates. See Section 2 for the list of variables included in X_i.

Figure 6: Evidence of Selection on Unobservables

(a) Cognitive Skills



(b) Noncognitive Skills



Notes: At $H_i = 0$, we report $\mathbb{E}[S_i - \lim_{h \to 0} \mathbb{E}[S_i|H_i = h, X_i]|H_i = 0]$, estimated as a simple average, where $\mathbb{E}[S_i|H_i = h, X_i]$ is estimated as a linear regression of S_i on H_i and X_i using only observations such that $H_i > 0$. For $H_i > 0$, we report $\mathbb{E}[S_i - \mathbb{E}[S_i|H_i = h, X_i]|H_i = h]$ for $h > 0$, estimated as a local linear regression of $S_i - \mathbb{E}[S_i|H_i = h, X_i]$ on H_i (with a bandwidth of 2 hours). The 95% confidence intervals are also reported. The list of controls X_i is given in Section 2. The p-value of a test for whether there is a discontinuity in the plot at $H_i = 0$ is shown on top of each panel.

Remark 3.1. *(Qualitative Findings) The right panel of Figure 6 provides evidence that the effect of video watching on noncognitive skills is negative. To see this, let $S_i(h)$ be the potential outcome of child i if they were treated at the level h. Now, consider the analogous decomposition we made above for any positive value of the treatment h:*

$$\overbrace{\mathbb{E}[S_i(h+1)|H_i = h+1] - \mathbb{E}[S_i(h)|H_i = h]}^{observed\ slope\ =\beta(h)+\delta(h)} = \overbrace{(\mathbb{E}[S_i(h+1)|H_i = h+1] - \mathbb{E}[S_i(h)|H_i = h+1])}^{treatment\ effect\ =\beta(h)} + \overbrace{(\mathbb{E}[S_i(h)|H_i = h+1] - \mathbb{E}[S_i(h)|H_i = h])}^{selection\ effect\ =\delta(h)},$$

where we consider only treatment increments of one unit to keep things simple. This decomposition states that the slope of the plot in Figure 6 at $H_i = h$ is equivalent to the treatment effect $\beta(h)$ plus the selection effect $\delta(h)$ at $H_i = h$. If we knew the value of the selection effect $\delta(h)$, then we would be able to obtain the value of $\beta(h)$ by subtracting $\delta(h)$ from the slope observed in the figure. We actually do not know $\delta(h)$, but we do know the sign of the selection effect at $h = 0$, $\delta(0)$. Indeed, the positive discontinuity in the outcome in the right panel from Figure 6 tells us that $\delta(0) > 0$. Thus, if we assume that the signs of the selection effects at $h > 0$ and at $h = 0$ are the same, then $\delta(h) > 0$, allowing us to conclude that the treatment effect $\beta(h)$ must be negative (since the slope of the plot in the right panel in Figure 6 is negative, implying $\beta(h) + \delta(h) < 0$).

Using the analogous logic, we can infer based on the left panel in Figure 6 that the effect of H_i on cognitive skills is near zero.[21] *Note that these qualitative conclusions can be reached with only two assumptions: continuity of treatment effects (so that the discontinuity in Figure 6 is only attributed to selection)*[22] *and the sign of the selection effect is the same for $h = 0$ and for $h > 0$. Below we make further assumptions in order to pin down the point estimates.*

3.2 Identification Strategy

To formalize the discussion above, we write the outcome equation of interest as:

$$S_i = \beta H_i + U_i$$

where S_i denotes the outcome, which is a child's skill measure (either a cognitive or noncognitive score measure), β is the parameter of interest, and U_i is the remainder of the equation, including a constant. In general, U_i may be a nonlinear function of the type H_i^*—i.e., it can be written generally as $U_i = \mathbb{E}[U_i|H_i^*] + \epsilon_i$, where ϵ_i is an idiosyncratic error, so that the slope of $\mathbb{E}[U_i|H_i^*]$ can change with each value of H_i^*, as discussed in the remark above.

In this section, we impose two complementary assumptions that allow us to express the nonparametric function $\mathbb{E}[U_i|H_i^*]$ in terms of quantities that can be identified. We start with Assumption 1 below, which imposes the restriction that $\mathbb{E}[U_i|H_i^*]$ is a linear function.

Assumption 1. *(Linearity in H_i^*)* $\mathbb{E}[U_i|H_i^*] = \alpha + \delta H_i^*$

This assumption essentially states that the covariance between H_i and the error U_i is zero once H_i^* is controlled for *linearly*. Remark 3.1 above discussed the weaker assumption that the selection effect at $H_i = 0$ has the same sign as the selection effect at $H_i > 0$. Assumption 1 is stronger because it requires the *value* (not just the sign) of the marginal selection effect, δ, to be the same across different values of H_i^*; that is, it assumes $\delta(h) = \delta(0)$ instead of $sign(\delta(h)) = sign(\delta(0))$. Section 3.4 below provides more detail and intuition about this assumption.

Figure 6 suggests that this assumption is quite reasonable in our case, especially after controls are accounted for. Indeed, the slope $\beta(h) + \delta(h)$ can be observed directly in that

[21] The discontinuity in the left panel in Figure 6 reveals that the selection effect is negative ($\delta(0) < 0$), and the slope of the plot is negative but of a modest magnitude for the vast majority of values of $H_i > 0$.

[22] Actually, continuity of treatment effects is technically not a necessary condition to determine the sign of the selection effect at $h = 0$, $\delta(0)$. Even if some of the discontinuity in the residualized outcome at $h = 0$ was attributed to treatment effect discontinuity, we would still arrive at the same qualitative conclusions about $\delta(0)$. Our conclusions about $\delta(0)$ would only change if the discontinuity due to treatment effects is in the same direction as the discontinuity in Figure 6 and so large that the remaining discontinuity (due to unobservables) would be in the opposite direction as the one observed in the figure.

figure, and it appears to be approximately constant over most values of h, except perhaps for extremely high values of h, which are too rare to matter empirically (explaining why the confidence interval in that range is so much larger).[23] In Sections 4.2.2 and 4.2.3, we show further evidence that the linearity assumption is reasonable in our context, including in the unobserved lower tail (where $H_i^* < 0$).

This assumption allows us to write the outcome equation:

$$S_i = \alpha + \beta H_i + \delta H_i^* + \epsilon_i, \qquad cov(H_i, \epsilon_i | H_i^*) = 0. \tag{3}$$

Next, we consider how the average of the outcome S_i changes with each value of H_i:[24]

$$\mathbb{E}[S_i|H_i] = \alpha + \beta H_i + \delta(H_i + \mathbb{E}[H_i^*|H_i = 0]\mathbf{1}(H_i = 0)). \tag{4}$$

This equation shows that, although H_i^* is not observed, we can proxy for it with $H_i + \mathbb{E}[H_i^*|H_i = 0]\mathbf{1}(H_i = 0)$. The only unknown quantity in this proxy is $\mathbb{E}[H_i^*|H_i = 0]$. As discussed in Section 3.1, we know that $\mathbb{E}[H_i^*|H_i = 0]$ is negative (since $H_i^* \leq 0$ among observations at $H_i = 0$, and $H_i^* < 0$ for some of these observations), but we do not know its exact value (see discussion of Figure 3). An approach that avoids any further assumption is to estimate β for all possible values of $\mathbb{E}[H_i^*|H_i = 0]$. This can be implemented by estimating equation (4) using OLS, regressing S_i on a constant, H_i, and the control function $H_i + \mathbb{E}[H_i^*|H_i = 0]\mathbf{1}(H_i = 0)$, for the full range of candidate values of $\mathbb{E}[H_i^*|H_i = 0]$. In Section 4.2.1, we implement this agnostic approach and show that the key qualitative findings of the paper do not depend on any further assumption, confirming the results from Remark 3.1. Moreover, we show that the reported point estimates are conservative, in the sense that there is little scope for the results to be less intense than reported, and much more scope for them to be more intense than reported.

While the agnostic approach is informative, reporting ranges of estimates for every outcome, sample, treatment definition, and subgroup would be impractical. We therefore follow Caetano, Caetano, and Nielsen (2024b) (also applied in Caetano, Caetano, and Nielsen 2024a and Caetano, Caetano, Nielsen, and Sanfelice 2025), which assumes that H_i^* has symmetric tails and allows us to report a single estimate of β for each specification.

Assumption 2. *(Tail Symmetry of H_i^*) The tails of the distribution of H_i^* below the $\mathbb{P}(H_i = 0)$-th quantile and above the $(1 - \mathbb{P}(H_i = 0))$-th quantile are symmetric.*

[23] It is in principle possible that both the treatment effect $\beta(h)$ and the selection effect $\delta(h)$ vary with h, but in offsetting ways so that their sum $\beta(h) + \delta(h)$ would look approximately constant. The sensitivity analysis reported in Section 4.2.3 rules out this knife-edge case.

[24] This equation can be obtained from equation (3) by noting that (a) $H_i^* = H_i + H_i^*\mathbf{1}(H_i = 0)$ (since $H_i^* = H_i$ when $H_i > 0$), and (b) $\mathbb{E}[\epsilon_i|H_i] = 0$, since by the iterated law of expectations $\mathbb{E}[\epsilon_i|H_i] = \mathbb{E}[\mathbb{E}[\epsilon_i|H_i^*]|H_i] = \mathbb{E}[0|H_i] = 0$.

This assumption requires that the lower tail of the distribution of H_i^* is the mirror image of the upper tail of the distribution of H_i^* (which is observed because $H_i^* = H_i$ when $H_i > 0$).[25] The full distribution of H_i can be seen in the right panel of Figure 3, where we also illustrate two hypothetical lower tails of H_i^*. Although H_i^* tends to be symmetrical in the middle range of the distribution, where the portions above and below the median can be observed, we cannot directly verify the tail symmetry property because we cannot compare the upper tail (which is observed) with the lower tail (which is unobserved).

Assumption 2 is justified on statistical grounds. Tail symmetry is more general than symmetry. In turn, symmetry is more general than many of the parametric distributional assumptions routinely made in economics, such as the normal, uniform, or logistic distributions, all of which are symmetric. Moreover, if H_i^* is a sum of many independent unobserved factors, like in equation (7) below (preference or constraint factors, as discussed in Section 3.1), with no one of them playing a predominant role, then the Central Limit Theorem tells us that H_i^* will approximate a normal distribution irrespective of the distributions of each of these factors. Indeed, normality is not rejected in our data. Of course, not rejecting normality does not mean normality is actually valid. Rather than relying on the stronger assumption of normality, we instead assume only tail symmetry, which accommodates departures from normality that may arise because the unobserved factors are dependent or because some play a more prominent role than others. We show below that the tail symmetry assumption also has the advantage of providing a conservative estimate of the effect of video watching, which tends to be less intense than the estimates implied by the normal or the uniform distributions.

Remark 3.2. *(δ is a Ratio of Two Discontinuities) The role of Assumption 1 becomes clearer if we temporarily allow the relationship between U_i and H_i^* to be more flexible:*

$$\mathbb{E}[U_i|H_i^*] = \alpha + \delta_0 H_i^* \mathbf{1}(H_i = 0) + \delta_+ H_i^* \mathbf{1}(H_i > 0). \tag{5}$$

Here, the selection effect is allowed to differ between children who watch no video and those who watch a positive amount of video. Assumption 1 holds if $\delta_0 = \delta_+$.

With this more flexible specification, equation (4) can be rewritten as:

$$\mathbb{E}[S_i|H_i] = \alpha + (\beta + \delta_+)H_i + \overbrace{\delta_0 \underbrace{\mathbb{E}[H_i^*|H_i = 0]}_{-\Delta_{H^*}}}^{-\Delta_S} \mathbf{1}(H_i = 0), \tag{6}$$

[25] With Assumption 2, we estimate $\hat{\mathbb{E}}[H_i^*|H_i = 0] = \hat{H}_q - \hat{\mathbb{E}}[H_i|H_i \geq \hat{H}_q]$, where $\hat{H}_q$ is the value of H_i corresponding to empirical quantile $q = 1 - \mathbb{P}(H_i = 0)$ and $\hat{\mathbb{E}}[H_i|H_i \geq \hat{H}_q]$ is the sample average over the observed upper tail of H_i. Once $\hat{\mathbb{E}}[H_i^*|H_i = 0]$ is obtained, we estimate β and δ via OLS, regressing S_i on a constant, H_i, and $H_i + \hat{\mathbb{E}}[H_i^*|H_i = 0]\mathbf{1}(H_i = 0)$ (see equation (4)).

which clarifies the connection between our identification strategy and the discontinuity plots in Figures 5 and 6. As discussed in Remark 3.1 above, for $H_i > 0$, $\mathbb{E}[S_i|H_i]$ varies with slope $\beta + \delta_+$, but it is discontinuous at $H_i = 0$, with discontinuity $\Delta_S = -\delta_0 \mathbb{E}[H_i^*|H_i = 0]$. Thus, δ_0 equals the ratio of two discontinuities, Δ_S and Δ_{H^*}, where Δ_{H^*} is defined in equation (2).[26] *Equation (6) also clarifies that the selection effect is separately identified from β only at $H_i = 0$ (the bunching point). Identifying β away from the bunching point therefore requires extrapolating the selection effect identified at $H_i = 0$ to $H_i > 0$, which is precisely the role of Assumption 1.*[27]

3.3 Adding Controls

As in Figure 6, in practice we use covariates X_i to relax the identification assumptions. Instead of assuming linearity of U_i in H_i^* (Assumption 1), we assume linearity of U_i in $H_i^* - \mathbb{E}[H_i^*|X_i]$ (the residualized version of H_i^* after controls X_i are accounted for), allowing for the part of U_i that is predicted by controls X_i to be arbitrarily nonlinear. Similarly, instead of assuming tail symmetry of H_i^* (Assumption 2), we assume tail symmetry of $\frac{H_i^* - \mathbb{E}[H_i^*|X_i]}{\sigma(H_i^*|X_i)}$ (the standardized version of H_i^* for each value of X_i, where $\sigma(H_i^*|X_i)$ is the standard deviation of the distribution of H_i^* conditional on X_i), allowing for the part of H_i^* that is predicted by controls X_i to be arbitrarily asymmetric. In the main results reported in Section 4, controls X_i enter linearly in the specifications, but in Section 4.2.3 we allow them to enter nonparametrically and the results do not change.

3.4 Understanding Assumption 1

In this section, we discuss Assumption 1 in more detail and describe diagnostic exercises designed to assess whether the assumption is a reasonable approximation in our context.

[26] There is a connection between this identification strategy and fuzzy regression discontinuity designs (RDDs), which also identify effects as Wald ratios of two discontinuities—a "reduced-form" discontinuity divided by a "first-stage" discontinuity. The two key distinctions here are that the selection effect δ_0, not the treatment effect β, is identified this way, and that the "first-stage" discontinuity Δ_{H^*} is not identified directly (since H_i^* is unobserved). See Caetano et al. (2025) for a detailed discussion of this connection.

[27] The linearity assumption is different from the assumptions required to extrapolate RDD effects away from the cutoff, such as in Angrist and Rokkanen (2015). Because our framework does not involve a running variable, there is no analogue of the conditional independence assumption used in RDD extrapolation, whereby the running variable is assumed to be unrelated to the potential outcomes after conditioning on additional covariates.

3.4.1 Multiple Confounding Factors

We first show that the linearity assumption does *not* restrict the number of confounding factors to be one. To see this, suppose that U_i is determined by multiple unobserved confounding factors beyond an idiosyncratic error:

$$U_i = \alpha + \sum_{j=1}^{J} \delta_j F_{ij} + \epsilon_i, \tag{7}$$

where each F_{ij} is a different confounding factor and δ_j represents the association between that factor and the outcome. Each factor might matter more depending on δ_j, which can be positive or negative. For Assumption 1 to be approximately valid, what matters is whether on the whole U_i is well approximated by a linear function in H_i^*. Thus, what matters is whether $\sum_j \delta_j \mathbb{E}[F_{ij}|H_i^*]$ is approximately linear, even if each of the confounding factors F_{ij} is a nonlinear function of H_i^*. Intuitively, Assumption 1 will be approximately valid if: (i) $\mathbb{E}[F_{ij}|H_i^*]$ is nearly linear in H_i^* across all factors F_{ij}; (ii) nonlinearities in different confounding factors offset one another so that $\mathbb{E}[U_i|H_i^*]$ is nearly linear; or (iii) nonlinear confounding factors have only a weak association with the outcome (their δ_js are small in magnitude). The example below discusses in more detail a violation of Assumption 1 using a simplified but concrete example with two confounding factors.

3.4.2 A Concrete Example Where Assumption 1 Is Violated

For concreteness, we consider an example with only two confounding factors, "preference for video watching", V_i, and "how overscheduled the child is with enrichment activities", O_i. We write:

$$U_i = \alpha + \delta_V V_i + \delta_O O_i + \epsilon_i, \tag{8}$$

where δ_V and δ_O capture the association between each confounder and the outcome. Controlling directly for V_i and O_i would solve the endogeneity concern, since ϵ_i is an idiosyncratic error. However, neither V_i nor O_i are observed.

Our identification strategy instead controls for a proxy of the unobserved type H_i^*. Using equation (8), we can write:

$$\mathbb{E}[U_i|H_i^*] = \alpha + \delta_V \mathbb{E}[V_i|H_i^*] + \delta_O \mathbb{E}[O_i|H_i^*]. \tag{9}$$

Thus, the shape of $\mathbb{E}[U_i|H_i^*]$ is determined by the shapes of $\mathbb{E}[V_i|H_i^*]$ and $\mathbb{E}[O_i|H_i^*]$, weighted by the selection effects of V_i and O_i (δ_V and δ_O). Assumption 1 is valid provided that both $\mathbb{E}[V_i|H_i^*]$ and $\mathbb{E}[O_i|H_i^*]$ are linear, or that nonlinearities in $\mathbb{E}[V_i|H_i^*]$ and $\mathbb{E}[O_i|H_i^*]$ offset each

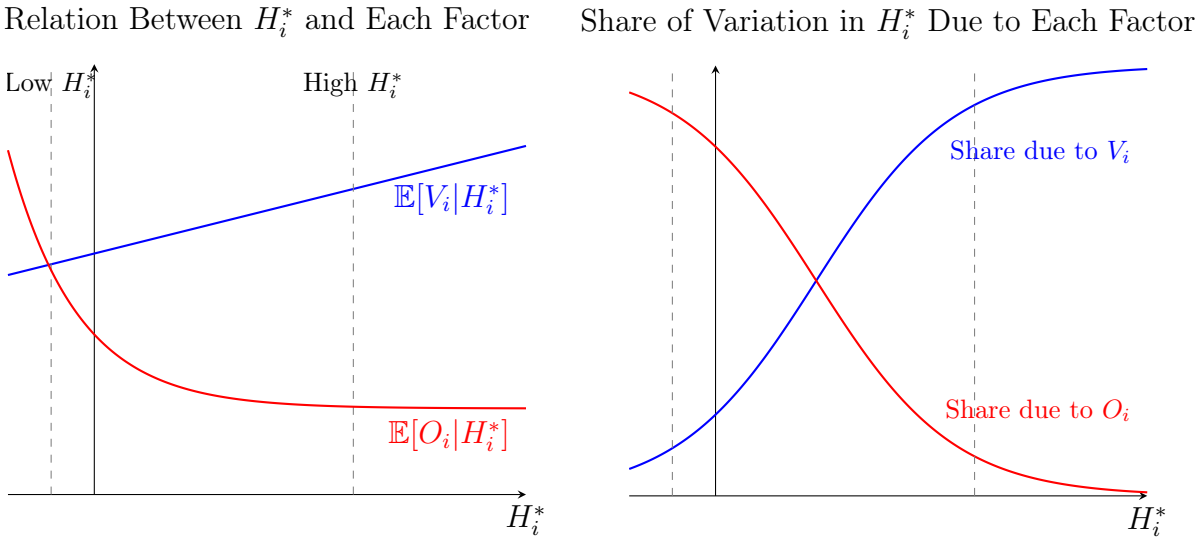
other so that $\delta_V \mathbb{E}[V_i|H_i^*] + \delta_O \mathbb{E}[O_i|H_i^*]$ is linear in H_i^*. Importantly, this conclusion does not depend on the sign or the magnitude of the endogeneity (i.e., irrespective of the signs or magnitudes of the slopes of the curves $\mathbb{E}[V_i|H_i^*]$ and $\mathbb{E}[O_i|H_i^*]$, and irrespective of the signs or magnitudes of δ_V and δ_O).

Figure 7 illustrates a case where Assumption 1 is violated. Here $\mathbb{E}[V_i|H_i^*]$ is a linear, increasing function of H_i^* while $\mathbb{E}[O_i|H_i^*]$ is a convex, decreasing function of H_i^*. Since $\delta_O \neq 0$, this convexity carries into $\mathbb{E}[U_i|H_i^*] = \delta_V \mathbb{E}[V_i|H_i^*] + \delta_O \mathbb{E}[O_i|H_i^*]$, leading $\mathbb{E}[U_i|H_i^*]$ to be nonlinear in H_i^*. This convexity of $\mathbb{E}[O_i|H_i^*]$ implies that the slope of $\mathbb{E}[O_i|H_i^*]$ is steep for low H_i^* and flattens for high H_i^*, while the slope of $\mathbb{E}[V_i|H_i^*]$ remains constant throughout, so that:

$$\left|\frac{\partial \mathbb{E}[V_i|H_i^*]}{\partial H_i^*}\right| < \left|\frac{\partial \mathbb{E}[O_i|H_i^*]}{\partial H_i^*}\right| \text{ for low } H_i^*, \text{ and } \left|\frac{\partial \mathbb{E}[V_i|H_i^*]}{\partial H_i^*}\right| > \left|\frac{\partial \mathbb{E}[O_i|H_i^*]}{\partial H_i^*}\right| \text{ for high } H_i^*.$$

Thus, a marginal increase in H_i^* leads to a different change in the mix of the underlying confounders depending on the level of H_i^*, as shown in the right panel of the figure. For low values of H_i^*, a small increase in H_i^* is associated primarily with changes in O_i. For high values of H_i^*, the same marginal increase is associated primarily with changes in V_i.

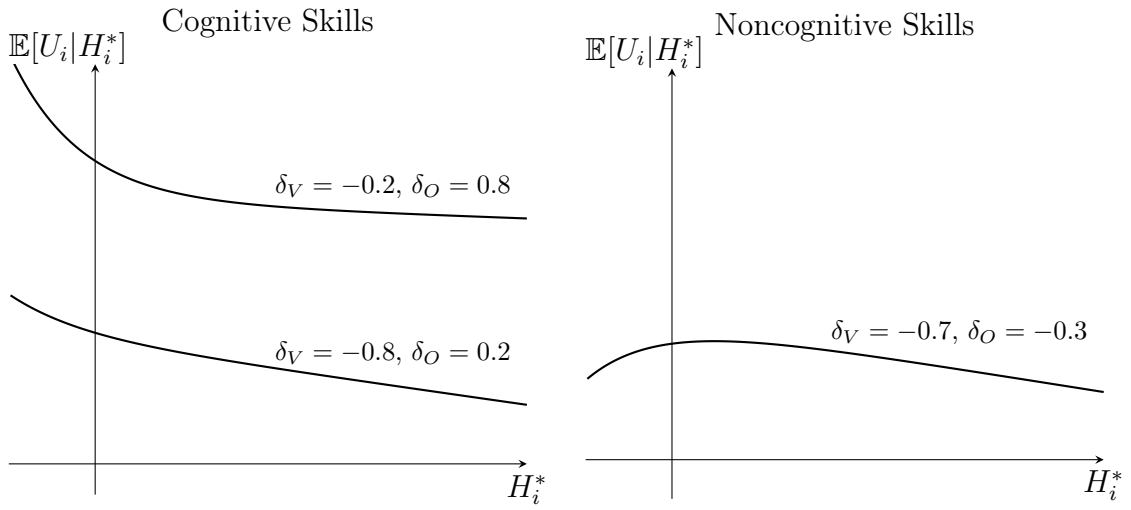
Figure 7: How Nonlinearity Leads to Different Mixes of Confounders



Note: The left panel plots the hypothetical scenario where $\mathbb{E}[V_i|H_i^*] = 56 + H_i^*$ and $\mathbb{E}[O_i|H_i^*] = 20 + 60e^{-0.25(H_i^*+5)}$ for the range $-5 \leq H_i^* \leq 25$. The right panel plots the share of local variation in the confounding factors accounted for by V_i and O_i, respectively. These shares are computed as $|d\mathbb{E}[V_i|H_i^*]/dH_i^*|/(|d\mathbb{E}[V_i|H_i^*]/dH_i^*| + |d\mathbb{E}[O_i|H_i^*]/dH_i^*|)$ and $|d\mathbb{E}[O_i|H_i^*]/dH_i^*|/(|d\mathbb{E}[V_i|H_i^*]/dH_i^*| + |d\mathbb{E}[O_i|H_i^*]/dH_i^*|)$.

Even though $\mathbb{E}[O_i|H_i^*]$ is strongly convex, it is not clear whether this nonlinearity leads to meaningful bias in our estimator of β. The left panel of Figure 8 shows two examples of how the shapes of $\mathbb{E}[V_i|H_i^*]$ and $\mathbb{E}[O_i|H_i^*]$ from Figure 7 translate into nonlinearities of $\mathbb{E}[U_i|H_i^*]$, depending on the values of δ_V and δ_O, according to equation (9). When the selection effect of V_i is four times as large in magnitude as the selection effect of O_i, the linearity of V_i in the left panel dominates the nonlinearity of O_i, leading $\mathbb{E}[U_i|H_i^*]$ to have an approximately linear shape. Although the linearity assumption is violated in this case, the consequence in terms of bias of the estimator is minor because the shape is almost linear. In contrast, when the selection effect of O_i is four times as large in magnitude as the selection effect of V_i, then the nonlinearity in O_i dominates the linearity in V_i, leading $\mathbb{E}[U_i|H_i^*]$ to have a strongly concave shape. In this case, the linearity assumption is strongly violated and the resulting bias in our estimator of β is large. Thus, the empirically relevant question is not whether nonlinear confounding factors exist, but whether they generate consequential nonlinearities in $\mathbb{E}[U_i|H_i^*]$.

Figure 8: Nonlinearity Examples



Note: Both panels use $\mathbb{E}[V_i|H_i^*] = 56 + H_i^*$ and $\mathbb{E}[O_i|H_i^*] = 20 + 60e^{-0.25(H_i^*+5)}$ for the range $-5 \leq H_i^* \leq 25$, as in Figure 7. The panels plot the resulting $\mathbb{E}[U_i|H_i^*] = \alpha + \delta_V \mathbb{E}[V_i|H_i^*] + \delta_O \mathbb{E}[O_i|H_i^*]$, with $\alpha = 80$, and with the values of δ_V and δ_O shown for each plot. In both panels, $\delta_V < 0$. The cognitive skill example sets $\delta_O > 0$, while the noncognitive skill example sets $\delta_O < 0$.

This figure also shows that the implications for the linearity assumption are unrelated to whether the associations between the confounders and outcome have the same or opposite signs. In the left panel, we consider examples that are more intuitive for cognitive skills, where we set $\delta_V < 0$ and $\delta_O > 0$ to reflect the idea that children who have lower preference

for video watching and are more overscheduled with enrichment activities tend to have higher cognitive skills. These signs lead to a negative selection effect around $H_i = 0$, consistent with the negative discontinuity in the left panel of Figure 6.

In contrast, in the right panel, we set $\delta_V < 0$ and $\delta_O < 0$ to reflect the idea that children who have lower preference for video watching and are less overscheduled with enrichment activities tend to have higher noncognitive skills.[28] To highlight a potentially concerning situation for our identification strategy, we set $\delta_V = -0.7$ and $\delta_O = -0.3$. Under these values, the selection effect is positive for $H_i^* \leq 0$ but mostly negative for $H_i^* > 0$, violating not only the linearity assumption, but even the weaker equal-sign assumption presented in Remark 3.1.

3.4.3 Detecting Violations of Assumption 1

The key practical implication of the nonlinearities illustrated in Figure 8 is that estimates of δ (and therefore the estimates of β) should vary depending on which portion of the support of H_i^* is used as the source of variation. This implication motivates the truncation exercise we present in Section 4.2.2, where we restrict the sample to observations with $H_i < H_{\max}$ and progressively increase the value of the truncation point $H_{\max}$. If strong nonlinearities are present, estimates of δ (and thus β) should change as $H_{\max}$ changes because the slope of the best linear approximation changes as different portions of the support are used for estimation.[29] In Section 4.2.2 we find that the estimates of β generally remain stable as $H_{\max}$ changes, providing evidence that nonlinear confounders, if present, are not empirically important.

The truncation exercise can detect only violations of linearity that show up in the range $H_i^* > 0$. However, it is still possible that nonlinearities arise primarily for $H_i^* < 0$, as illustrated in the example from Figure 8. For this reason, we consider a complementary diagnostic in Section 4.2.3, in which we allow α (the constant in our estimating equation) and δ (the selection effect) to vary across different clusters of the control vector X_i. As we show later, these clusters themselves are associated with different ranges of H_i^*, including negative values of H_i^*.[30] In the case of nonlinearities, the selection effect would vary across

[28] Caetano et al. (2024a) suggest that children who are overscheduled with enrichment activities tend to have lower levels of noncognitive skills.

[29] Let $\delta_+(H_{\max})$ denote the slope of the best linear projection of the selection over $0 < H_i < H_{\max}$ (i.e., the slope of the best linear approximation to the curve in the right panel of Figure 8). Under nonlinearities, $\delta_+(H_{\max})$ varies with the truncation point $H_{\max}$. The bias of the estimator of β is $\delta_0 - \delta_+(H_{\max})$ when Assumption 1 is violated. Because δ_0 is always identified at the bunching point (see equation (6)), it does not vary with $H_{\max}$. Consequently, if nonlinearities are empirically important, the estimates of β should vary with $H_{\max}$.

[30] Using notation analogous to Footnote 29, the contribution of cluster k to the bias is proportional to

different ranges of H_i^*, and allowing α and δ to vary across clusters absorbs some of this nonlinear variation. Thus, if nonlinearities were empirically consequential, the estimate of β would tend to change as we allow for increasingly flexible specifications where α and δ vary across clusters, since the portion of variation in the outcome attributed to selection rather than treatment would increase. Reassuringly, in Section 4.2.3, we find that the estimates of β are stable as the number of clusters grows from $K = 1$ to $K = 50$.

Together, these exercises provide quite a comprehensive set of diagnostics for assessing whether the linearity assumption is a good approximation in our context (and, in turn, whether the weaker equal-sign assumption introduced in Remark 3.1 is likely to hold). Note that the truncation and clustering exercises are complementary in their coverage of H_i^*. The truncation exercise is informative about any nonlinearity located in the range $H_i^* > 0$, regardless of its relationship to X_i, because it varies the portion of that range used for identification. The clustering exercise exploits variation across clusters defined by X_i. Because these clusters are associated with different ranges of H_i^*—including ranges with $H_i^* < 0$—the exercise can detect nonlinearities in regions that the truncation exercise cannot. Taken together, these diagnostics imply that a nonlinearity escaping detection by both exercises would have to be confined to values of $H_i^* < 0$ that are *not* differentially associated with X_i. Accordingly, for Assumption 1 to fail in a way that meaningfully affects our conclusions, the nonlinearity would have to be simultaneously (i) large enough to materially affect the estimates; (ii) confined to $H_i^* < 0$, so that it escapes the truncation exercise in Section 4.2.2; and (iii) unrelated to the observed characteristics X_i, so that it also escapes detection by the clustering exercise in Section 4.2.3. We view this combination as unlikely in our setting.

4 Results

4.1 Main Results

In Table 7, we present results for the full sample. The first two columns show estimates from linear regression models without addressing the potential endogeneity issue. Column (1), which does not control for any observed covariates, shows that both cognitive and noncognitive skills are negatively correlated with time spent watching video. Column (2), which adds our full set of controls, shows that observed covariates explain part of the correlation between video watching time and cognitive skills but the relationship remains negative and statistically significant. For noncognitive skills, when we include the full set of controls, the

$\delta_0(k) - \delta_+(k)$, where $\delta_0(k)$ is the selection effect at $H_i = 0$ and $\delta_+(k)$ is the selection effect at $H_i > 0$ among observations in cluster k (see equation (6)). Both $\delta_0(k)$ and $\delta_+(k)$ vary across clusters because the distribution of H_i^* differs across clusters.

correlation is close to zero and insignificant.

Table 7: The Effect of Daily Weekday Video Watching on Cognitive and Noncognitive Skills

Dependent Variables	(1) Uncorrected No Controls	(2) Uncorrected With Controls	(3) Bunching Approach
Cognitive Skill	–0.034***	–0.021***	0.026
(N = 5,274)	(0.009)	(0.006)	(0.020)
Noncognitive Skill	–0.019**	–0.0004	–0.091***
(N = 5,706)	(0.008)	(0.008)	(0.033)

Notes: Each cell shows the effect of an additional hour of weekday video watching on children's skills. Bootstrapped standard errors are in parentheses (500 iterations). Columns (2) and (3) include the control variables described in Section 2. Column (3) uses the tail symmetry assumption. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Column (3) presents the estimates of β using our bunching approach. For cognitive skills, the point estimate is positive but not statistically significant. The pattern of results for cognitive skills—from a negative and significant coefficient in column (1) to a negative and significant but attenuated coefficient in column (2) to a positive and insignificant coefficient in column (3)—suggests selection on observables and unobservables matter and operate in the same direction. That is, selection on observables and unobservables are both negative for cognitive skills. These results are consistent with our findings in the left panels from Figures 5 and 6, since the negative discontinuity reduces in magnitude when controls are added.

For noncognitive skills, the point estimate in column (3) is negative and significant (p-value < 0.01) and implies that an additional hour of watching video on a weekday reduces noncognitive scores by 0.091 standard deviations. The shift from a negative coefficient in column (1) to a coefficient close to zero and not statistically significant in column (2) to a coefficient in column (3) that is statistically significant, negative, and larger in magnitude compared to that in column (1) suggests that selection on unobservables matters and does so in the opposite direction as selection on observables. That is, selection on observables is negative, whereas selection on unobservables is positive for noncognitive skills. The fact that selection on observables and on unobservables operate in opposite directions in the noncognitive skill case is consistent with our findings in the right panels from Figures 5 and 6, since the positive discontinuity increases when controls are added.

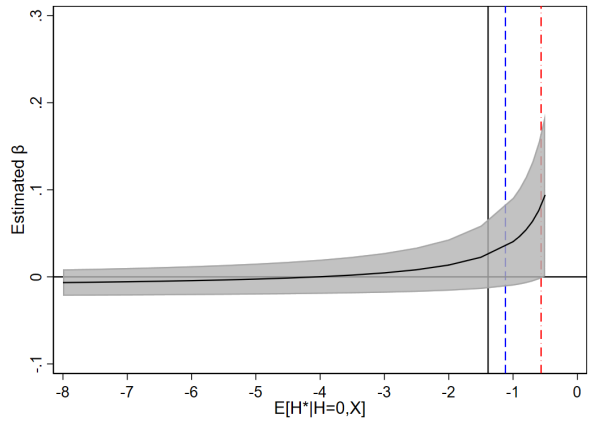
4.2 Robustness

4.2.1 No Distributional Assumption

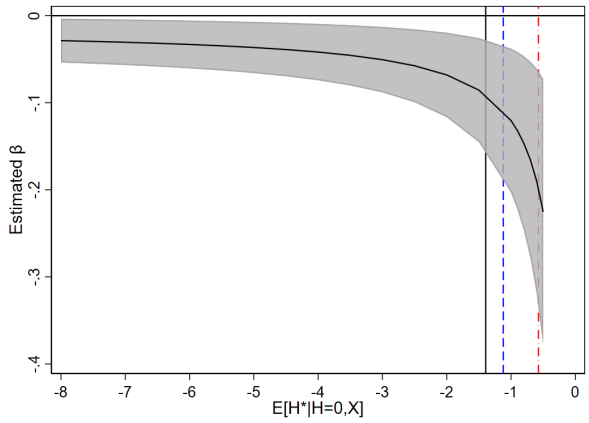
In the results presented in column (3) of Table 7, we imposed the tail symmetry assumption to identify $\mathbb{E}[H_i^*|H_i = 0]$, as described in Assumption 2. We next provide results under no distributional assumptions. Figure 9 shows different values of $\hat{\beta}$ under different possible values of $\mathbb{E}[H_i^*|H_i = 0]$. For reference, we add vertical lines that correspond to the estimates of the expectation under tail symmetry (solid black), normal (dashed blue), and uniform (dot-dashed red) distributional assumptions. Our finding of an insignificant effect of video watching time on cognitive skills generally holds across different possible values of the expectation. If anything, there is some scope for the effect of an additional hour of video watching to be more positive if the expectation is closer to zero than that implied by the tail symmetry assumption. For noncognitive skills, there is more scope for β to be more negative than the estimate in column (3) of Table 7 and little scope for it to be less negative. In other words, there is more scope for the tail symmetric results to underestimate the magnitude of β than overestimate it. In the remaining results, we report the estimates using the tail symmetry assumption as it reflects the distribution that yields the most conservative estimates among the options we considered.

Figure 9: Results Under No Distributional Assumption

(a) Cognitive Skills



(b) Noncognitive Skills



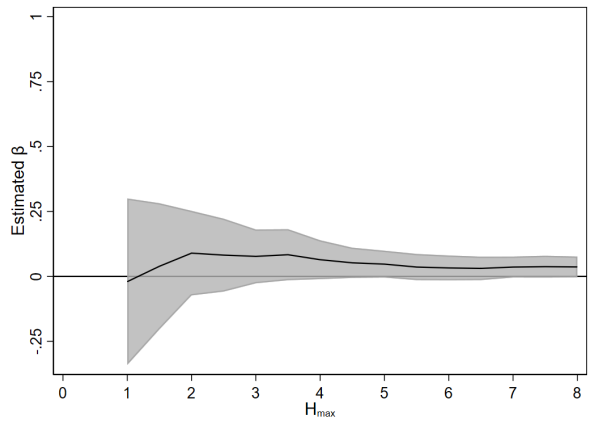
Notes: Each panel shows different values of $\hat{\beta}$ under different possible values of $E[H_i^*|H_i = 0]$, along with 95% confidence intervals (using 500 bootstrap iterations). For context, we also show the vertical lines implied by the expectation estimated under tail symmetry (solid black), normal (dashed blue) and uniform (dot-dashed red) distributions.

4.2.2 Local vs. Global Linearity Assumption

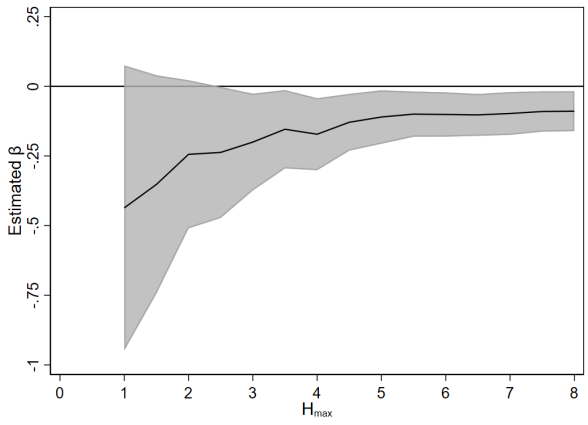
The linearity assumption, Assumption 1, restricts the selection effect to be the same across different values of H_i^*. If there are confounding factors that vary nonlinearly with H_i^* (and if the nonlinearity of other confounding factors does not offset that nonlinearity), then the linearity assumption may not be appropriate. Figure 6 above already suggested that the linearity assumption is reasonable because the plots for $H_i > 0$, which allow for nonlinearities, tend to be approximately linear, except for very high values of H_i which are too rare to matter empirically. Next, we provide more direct evidence by presenting results from a series of truncated samples. That is, we restrict the sample to observations with $H_i < H_{\max}$ and progressively increase the value of $H_{\max}$. As discussed in Section 3.4.3, varying $H_{\max}$ changes the portion of the support of H_i^* used for identification. If empirically important nonlinearities are present over $H_i^* > 0$, the implied bias of the estimator should change with $H_{\max}$, and the estimated values of β should therefore vary as $H_{\max}$ increases. In Figure 10, we show how the estimates of β change for different truncations of the sample ranging from $H_{\max} = 1$ to $H_{\max} = 8$.[31] For cognitive skills, the estimates of β are generally quite similar across values of $H_{\max}$, though confidence intervals widen with smaller values of $H_{\max}$. For noncognitive skills, point estimates are always negative, but they tend to be more intense and less precisely estimated for low values of $H_{\max}$.

Figure 10: Results for Different Truncations of the Data by $H_{\max}$

(a) Cognitive Skills



(b) Noncognitive Skills



Notes: Each panel shows different values of $\hat{\beta}$ from different samples with $H_i < H_{\max}$, along with 95% confidence intervals using the tail symmetry assumption (500 bootstrap iterations).

[31] As $H_{\max}$ changes, we maintain the same estimate of $\mathbb{E}[H_i^*|H_i = 0]$ using the tail symmetry assumption. In this way, any change in $\hat{\beta}$ as $H_{\max}$ changes reflects nonlinearities.

4.2.3 Relaxing Specification of Controls to Relax Assumptions 1 and 2

We next use controls more flexibly to further relax Assumptions 1 and 2. We rewrite equation (4) as:

$$\mathbb{E}[S_i|H_i, X_i] = \alpha(X_i) + \beta H_i + \delta(X_i)(H_i + \mathbb{E}[H_i^*|H_i = 0, X_i]\mathbf{1}(H_i = 0)), \tag{10}$$

where three quantities now depend on the controls X_i: $\alpha(X_i)$, $\delta(X_i)$ and $\mathbb{E}[H_i^*|H_i = 0, X_i]$. We specify each of these quantities nonparametrically using hierarchical clusters.

To fix ideas, when X_i is discrete and takes a finite number of values, $X_i \in \{x_1, ..., x_K\}$, then it is straightforward to estimate each quantity separately for each value of X_i using fully saturated specifications:

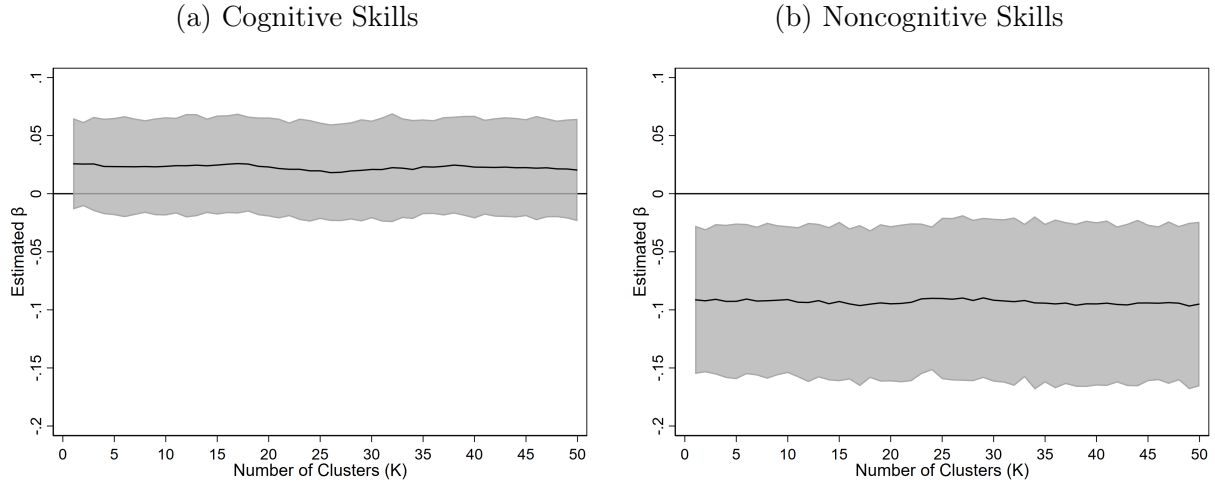
1. $\alpha(X_i) = \sum_{k=1}^{K} \alpha_k \mathbf{1}(X_i = x_k)$.
2. $\delta(X_i) = \sum_{k=1}^{K} \delta_k \mathbf{1}(X_i = x_k)$.
3. $\mathbb{E}[H_i^*|H_i = 0, X_i] = \sum_{k=1}^{K} \mathbb{E}[H_i^*|H_i = 0, X_i = x_k]\mathbf{1}(X_i = x_k)$.

However, X_i takes too many values to implement this strategy. Thus, following Caetano et al. (2024b), we discretize X_i by assigning observations with similar values of X_i to the same cluster, and then treat each cluster k as if it had a distinct value x_k, as in the discrete case discussed above. This means, for example, that we allow for α_k to vary only across clusters, but not across different values of X_i within the same cluster. Although observations within a given cluster may differ in their values of X_i, this within-cluster variation becomes smaller as the number of clusters allowed increases. We also allow for a linear term in $\alpha(X_i)$ (i.e., $X_i'\alpha$) to absorb some of this residual within-cluster variation in our empirical specifications.

There are many methods for grouping observations according to the similarity of their values of X_i. We use hierarchical clustering (with the Gower measure of distance and Ward's linkage) for its stability and ease of interpretation as we vary the number of clusters (Hastie, Tibshirani, and Friedman 2009). The results reported in the main text thus far correspond to $K = 1$, so controls enter only linearly in $\alpha(X_i)$. Below, we show that allowing $\alpha(X_i)$, $\delta(X_i)$ and $\mathbb{E}[H_i^*|H_i = 0, X_i]$ to vary across clusters has little effect on our results, even when we allow for a large number of clusters.

Figure 11 shows that our estimates of β are stable irrespective of the value of K. To understand why this stability provides evidence in favor of Assumptions 1 and 2, we argue below that increasing the number of allowed clusters from $K = 1$ to $K = 50$ progressively weakens both assumptions.

Figure 11: Results for Different Numbers of Allowed Clusters in Specification of $\alpha(X_i)$, $\delta(X_i)$, and $E[H_i^*|H_i = 0, X_i]$



Notes: Estimates of β using the tail symmetry assumption and corresponding 95% confidence intervals shown for different numbers of clusters used in the analysis (500 bootstrap samples).

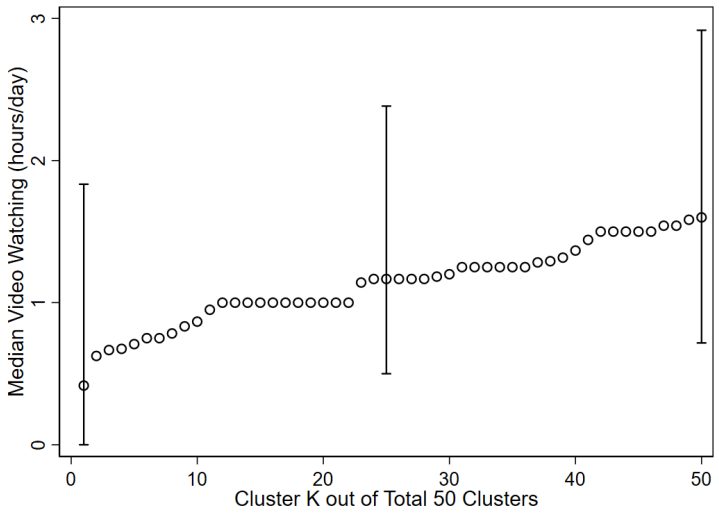
To understand how the linearity assumption (Assumption 1) is relaxed as we allow the number of clusters to increase, recall that the linearity assumption, $\mathbb{E}[U_i|H_i^*] = \alpha + \delta H_i^*$, restricts U_i to vary with H_i^* in a linear fashion. When we instead specify $\mathbb{E}[U_i|H_i^*, X_i] = \alpha(X_i) + \delta(X_i)H_i^*$, we allow both α and δ to change with covariates X_i. Because different clusters of X_i are associated with different ranges of H_i^*, allowing $\alpha(X_i)$ and $\delta(X_i)$ to vary across clusters indirectly allows them to vary with H_i^*, thereby relaxing the linearity assumption (as discussed in Section 3.4.3).

Figure 12 shows evidence that different clusters of X_i are indeed associated with different ranges of H_i^*. The figure considers the case of $K = 50$ clusters and shows the median of $H_i^* = H_i$ for each of the 50 clusters, sorted by their median, as well as the interquartile distribution of H_i for three selected clusters.[32] The median of H_i^* tends to change across clusters. In this sense, the clusters can be viewed as grouping observations from different portions of the support of H_i^*, with $\alpha(X_i)$ and $\delta(X_i)$ partially capturing differences in the average and slope of selection across these ranges of H_i^*, thereby partially absorbing nonlinearities. Thus, as we increase the number of clusters from $K = 1$ (where α and δ are not allowed to change across X_i) to $K = 50$ (where α and δ are allowed to change across the 50 clusters of X_i), we relax the linearity assumption. And yet, Figure 11 shows that the main estimates barely change when we do so.[33]

[32]Note that although H_i^* is not directly observed, the median of H_i^* is observed in our case, since there is less than 50% bunching across all clusters.

[33]Because all elements of X_i are pre-determined with respect to H_i, increasing K from 1 to 50 clusters accommodates nonlinearities only in the selection component of the outcome, not in the causal component

Figure 12: Evidence of Heterogeneity in H_i^* by Cluster



Notes: This plot shows the median within cluster for each of the clusters when $K = 50$. Clusters are sorted from left to right based on the median of $H_i^* = H_i$ within cluster. The interquartile ranges within cluster for selected clusters are also shown.

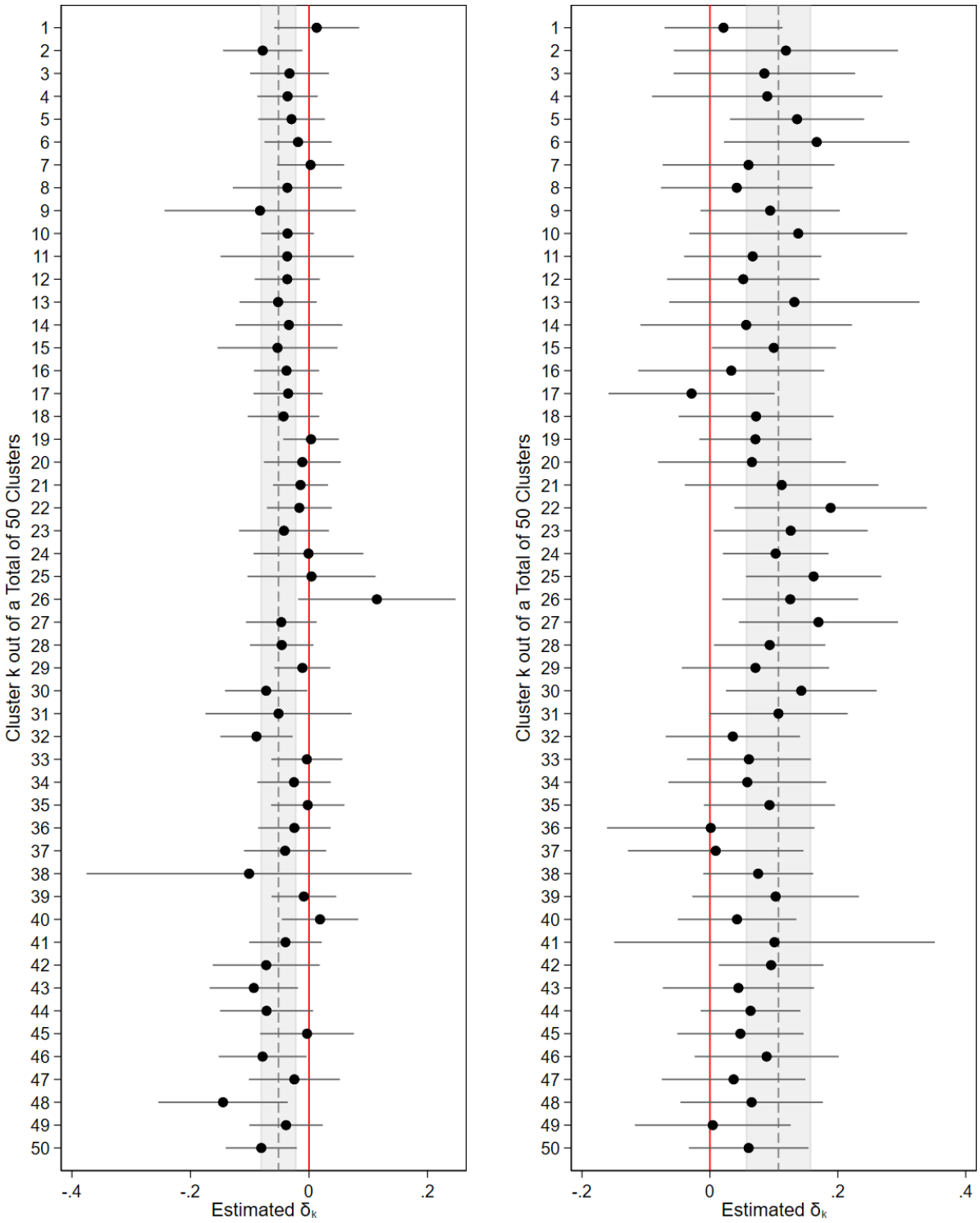
The previous figure showed that clusters of X_i correspond to different regions of the support of H_i^*. Next, we explore whether the estimated selection effect differs across those clusters, and therefore, indirectly across those regions of H_i^*. Figure 13 reports the cluster-specific estimates of the selection effect, δ_k, one for each of the $K = 50$ clusters. The dashed vertical line indicates the estimate of δ from the baseline specification with $K = 1$, and the shaded band shows its 95% confidence interval. The left panel presents the results for cognitive skills, and the right panel presents the results for noncognitive skills. We order the clusters from the lowest to the highest median of H_i^*, as in Figure 12. There is little evidence that δ_k systematically varies across clusters in either the cognitive or noncognitive case, suggesting that indeed the linearity assumption is approximately valid. Figure 14 reinforces this conclusion by showing the analogous results for $K = 10$, where the cluster-specific estimates of δ_k are naturally more precise.

itself. Thus, these results rule out the knife-edge case discussed in Footnote 23.

Figure 13: Estimates of the 50 δ_ks for $K = 50$

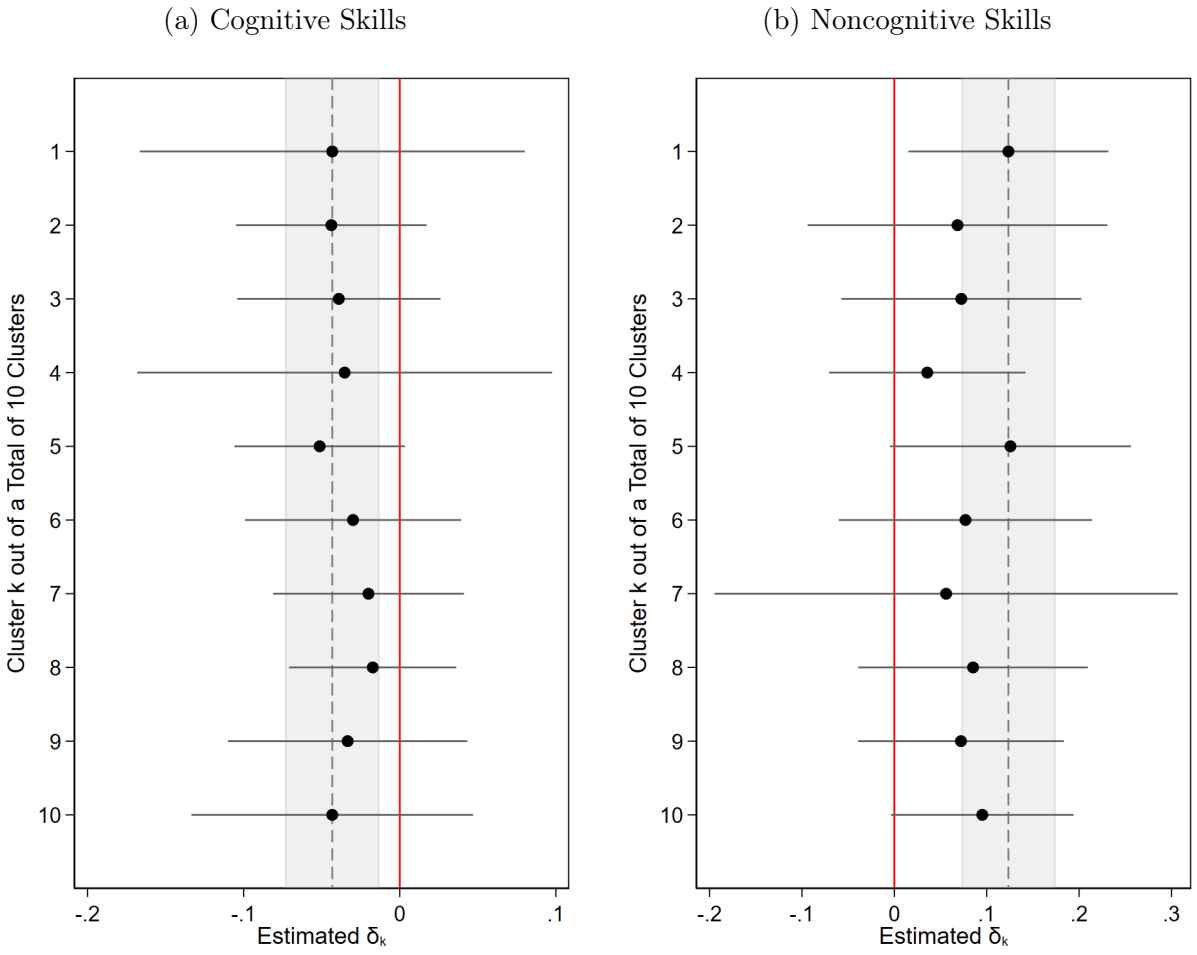
(a) Cognitive Skills

(b) Noncognitive Skills



Notes: The plots show the cluster-specific estimates of the selection effect δ_k, one for each of the $K = 50$ clusters, under the tail symmetry assumption. 95% confidence intervals are based on 500 bootstrap samples. The dashed vertical line indicates the estimate of δ from the baseline specification with $K = 1$, and the shaded band shows its 95% confidence interval. The 50 clusters are ordered from lowest to highest median of $H_i^* = H_i$, as in Figure 12.

Figure 14: Estimates of the 10 δ_ks for $K = 10$



Notes: The plots show the cluster-specific estimates of the selection effect δ_k, one for each of the $K = 10$ clusters, under the tail symmetry assumption. 95% confidence intervals are based on 500 bootstrap samples. The dashed vertical line indicates the estimate of δ from the baseline specification with $K = 1$, and the shaded band shows its 95% confidence interval. The 10 clusters are ordered from lowest to highest median of $H_i^* = H_i$, as in Figure 12.

To understand how the distributional assumption (Assumption 2) is relaxed as we allow the number of clusters to increase, note that *the whole distribution* of H_i^* can vary across clusters. Indeed, Figure 12 shows that not only the median, but also other features of the distribution change across clusters. For instance, the interquartile range is smaller for clusters 1 and 25 than for cluster 50. The median is closer to the halfway point of the interquartile range for cluster 25 than for cluster 1, showing that even the degree of asymmetry within cluster may vary across clusters. As we increase the number of clusters from $K = 1$ to $K = 50$, we replace a single distributional assumption for the entire sample with separate distributional assumptions within increasingly homogeneous clusters. That is, we allow $\mathbb{E}[H_i^*|H_i = 0, X_i]$ to vary across clusters. Consequently, asymmetries in the distribution of H_i^*

that are associated with X_i are allowed to differ across clusters rather than being constrained by a common distributional assumption for the full sample. In this sense, increasing the number of clusters relaxes Assumption 2. Thus, the stability of the estimates of β across values of K provides further evidence that our results are not an artifact of the distributional assumption.

4.2.4 Further Robustness

We estimate several specifications to further probe the robustness of our results. First, we consider alternative treatment definitions that aggregate video watching time over different horizons, namely daily video watching during the whole week as well as daily video watching on weekends. Second, we estimate the effect of video watching during weekdays on two different subsamples, one using the 1997, 2002, and 2007 CDS waves and the other using the 2014 and 2019 CDS waves. We split the sample in this way because the 2014 and 2019 CDS waves follow a different cohort of children than the earlier CDS waves, there was a wording change for the description of video watching in the time diary in the 2014 and 2019 CDS waves, and video watching on devices other than television sets becomes particularly more common in the 2010s. Third, we estimate effects of weekday video watching across samples that differ in how typical the weekday is deemed to be. Respondents describe each diary day on a scale of 1 ("very typical") to 5 ("not at all typical"). In this robustness exercise, we first estimate our model only on those whose weekday time diary was described as "very typical" and incrementally add observations with diaries that are less and less representative of a typical weekday. At the extreme, we include those who describe their weekday diary as "not at all typical."[34] Last, we estimate regression models including additional controls, namely lagged skill and lagged time spent watching video. Having shown that our results are largely unchanged when the number of clusters is increased, all robustness exercises below use the baseline specification, which relies on the tail symmetry assumption and a single cluster ($K = 1$). Results are presented in Figure 15. To facilitate comparisons with our main results, we show the point estimate from column (3) in Table 7 as a horizontal dashed line, with the 95% confidence interval shaded.

We generally find near zero or positive statistically insignificant effects of video watching time on cognitive skills when we consider daily video watching during the weekend, across the whole week, or separately by early versus later survey waves. For noncognitive skills, the

[34]Our baseline sample includes observations that describe neither their weekday nor weekend diary as "not at all typical." For the exercise in this section where we incrementally change the weekday diaries that are included, we do not put any restrictions on the weekend diary. Furthermore, for these analyses, we standardize the skill measures using the mean and standard deviation from our main analysis samples to maintain comparability of the estimates.

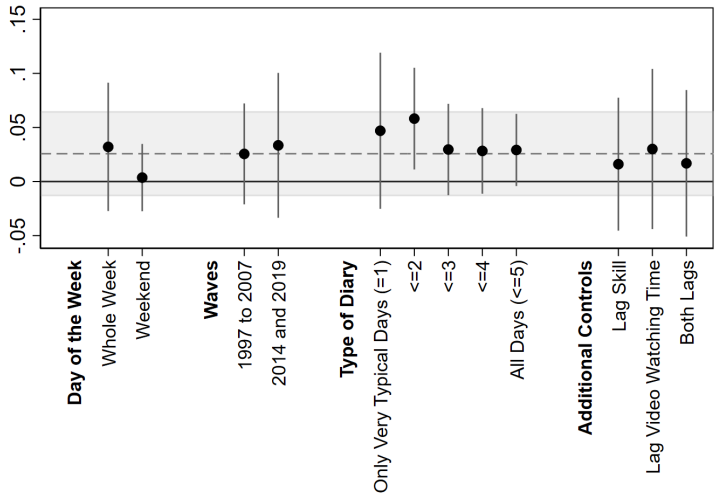
effect of daily weekend video watching is negative and statistically significant but somewhat smaller in magnitude than the effects of daily video watching on a weekday or during the whole week. The effects in the later waves are modestly smaller in magnitude and less precisely estimated compared to the early waves, though we cannot reject equality of the estimates at conventional significance levels. The general stability of the effects across the early and later survey waves is notable given the changes in sample composition, content, technology, and the way individuals access audiovisual content over time.[35]

Thus far, we have focused on the effects of daily video watching for those with a *typical* diary week, defined as one in which both the weekday and weekend days were *not* described as "not at all typical." Our estimated effects of an additional hour of video watched are robust to a range of choices, from limiting the sample to those with weekday diary days described only as "very typical" to including even those with "not at all typical" weekday diary days. The stability of the results may be due to the fact that respondents were asked to indicate whether the *day* was typical and not whether time spent on a particular *activity* was typical. Even on a very atypical day (perhaps because of irregular sleep or school-related activities that more strongly influence how respondents judge the day's typicality), video watching time may remain fairly typical. Overall, the stability of our results irrespective of how stringent is the sample restriction related to the typicality of the day suggests that there are likely no major issues related to measurement error of the treatment variable. We discuss more subtle potential measurement error issues in Section 4.2.5.

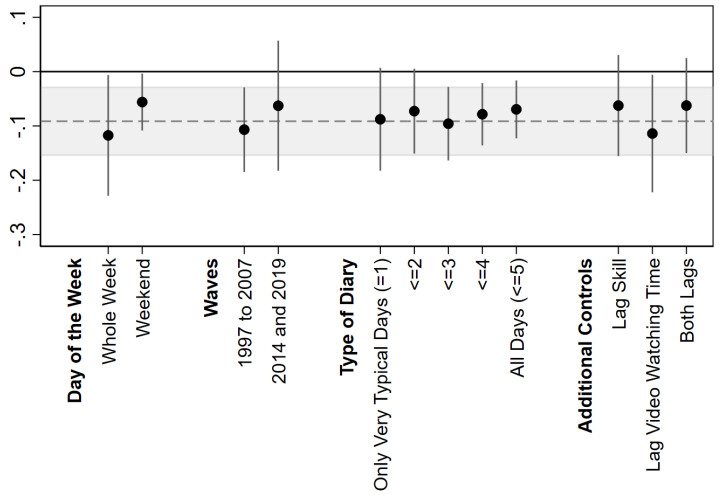
[35]We note that the age distribution of children differs in each wave. In Section 5.2, we consider effects among similarly aged children when we examine effects by grade level (elementary, middle, and high school).

Figure 15: Robustness of the Effects of Daily Video Watching on Skill

(a) Cognitive Skills



(b) Noncognitive Skills



Notes: Point estimates and corresponding 95% confidence interval bars shown using the tail symmetry assumption (500 bootstrap samples). We show the point estimate from column (3) in Table 7 as a horizontal dashed line, with the 95% confidence interval shaded.

Last, for children observed more than once in our sample, we exploit the panel aspect of the data and estimate specifications using lagged measures as additional controls. We show results adding just the lag of the outcome variable (lagged skill measure) as a control, just the lag of the treatment variable (lagged time spent watching video) as a control, and both lagged variables as controls. This exercise is helpful to ascertain that the bunching approach leaves no remaining potential confounders that are correlated with the lagged versions of the treatment or the outcome. The adverse noncognitive effects of daily weekday video

watching are similar to our main estimate, though less precisely estimated. The decreased precision is not surprising given we only include the children observed more than once in the sample. Furthermore, we cannot reject equality of the estimates with and without these lagged measures at conventional significance levels.

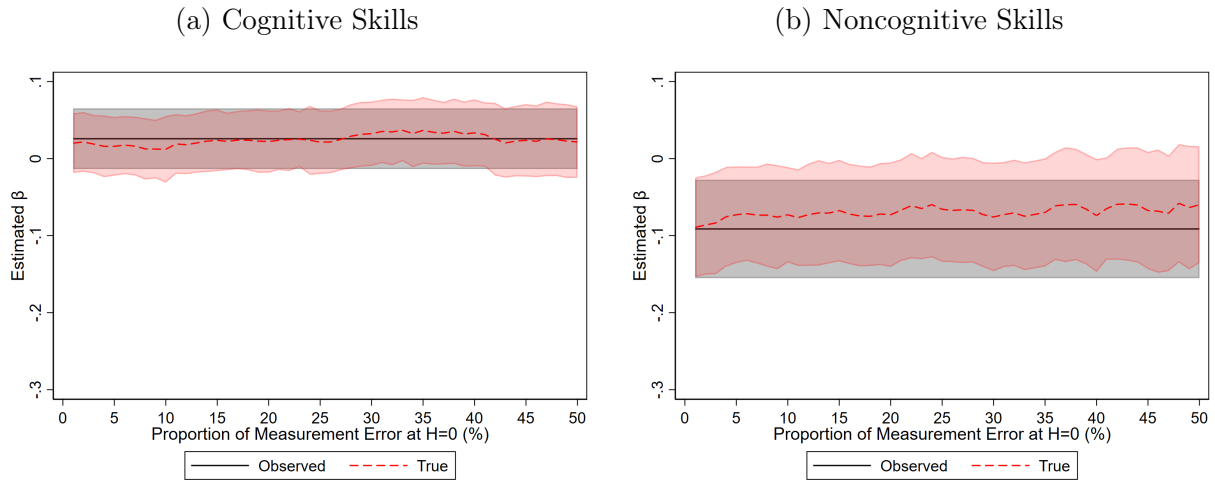
Overall, these exercises show that our results are quite robust to a variety of alternative treatment definitions, samples, and controls.

4.2.5 Measurement Error

In this section, we consider the possibility of measurement error in our treatment variable H_i. As we have discussed, we restrict our analysis to time diary data for days that are not deemed to be "not at all typical" by the respondent. Above, we conducted a sensitivity analysis where we restricted the sample to only days deemed to be "very typical," obtaining similar results. While that exercise suggests that major measurement error issues are unlikely, smaller issues may still remain.

Here, we consider whether our results are robust to what we believe to be the most plausible source of measurement error: some of the children observed in the data at $H_i = 0$ might in reality spend some small positive amount of time watching video. Denote $H_{i,\text{True}}$ the true treatment variable—the amount of video typically watched by the child in reality—and H_i our observed treatment variable. We simulate the true data that differ slightly from the observed data by allowing $H_{i,\text{True}} \neq H_i$ for a subset of observations, and then re-estimate our models using this modified dataset. Specifically, for a randomly selected proportion of individuals with $H_i = 0$, we assign $H_{i,\text{True}}$ values drawn from a uniform distribution between 0 and 1 hour per day, and then conduct the whole estimation and inference procedure discussed in Section 3 for $H_{i,\text{True}}$ instead of H_i.

Figure 16: Robustness to Measurement Error



Notes: Each panel shows different values of $\hat{\beta}$ under different measurement error scenarios (dashed red line), along with 95% confidence intervals (shaded red) using the tail symmetry assumption (500 bootstrap iterations). We also show the value of $\hat{\beta}$ from our observed data (solid black line) with the corresponding 95% confidence interval (shaded gray) using the tail symmetry assumption. The value on the horizontal axis refers to the percentage of observations at $H_i = 0$ assumed to be mismeasured in each scenario, with $H_{i,\text{True}}$ randomly drawn from the interval $(0, 1)$.

Figure 16 shows how our estimates of β change if we use $H_{i,\text{True}}$ as our treatment variable instead of H_i, varying the proportion of observations at $H_i = 0$ that are measured with error from 1% to 50%. For both cognitive and noncognitive skills, the results are fairly stable as the proportion of mismeasured observations grows, suggesting that this kind of measurement error is not consequential to our conclusions.

How is it possible that our estimates do not change much with this measurement error? As explained in detail in Appendix B, this measurement error leads to a violation of both Assumptions 1 and 2. The violations of both assumptions lead to biases in different directions that tend to offset each other, which explains why the β estimates are relatively stable as the measurement error becomes more prevalent. Specifically, in our setting, the violation of the linearity assumption will intensify our β estimate, while the violation of the distributional assumption will attenuate our β estimate. Appendix B discusses in detail the violations of both assumptions and why they tend to offset each other.

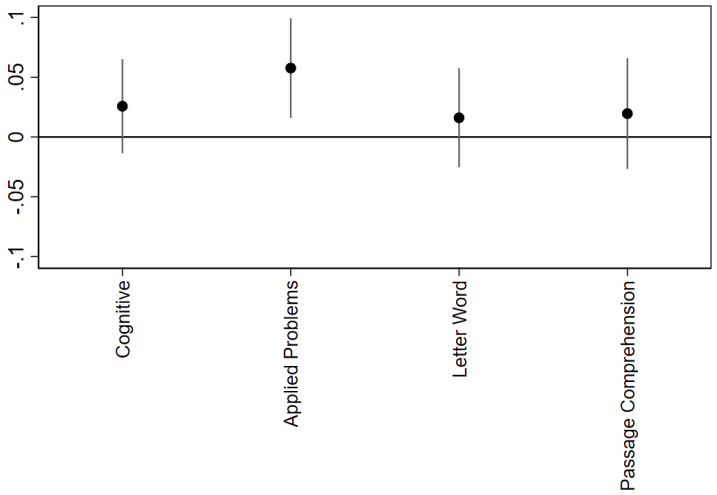
5 Heterogeneous Results

Unless otherwise noted, all estimates in this section use the baseline specification, which relies on the tail symmetry assumption and a single cluster ($K = 1$).

5.1 Effects on Alternative Measures of Skills

In this section, we estimate the effect of video watching time on the components of cognitive skill—letter word, applied problems, and passage comprehension tests. Each component has been standardized to have mean zero and standard deviation equal to one. Results are shown in Figure 17. Point estimates are positive for all skill components, but are not statistically significant in the case of the letter word and passage comprehension components. We find a positive and statistically significant effect on the applied problems component. An additional hour of video watched on a weekday increases performance on the applied problems test by 0.062 standard deviations.

Figure 17: Effects of Daily Weekday Video Watching on Other Measures of Cognitive Skills

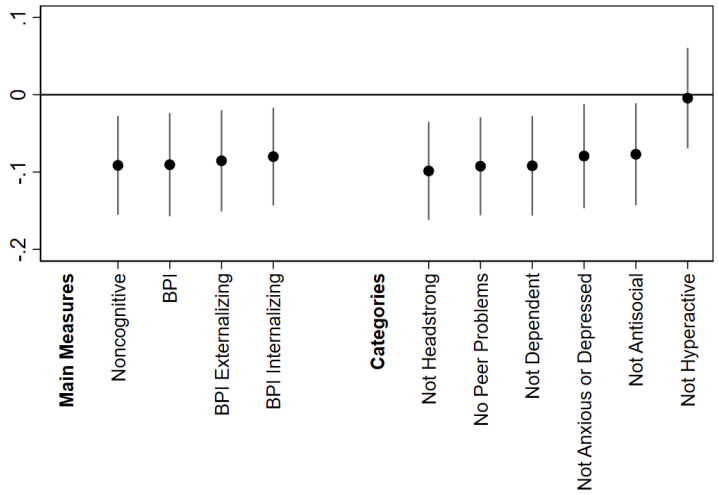


Notes: Point estimates and 95% confidence interval bars shown using the tail symmetry assumption (500 bootstrap samples).

In Figure 18, we present the estimated effects of an additional hour of video watching on alternative measures of noncognitive skill. We first consider the Behavior Problem Index (BPI) and its internalizing and externalizing subscales described in Section 2, where, as before, higher values indicate better development, and each subscale is normalized to have mean zero and standard deviation equal to one. An additional hour spent watching video adversely impacts both the internalizing (-0.080 s.d.) and externalizing (-0.085 s.d.) measures. We next consider an alternative form of aggregation of the 32 questions answered by parents about their child's behavior following Zill (1990). In particular, we examine six domains: not antisocial, not anxious or depressed, not headstrong, not hyperactive, not immaturely dependent, and no peer conflict. For ease of comparison with the results presented so far, we normalized these outcomes like the previous ones. We again find that time spent watch-

ing video negatively affects noncognitive skill, especially not headstrong (-0.098 s.d.), no peer conflict (-0.092 s.d.) and not immaturely dependent (-0.092 s.d.). Overall, the results suggest that increased time spent watching video has a robust negative effect on a range of noncognitive skill measures. These adverse effects seem to be concentrated on measures related to mood and interpersonal skills, rather than ability to concentrate or hyperactivity.

Figure 18: Effects of Daily Weekday Video Watching on Other Measures of Noncognitive Skills



Notes: Point estimates and 95% confidence interval bars shown using the tail symmetry assumption (500 bootstrap samples).

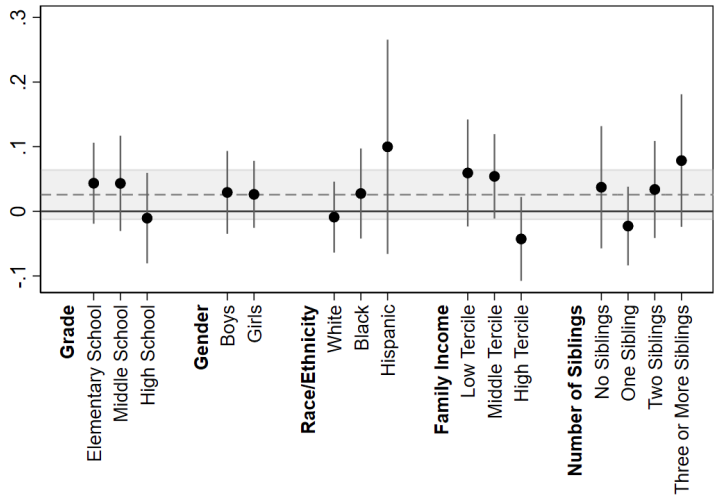
5.2 Treatment Effect Heterogeneity by Child and Family Characteristics

We examine effect heterogeneity by characteristics of the children, including grade level (elementary, middle, and high school), gender, and race/ethnicity (Non-Hispanic White, Non-Hispanic Black, and Hispanic).[36] The results are presented in the first columns of Figure 19.

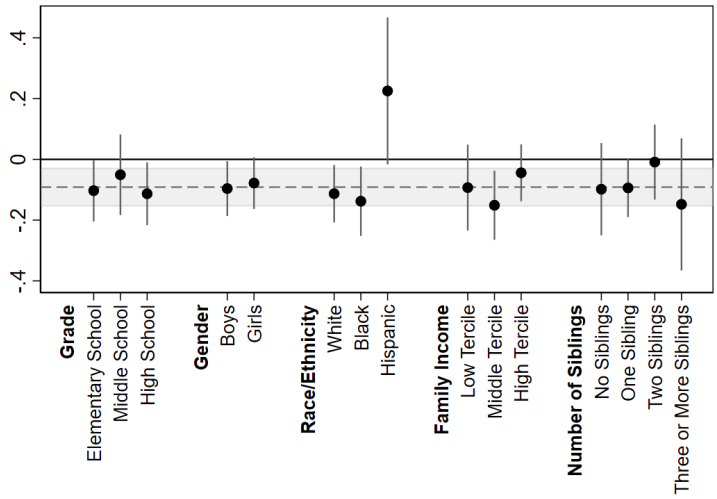
[36]We do not have a large enough sample to estimate the model on children of "other" race or ethnicity.

Figure 19: Effects of Daily Weekday Video Watching by Child and Family Characteristics

(a) Cognitive Skills



(b) Noncognitive Skills



Notes: Point estimates and 95% confidence interval bars shown using the tail symmetry assumption (250 bootstrap samples). We show the point estimate from column (3) in Table 7 as a horizontal dashed line, with the 95% confidence interval shaded.

We find similar positive but insignificant effects on cognitive skill for children in elementary and middle school, but the point estimate for high school students is very close to zero. The effects of an additional hour of video watching on noncognitive skill are more similar across grade levels, though they are not statistically significant for children in middle school. The effects on both cognitive and noncognitive skills are remarkably similar across gender. Compared to non-Hispanic White and Black children, the effects of video watching on Hispanic children are substantially more positive for both cognitive and noncognitive

skills.

We also consider effect heterogeneity across characteristics of the child's family, including household income tercile and number of biological siblings in the household. Results are shown in the remainder of Figure 19. For cognitive skills, the effects are positive but not significant for children with household income in the lowest and middle terciles, while they are negative and insignificant for children from high-income families. We can formally reject that the effects among children with household income in the third tercile are the same as those with income in the first and middle terciles at the 10% significance level. We generally find negative effects on noncognitive skills across the household income distribution and cannot reject equality of the estimates across those groups. Likewise, we find little evidence of systematic effect heterogeneity by the number siblings in the household for both cognitive and noncognitive skills.

Next, we zoom in to understand whether the positive effects for Hispanic children are heterogeneous depending on the specific components of cognitive and noncognitive skill. The results in Figure 20 suggest that the effect heterogeneity along race/ethnicity is very striking: while the results for non-Hispanic White and Black children tend to be similar, the results for Hispanic children are more positive in general. For this latter group, effects of video watching are stable and positive across the different dimensions of cognitive skills (applied problems, letter word, and passage comprehension) though not statistically significant. The negative effects on noncognitive skills are larger in magnitude and statistically significant for the externalizing scale, which is consistent with the idea that video watching facilitates cultural assimilation (Dalisay 2012 and Li and Tsai 2015). It is also worth mentioning that, relative to other races/ethnicities, Hispanics spend a higher proportion of their video watching time in the company of others (see Table C.1 in the Appendix). This pattern may help explain these findings if, all else constant, the effect of watching video on skills is particularly negative when done alone.

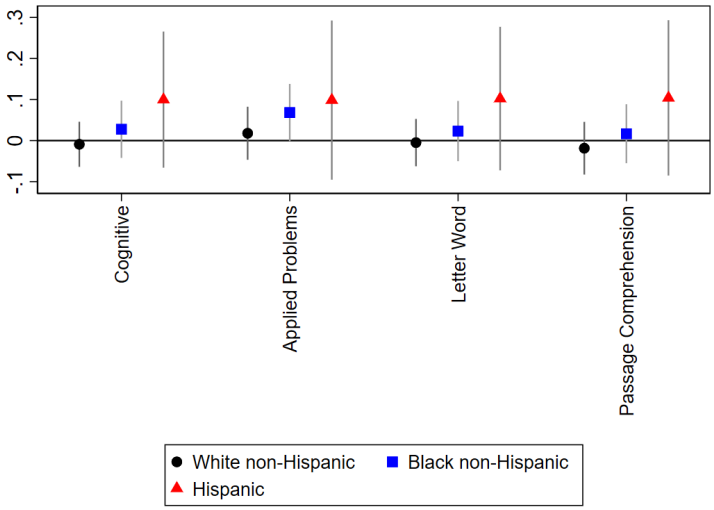
We also explore effect heterogeneity across household income terciles on the components of skill. The results in Figure 21 show that the effects on any particular component of cognitive skill are similar for children from low- and middle-income households, with the positive effects on applied problems being more pronounced and statistically significant. Children from high-income households generally see no statistically significant effect of an additional hour of video watching on the cognitive skill measures, which is consistent with the idea that this group might have a higher opportunity cost of watching video (since video watching is more likely to crowd out a cognitive-enhancing activity). For noncognitive skills, we tend to find less intense negative effects for low- and high-income children, while we find more intense (and significant) negative effects for middle-income children. Although we

cannot reject equality of the effects across the income terciles, this non-monotonic pattern aligns well with heterogeneous trade-offs shaped by opportunities and constraints (Caetano et al. 2024a). Relative to low-income children, video watching is more likely to crowd out activities that foster noncognitive skills among middle-income children.[37] Relative to high-income children, middle-income children may have less capacity to buffer or compensate for potential stressors—e.g., through parental resources, mentorship, access to knowledge or professionals to help with socioemotional support (Lareau 2003 and Mollborn et al. 2022). Unfortunately, the available data do not provide enough detail to examine these conjectures in depth, so investigating these potential mechanisms is left for future research.

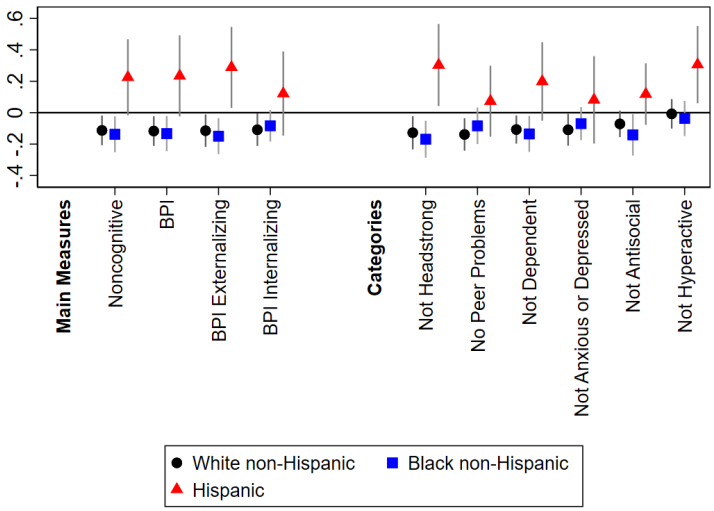
[37]Moreover, low-income children appear to spend a higher proportion of their video watching time in the company of others (see Table C.2 in the Appendix), which can help explain these findings if, all else constant, the effect of watching video on skills is particularly negative when done alone.

Figure 20: Effects of Daily Weekday Video Watching on Other Measures of Cognitive and Noncognitive Skills by Race/Ethnicity

(a) Cognitive Skills



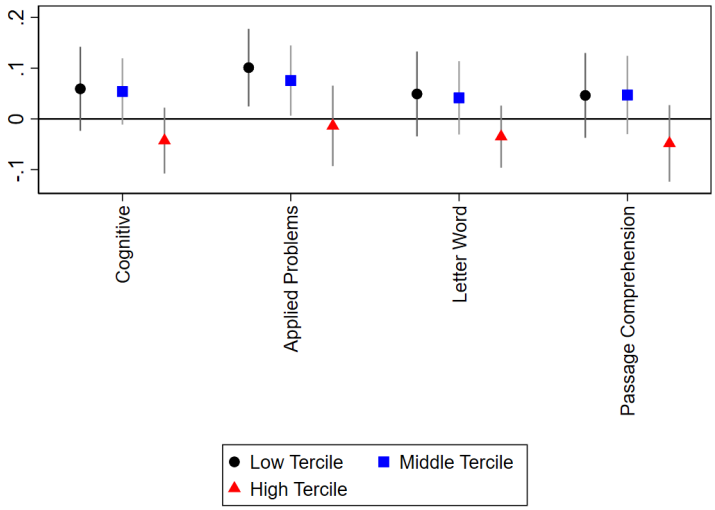
(b) Noncognitive Skills



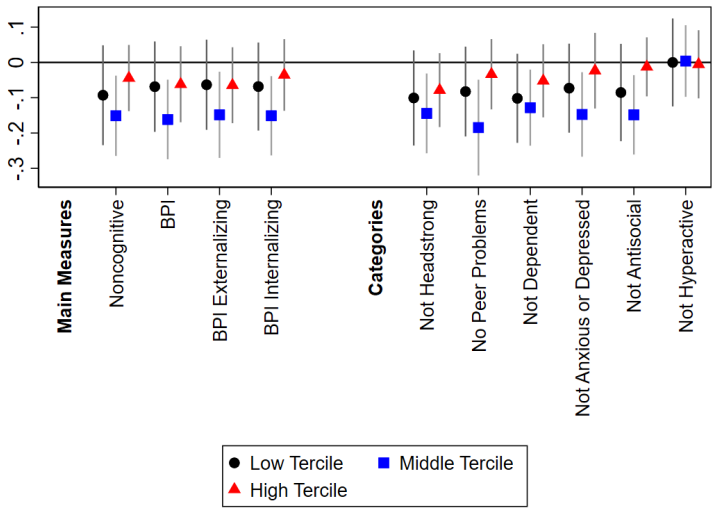
Notes: Point estimates and 95% confidence interval bars shown using the tail symmetry assumption (500 bootstrap samples).

Figure 21: Effects of Daily Weekday Video Watching on Other Measures of Cognitive and Noncognitive Skills by Income Tercile

(a) Cognitive Skills



(b) Noncognitive Skills



Notes: Point estimates and 95% confidence interval bars shown using the tail symmetry assumption (500 bootstrap samples).

5.3 Nonlinear Effects

Finally, we consider the possibility that the marginal effect of video watching hours on skills might be different depending on the number of hours the child watches video. Specifically, we allow for the average causal function to take a quadratic form instead of a linear form by rewriting equation (3) as

$$S_i = \alpha + \beta_0 H_i + \beta_H H_i^2 + \delta H_i^* + \epsilon_i, \qquad cov[H_i, \epsilon_i | H_i^*] = 0. \tag{11}$$

Here there are two causal coefficients instead of just one: β_0, representing the linear coefficient, as before, and β_H, representing the quadratic coefficient. This specification allows for convexity (if $\beta_H > 0$) or concavity (if $\beta_H < 0$).

The results of this exercise are shown in Table 8. We find no evidence of nonlinear effects, as $\hat{\beta}_H$ from the bunching approach is very small and not statistically significant. These results are in stark contrast to recommendations that screen time be capped. Such recommendations implicitly assume a nonlinear relation, whereby moderate video watching is deemed benign (or at least not harmful) but excessive video watching is viewed as harmful (e.g., Vandewater, Bickham, and Lee 2007 and CCM 2016).

Our finding that the average causal function is approximately linear is consistent with the idea that individuals might endogenously adjust along other behavioral margins, reducing the potentially harmful effects of intense video watching. As viewing time increases, they may pay less attention, multitask, or select less demanding content, thereby modulating their investment in cognitive engagement. This interpretation aligns with rational inattention models (Sims 2003 and Matějka and McKay 2015), in which agents optimally allocate limited attention, and with evidence that individuals adjust the intensity or quality of time inputs when faced with competing demands (Becker 1965 and Guryan, Hurst, and Kearney 2008). It also aligns with the psychological literature on self-regulation, which shows that media use often serves compensatory or affective balancing functions rather than being purely passive (Carver and Scheier 1998 and Reinecke and Hofmann 2016). These behavioral margins are difficult to observe in the available data, and exploring them in greater detail is left for future research.

Table 8: The Nonlinear Effect of Daily Weekday Video Watching on Cognitive and Noncognitive Skills

Dependent Variable		(1) Uncorrected No Controls	(2) Uncorrected With Controls	(3) Bunching Approach
Cognitive Skill	β_0	–0.121***	–0.028**	0.042
(N = 5,274)		(0.021)	(0.013)	(0.036)
	β_H	0.016***	0.001	–0.002
		(0.004)	(0.003)	(0.003)
Noncognitive Skill	β_0	–0.004	0.027	–0.086*
(N = 5,706)		(0.019)	(0.018)	(0.050)
	β_H	–0.003	–0.005*	0.000
		(0.003)	(0.003)	(0.004)

Notes: Bootstrapped standard errors in parentheses (500 iterations). Columns (2) and (3) include the control variables described in Section 2. Column (3) uses the tail symmetry assumption. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

6 Conclusion

This paper documents the patterns of video watching behavior of children ages 6–18 in the United States, and estimates the consequences of video watching to their skill development. We use the CDS-PSID, which provides detailed data on both video watching and several measures of cognitive and noncognitive skills. The descriptive patterns reveal that video watching is generally stable across demographics such as gender, age, race/ethnicity, household income, and number of siblings, with the average child watching around 1.5 hours daily on weekdays and 2.7 hours daily on weekends.

The findings from our bunching identification strategy reveal a nuanced picture. Video watching has a significant negative impact on children's noncognitive skills, but an important exception is observed among Hispanic children. For these children, video watching appears to have positive effects on noncognitive skills, especially externalizing behaviors. In terms of cognitive skills, we find some small, positive but mostly insignificant effects. The effects tend to be more positive for quantitative reasoning and for Hispanic children. That Hispanic children have more positive effects of video watching on skill development is consistent with

the idea that video watching promotes cultural assimilation. This is further supported by our finding that the effects on noncognitive skill concentrate on the externalizing scale, which is more associated with socialization. We also find suggestive evidence that the adverse effects of video watching on noncognitive skills are most pronounced among middle-income children. This pattern may reflect a trade-off: while they have enough opportunities (relative to low-income children) for video watching to crowd out activities that foster socioemotional development, they often lack the resources (relative to high-income children) to offset the resulting behavioral costs (e.g., through parental time, guidance, or professional support).

Importantly, we find no evidence of nonlinear effects of video watching on developmental outcomes. While this finding aligns with the American Academy of Pediatrics' most recent guidance that refrains from providing specific time limits on screen time (AAP 2025), it contrasts with historical recommendations and general intuitions frequently expressed in popular media, which suggest that screen time may be acceptable in moderation but becomes harmful beyond a certain threshold. Such views, however, typically overlook the possibility that children and families make additional endogenous adjustments along other behavioral margins to accommodate cognitive and time constraints. Further research with richer measures of engagement, content, and within-day attention allocation would help clarify the extent to which these endogenous adjustments explain the approximate linearity we observe.

References

AAP (2025). American Academy of Pediatrics: "Screen Time Guidelines". Available here. Accessed on 07/24/2025.

Angrist, J. D. and M. Rokkanen (2015). Wanna get away? regression discontinuity estimation of exam school effects away from the cutoff. *Journal of the American Statistical Association 110*(512), 1331–1344.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal 75*(299), 493–517.

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

Caetano, C., G. Caetano, and E. Nielsen (2024a). Are Children Spending Too Much Time on Enrichment Activities? *Economics of Education Review 98*, 102503.

Caetano, C., G. Caetano, and E. Nielsen (2024b). Correcting for Endogeneity in Models with Bunching. *Journal of Business & Economic Statistics 24*(3), 851–863.

Caetano, C., G. Caetano, E. Nielsen, and V. Sanfelice (2025). The Effect of Maternal Labor Supply on Children: Evidence from Bunching. *Forthcoming, Journal of Labor Economics*.

Caetano, G., J. Kinsler, and H. Teng (2019). Towards Causal Estimates of Children's Time Allocation on Skill Development. *Journal of Applied Econometrics 34*(4), 588–605.

Carver, C. S. and M. F. Scheier (1998). *On the Self-Regulation of Behavior*. New York: Cambridge University Press.

CCM (2016). Council on Communications and Media: "Media Use in School-Aged Children and Adolescents". *Pediatrics 138*(5), e20162592.

Colombo, K. and E. Failache (2023). Exposure to High-Speed Internet and Early Childhood Development. Working Paper.

Colombo, K., E. Failache, and M. Querejeta (2025). High-Speed Internet and Socioemotional Wellbeing in Adolescence and Youth. *Journal of Population Economics 38*(1), 1–29.

Dalisay, F. (2012). Media Use and Acculturation of New Immigrants in the United States. *Communication Research Reports 29*(2), 148–160.

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

Gentzkow, M. and J. M. Shapiro (2008). Preschool Television Viewing and Adolescent Test Scores: Historical Evidence from the Coleman Study. *The Quarterly Journal of Economics 123*(1), 279–323.

Guryan, J., E. Hurst, and M. Kearney (2008). Parental Education and Parental Time with Children. *Journal of Economic Perspectives 22*(3), 23–46.

Hastie, T., R. Tibshirani, and J. Friedman (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer Science & Business Media*.

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

Hernæs, Ø., S. Markussen, and K. Røed (2019). Television, Cognitive Ability, and High School Completion. *Journal of Human Resources 54*(2), 371–400.

Hill, D. et al. (2016). Council on Communications and Media: "Media and Young Minds". *Pediatrics 138*(5), e20162591.

Huang, F. and M.-J. Lee (2010). Dynamic Treatment Effect Analysis of TV Effects on Child Cognitive Development. *Journal of Applied Econometrics 25*(3), 392–419.

Jaeger, D. A., T. J. Joyce, and R. Kaestner (2020). A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing? *Journal of Business & Economic Statistics 38*(2), 317–326.

Kearney, M. S. and P. B. Levine (2015). Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing. *American Economic Review 105*(12), 3597–3632.

Kearney, M. S. and P. B. Levine (2019). Early Childhood Education by Television: Lessons from Sesame Street. *American Economic Journal: Applied Economics 11*(1), 318–350.

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

Li, C. and W.-H. S. Tsai (2015). Social Media Usage and Acculturation: A Test with Hispanics in the US. *Computers in Human Behavior 45*, 204–212.

Matějka, F. and A. McKay (2015). Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. *American Economic Review 105*(1), 272–298.

Meroni, E. C., D. Piazzalunga, and C. Pronzato (2022). Allocation of Time and Child Socio-Emotional Skills. *Review of Economics of the Household 20*(4), 1155–1192.

Mollborn, S., A. Limburg, J. Pace, and P. Fomby (2022). Family Socioeconomic Status and Children's Screen Time. *Journal of Marriage and Family 84*(4), 1129–1151.

Nieto, A. and M. Suhrcke (2021). The Effect of TV Viewing on Children's Obesity Risk and Mental Well-Being: Evidence from the UK Digital Switchover. *Journal of Health Economics 80*, 102543.

Reinecke, L. and W. Hofmann (2016). Slacking Off or Winding Down? An Experience Sampling Study on the Drivers and Consequences of Media Use for Recovery. *Journal of Communication 66*(4), 529–559.

Sims, C. A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics 50*(3), 665–690.

Survey Research Center (1996-2019). Panel Study of Income Dynamics. Institute for Social Research, University of Michigan.

Trudeau, J. (2016). The Role of New Media on Teen Sexual Behaviors and Fertility Outcomes: The Case of 16 and Pregnant. *Southern Economic Journal 82*(3), 975–1003.

Vandewater, E. A., D. S. Bickham, and J. H. Lee (2007). Time Well Spent? Relating Television Use to Children's Free-Time Activities. *Pediatrics 119*(2), e314–e319.

Vigdor, J. L., H. F. Ladd, and E. Martinez (2014). Scaling the Digital Divide: Home Computer Technology and Student Achievement. *Economic Inquiry 52*(3), 1103–1119.

Zill, N. (1990). Behavior Problems Index Based on Parent Report. Technical report, Child Trends, Inc, Washington, D.C.

A Factor Loadings for Cognitive and Noncognitive Scores

Table A.1: Noncognitive Questions and Categorizations

Wording of the questions as they appear in the raw data	Zill (1990) Category
Has trouble getting along with other children	Peer Problems
Is not liked by other children	Peer Problems
Feels others out to get them	Peer Problems
Is withdrawn, does not get involved with others	Peer Problems
Cheats or tells lies	Anti-social
Breaks things on purpose or deliberately destroys (their) own or another's things	Anti-social
Bullies or is cruel or mean to others	Anti-social
Does not seem to feel sorry after (they) misbehaves	Anti-social
Has trouble getting along with teachers	Anti-social
Is disobedient at school	Anti-social
Hangs around with kids who get in trouble	Anti-social
Is too fearful or anxious	Anxious or Depressed
Has sudden changes in mood or feeling	Anxious or Depressed
Feels or complains that no one loves them	Anxious or Depressed
Is unhappy, sad or depressed	Anxious or Depressed
Worries too much	Anxious or Depressed
Feels worthless or inferior	Anxious or Depressed
Demands a lot of attention	Dependent
Clings to adults	Dependent
Cries too much	Dependent
Is too dependent on others	Dependent
Argues too much	Headstrong
Is disobedient	Headstrong
Is rather high strung, tense and nervous	Headstrong
Is stubborn, sullen, or irritable	Headstrong
Has a very strong temper and loses it easily	Headstrong
Has difficulty concentrating, cannot pay attention for long	Hyperactive
Is easily confused or seems to be in a fog	Hyperactive
Is impulsive or acts without thinking	Hyperactive
Has a lot of difficulty getting (their) mind off certain thoughts	Hyperactive
Is restless or overly active or cannot sit still	Hyperactive
Is secretive, keeps things to themselves	-

Notes: The first column of the table provides the wording of the 32 questions that compose the Behavior Problems Index. The second column shows the domain from Zill (1990) that each question corresponds to. "Is secretive, keeps to themselves" is not categorized in Zill (1990), so it is not categorized in our analysis. Individual items are recoded such that an increase means better development. Thus, we interpret the Zill (1990) categories as "no peer problems," "not anti-social," etc.

Table A.2: Cognitive and Noncognitive Factor Loadings

	1997	2002	2007	2014	2019
Cognitive Skills					
Letter Word	0.320	0.346	0.345	0.418	0.249
Applied Problems	0.185	0.192	0.224	0.153	0.174
Passage Comprehension	0.517	0.482	0.471	0.450	0.596
Noncognitive Skills					
Does not have sudden changes in mood or feeling	0.062	0.061	0.061	0.059	0.058
Does not feel or complain that no one loves them	0.049	0.060	0.061	0.072	0.047
Is not rather high strung, tense and nervous	0.050	0.058	0.054	0.064	0.054
Does not cheat or tell lies	0.051	0.060	0.061	0.056	0.049
Is not too fearful or anxious	0.046	0.050	0.052	0.046	0.046
Does not argue too much	0.072	0.055	0.056	0.057	0.063
Does not have difficulty concentrating, can pay attention for long	0.070	0.069	0.062	0.049	0.058
Is not easily confused, does not seem to be in a fog	0.055	0.056	0.049	0.049	0.046
Does not bully or is not cruel or mean to others	0.072	0.061	0.050	0.045	0.046
Is not disobedient	0.050	0.062	0.055	0.054	0.064
Seems to feel sorry after they misbehaves	0.040	0.043	0.036	0.041	0.039
Does not have trouble getting along with other children	0.069	0.073	0.061	0.081	0.076
Is not impulsive, or does not act without thinking	0.076	0.073	0.067	0.069	0.067
Does not feel worthless or inferior	0.053	0.061	0.069	0.045	0.057
Is liked by other children	0.039	0.037	0.043	0.043	0.045
Does not have a lot of difficulty getting their mind off certain thoughts	0.064	0.057	0.066	0.063	0.065
Is not restless or overly active, can sit still	0.075	0.054	0.047	0.051	0.055
Is not stubborn, sullen, or irritable	0.079	0.071	0.068	0.069	0.083
Does not have a very strong temper and does not lose it easily	0.076	0.083	0.077	0.073	0.073
Is not unhappy, sad or depressed	0.069	0.067	0.067	0.072	0.058
Is not withdrawn, gets involved with others	0.037	0.041	0.035	0.048	0.034
Does not break things on purpose or does not deliberately destroy their own or another's things	0.047	0.048	0.045	0.049	0.044
Does not cling to adults	0.031	0.032	0.027	0.025	0.027
Does not cry too much	0.042	0.031	0.032	0.030	0.028
Does not demand a lot of attention	0.072	0.060	0.055	0.043	0.041
Is not too dependent on others	0.045	0.048	0.047	0.047	0.028
Does not feel others out to get them	0.065	0.052	0.064	0.047	0.053
Does not hang around with kids who get in trouble	0.042	0.038	0.038	0.038	0.036
Is not secretive, keeps things to themselves	0.049	0.036	0.038	0.039	0.032
Does not worry too much	0.031	0.034	0.040	0.049	0.043
Is not disobedient at school	0.047	0.053	0.049	0.050	0.057
Does not have trouble getting along with teachers	0.042	0.047	0.037	0.041	0.052

Note: Questions related to noncognitive skills represent the rescaled version of the true question wording, which is shown in Table A.1.

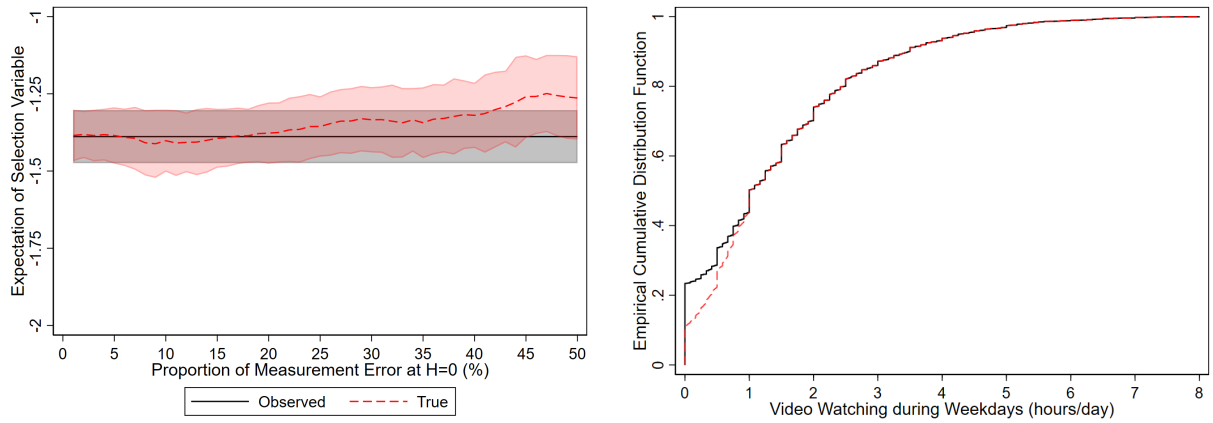
B Understanding the Consequences of Measurement Error

In this section, we discuss why the measurement error case discussed in Section 4.2.5 leads to violations of Assumptions 1 and 2. We also show that these violations tend to offset each other, which explains why our estimates of β are relatively stable as the measurement error problem becomes more prevalent (Figure 16).

Measurement error affects the tail symmetry assumption because it changes the observed amount of bunching at zero, and, consequently the quantile used to estimate $\mathbb{E}[H_i^*|H_i = 0]$. Because this simultaneously changes both the threshold for the upper-tail sample average and the observations included in that average, the net effect on the estimate of $\mathbb{E}[H_i^*|H_i = 0]$ is not obvious *a priori*. The left panel of Figure B.1 shows the estimated value of $\mathbb{E}[H_{i,True}^*|H_{i,True} = 0]$ for different hypothetical amounts of measurement error as well as the estimated expectation using the observed data. In our setting, the estimated true expectation becomes gradually less negative as the amount of hypothetical measurement error increases.

The right panel of Figure B.1 illustrates these changes for the case where 50% of the observations at $H_i = 0$ are assumed to be mismeasured. The black solid plot shows the empirical CDF of H_i as observed in the data, and the red dashed plot represents the CDF of the true treatment variable $H_{i,\text{True}}$, according to this model of measurement error. Relative to the true distribution, the observed distribution exhibits excess bunching at zero. Specifically, we wrongly assume a bunching rate a little over 20%, when in fact it is a little over 10%. This causes us to use a different quantile when constructing the unobserved lower tail, and, consequently, a different set of observations from the observed upper tail when estimating $\mathbb{E}[H_i^*|H_i = 0]$. In our setting, this leads to an estimate of $\mathbb{E}[H_i^*|H_i = 0]$ that is more negative than the true expectation under $H_{i,\text{True}}$. As shown in Figure 9, if our expectation estimate is farther from zero than the true value of the expectation, this by itself will lead our estimate of β to be closer to zero.

Figure B.1: Understanding How Measurement Error Violates the Distributional Assumption



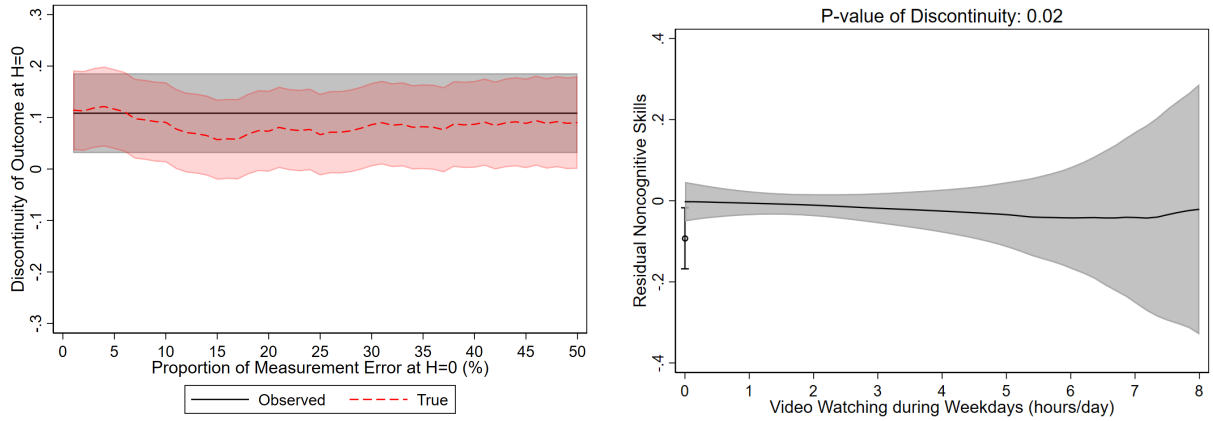
Notes: The left panel shows $\hat{\mathbb{E}}[H_i^*|H_i = 0]$ from our observed data (solid black horizontal line) with the corresponding 95% confidence interval (shaded gray) using the tail symmetry assumption. It also shows different values of $\hat{\mathbb{E}}[H_{i,True}^*|H_{i,True} = 0]$ under different measurement error scenarios (dashed red line), along with 95% confidence intervals (shaded red) using the tail symmetry assumption. The value on the horizontal axis refers to the percentage of observations at $H_i = 0$ assumed to be mismeasured in each scenario, with $H_{i,True}$ randomly drawn from the interval $(0, 1)$. The right panel shows the empirical CDF of H_i (solid black line) and the empirical CDF of $H_{i,True}$ (dashed red line) under the assumption that 50% of observations at $H_i = 0$ are mismeasured, with $H_{i,True}$ randomly drawn from the interval $(0, 1)$.

Measurement error also violates the linearity assumption because it changes the relationship between the residualized outcome and the treatment variable near $H_i = 0$. The redistribution of observations from $H_i = 0$ to $H_{i,\text{True}}$, where $0 < H_{i,\text{True}} < 1$, affects the slope of $\mathbb{E}[S_i|H_{i,\text{True}} = h, X_i]$ as h approaches 0, which in turn changes the discontinuity in the outcome at the bunching point. The left panel of Figure B.2 shows the estimated discontinuity in noncognitive skill ($\Delta_{S_{\text{True}}}$) at $H_{i,True} = 0$ for different assumed amounts of measurement error as well as the estimated discontinuity using the observed data. In our setting, the estimated true discontinuity is generally smaller in magnitude in the presence of measurement error.

The mechanism behind the smaller discontinuity can be seen by comparing the right panel of Figure B.2, which uses $H_{i,True}$ as the treatment variable under the assumption that 50% of the observations at $H_i = 0$ are mismeasured, with the right panel of Figure 6. The slope is a bit flatter as h approaches 0 for the true treatment variable than for the mismeasured treatment variable, because some of the observations at $H_i = 0$ end up randomly redistributed to values just above zero. These reassigned observations have lower residualized outcomes, on average, than those originally located in that region, so they lower the average of the residualized outcome among observations just to the right of zero, as shown in the right panel of Figure B.2. Because the reassignment is random, however, it does not change the average residualized outcome among the observations that remain at

$H_i = 0$. As a result, the discontinuity under the mismeasured treatment variable Δ_S is larger than the discontinuity $\Delta_{S_{\text{True}}}$ under the true treatment variable. This alone will lead us to overstate the magnitude of β.

Figure B.2: Understanding How Measurement Error Violates the Linearity Assumption



Notes: The left panel shows $\hat{\Delta}_S$ at $H_i = 0$ (i.e., the discontinuity in noncognitive skill at the bunching point) from our observed data (solid black horizontal line) with the corresponding 95% confidence interval (shaded gray). It also shows different values of $\hat{\Delta}_{S_{\text{True}}}$ at $H_{i,True} = 0$ under different measurement error scenarios (dashed red line), along with 95% confidence intervals (shaded red). The value on the horizontal axis refers to the percentage of observations at $H_i = 0$ assumed to be mismeasured in each scenario, with $H_{i,True}$ randomly drawn from the interval $(0, 1)$. The right panel is analogous to the right panel of Figure 6 but uses $H_{i,True}$ as the treatment variable and assumes 50% measurement error at $H_i = 0$.

The offsetting of the biases generated by the violations of the two assumptions can be seen by invoking Remark 3.2. While $\delta = \frac{\Delta_S}{\Delta_{H^*}}$ according to the mismeasured H_i, $\delta_{\text{True}} = \frac{\Delta_{S_{\text{True}}}}{\Delta_{H^*_{\text{True}}}}$ according to the true treatment variable, $H_{i,\text{True}}$. Because the discontinuity estimates of Δ_S and $\Delta_{H^*} = -\mathbb{E}[H_i^*|H_i = 0]$ both tend to be larger in magnitude than they would be under the true treatment variable, δ and δ_{True} tend to be similar. This explains why the estimates of β are fairly stable when measurement error is introduced.

C Additional Summary Statistics

Table C.1: Time Watching Video by Characteristics of Video Watching Behavior and by Race/Ethnicity

Hours/day	All Sample	Non-Hispanic White	Non-Hispanic Black	Hispanic	Other Race/Ethnicity
Video Watching	1.47 (1.52)	1.36 (1.47)	1.74 (1.55)	1.77 (1.57)	1.53 (1.84)
Period of the Day					
6 a.m. to 3 p.m.	0.21 (0.65)	0.21 (0.66)	0.22 (0.65)	0.24 (0.65)	0.21 (0.58)
3 p.m. to 8 p.m.	0.85 (1.06)	0.76 (1.01)	1.04 (1.13)	1.06 (1.12)	0.93 (1.11)
8 p.m. to 6 a.m.	0.41 (0.70)	0.39 (0.66)	0.47 (0.77)	0.47 (0.81)	0.40 (0.80)
Company					
Alone	0.56 (1.08)	0.51 (1.03)	0.76 (1.20)	0.50 (1.04)	0.62 (1.27)
Not alone	0.92 (1.21)	0.85 (1.15)	0.97 (1.25)	1.27 (1.37)	0.91 (1.37)
Missing	0.00 (0.06)	0.00 (0.04)	0.00 (0.02)	0.01 (0.14)	0.00 (0.00)
Device					
TV, VHS, DVD player	1.35 (1.49)	1.25 (1.42)	1.58 (1.52)	1.60 (1.55)	1.45 (1.82)
Computer, tablet, phone, and others	0.08 (0.46)	0.08 (0.45)	0.10 (0.47)	0.11 (0.51)	0.04 (0.27)
Missing	0.04 (0.25)	0.03 (0.21)	0.06 (0.32)	0.06 (0.34)	0.05 (0.26)
Show Genre					
Informative/educational	0.09 (0.31)	0.09 (0.31)	0.07 (0.28)	0.11 (0.36)	0.06 (0.27)
Fictional entertainment	0.76 (1.12)	0.73 (1.08)	0.80 (1.17)	0.90 (1.18)	0.76 (1.23)
Reality/variety entertainment	0.20 (0.57)	0.18 (0.51)	0.31 (0.75)	0.21 (0.61)	0.22 (0.58)
Missing	0.42 (0.90)	0.36 (0.85)	0.56 (0.96)	0.55 (1.00)	0.49 (0.99)
Targeted Audience					
Age appropriate	0.58 (0.93)	0.56 (0.89)	0.66 (1.06)	0.62 (0.98)	0.50 (0.82)
Not age appropriate	0.56 (1.05)	0.53 (1.01)	0.62 (1.04)	0.66 (1.13)	0.66 (1.34)
Missing	0.33 (0.81)	0.27 (0.76)	0.45 (0.87)	0.48 (0.96)	0.37 (0.79)
Observations	5,706	2,774	2,252	426	227

Notes: The sample includes 5,706 child-wave observations from the 1997, 2002, 2007, 2014, and 2019 CDS-PSID waves. Sample weights are applied and standard deviations are shown in parentheses.

Table C.2: Time Watching Video by Characteristics of Video Watching Behavior and by Family Income

Hours/day	All Sample	Lower Tercile	Middle Tercile	High Tercile	Missing Income
Video Watching	1.47 (1.52)	1.68 (1.60)	1.55 (1.55)	1.28 (1.44)	1.65 (1.47)
Period of the Day					
6 a.m. to 3 p.m.	0.21 (0.65)	0.26 (0.74)	0.23 (0.69)	0.18 (0.58)	0.15 (0.38)
3 p.m. to 8 p.m.	0.85 (1.06)	0.98 (1.09)	0.92 (1.08)	0.70 (1.01)	0.99 (0.93)
8 p.m. to 6 a.m.	0.41 (0.70)	0.44 (0.74)	0.40 (0.69)	0.40 (0.68)	0.51 (0.83)
Company					
Alone	0.56 (1.08)	0.57 (1.06)	0.57 (1.17)	0.54 (1.03)	0.43 (0.74)
Not alone	0.92 (1.21)	1.11 (1.36)	0.98 (1.22)	0.74 (1.07)	1.22 (1.40)
Missing	0.00 (0.06)	0.00 (0.10)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
Device					
TV, VHS, DVD player	1.35 (1.49)	1.51 (1.55)	1.43 (1.53)	1.18 (1.40)	1.60 (1.48)
Computer, tablet, phone, and others	0.08 (0.46)	0.11 (0.54)	0.09 (0.48)	0.07 (0.38)	0.00 (0.00)
Missing	0.04 (0.25)	0.06 (0.32)	0.03 (0.22)	0.03 (0.22)	0.05 (0.25)
Show Genre					
Informative/educational	0.09 (0.31)	0.10 (0.33)	0.09 (0.32)	0.08 (0.29)	0.04 (0.16)
Fictional entertainment	0.76 (1.12)	0.87 (1.24)	0.79 (1.08)	0.66 (1.04)	1.03 (1.38)
Reality/variety entertainment	0.20 (0.57)	0.23 (0.63)	0.20 (0.55)	0.19 (0.56)	0.11 (0.34)
Missing	0.42 (0.90)	0.48 (0.90)	0.46 (1.01)	0.34 (0.80)	0.47 (0.93)
Targeted Audience					
Age appropriate	0.58 (0.93)	0.62 (1.00)	0.61 (0.91)	0.53 (0.89)	0.62 (0.96)
Not age appropriate	0.56 (1.05)	0.68 (1.22)	0.56 (1.01)	0.50 (0.97)	0.59 (0.94)
Missing	0.33 (0.81)	0.39 (0.81)	0.38 (0.94)	0.25 (0.67)	0.45 (0.93)
Observations	5,706	1,848	1,853	1,864	141

Notes: The sample includes 5,706 child-wave observations from the 1997, 2002, 2007, 2014, and 2019 CDS-PSID waves. Sample weights are applied and standard deviations are shown in parentheses.