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	No 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Is withdrawn, does not get involved with others	Peer Problems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 Details of Estimation Method with Controls

When adding controls, we rewrite equation (4) as

$$\mathbb{E}[S_i|H_i] = \alpha(X_i) + \beta H_i + \delta(X_i)(H_i + \mathbb{E}[H_i^*|H_i = 0, X_i]\mathbf{1}(H_i = 0)), \tag{7}$$

where now there are three quantities that depend on 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 with the use of 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 trivial to estimate each of these quantities separately for each value of X_i with 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. $\mathbb{E}[H_i^*|H_i = 0, X_i] = \sum_{k=1}^{K} \mathbb{E}[H_i^*|H_i = 0, X_i = x_k]\mathbf{1}(X_i = x_k)$.

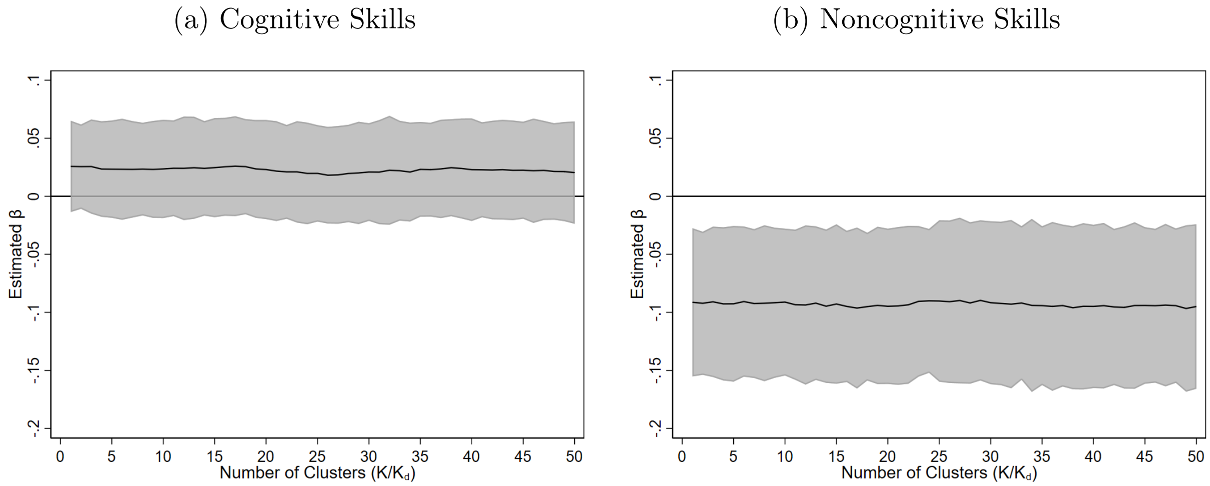
As shown in Section 2, X_i takes too many values to feasibly implement this discrete strategy. Thus, we follow Caetano et al. (2024b) and first discretize X_i by assigning each observation to one of K clusters, based on their similarity, and then treat each cluster k as if it had a different value x_k, as in the discrete case discussed above. This means, for example, that we allow for α_k to vary only across clusters indexed by k, not across different values of X_i inside the same cluster. While observations within a given cluster may have different values of X_i, this variation is limited, more so the larger is the number of clusters. 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 available methods for grouping observations according to the similarity in their value of X_i. We report results using 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). For reference, our results reported in the text assume $K = 1$ (so controls are only added linearly in $\alpha(X_i)$), but the results below show that the choice of K in the specification of the three quantities ($\alpha(X_i)$, $\delta(X_i)$ and $\mathbb{E}[H_i^*|H_i = 0, X_i]$) is irrelevant in our context.

Further Evidence of Robustness

Figure B.1 shows that our estimates of β are stable irrespective of the value of K. To understand why this stability should be understood as evidence in favor of Assumptions 1 and 2, we argue below that, as we allow for the number of clusters to grow from $K = 1$ to $K = 50$, the linearity and distributional assumptions (Assumptions 1 and 2) become strictly weaker.

Figure B.1: Number of Clusters in Specification of $\alpha(X_i)$, $\delta(X_i)$, and $E[H_i^*|H_i = 0, X_i]$

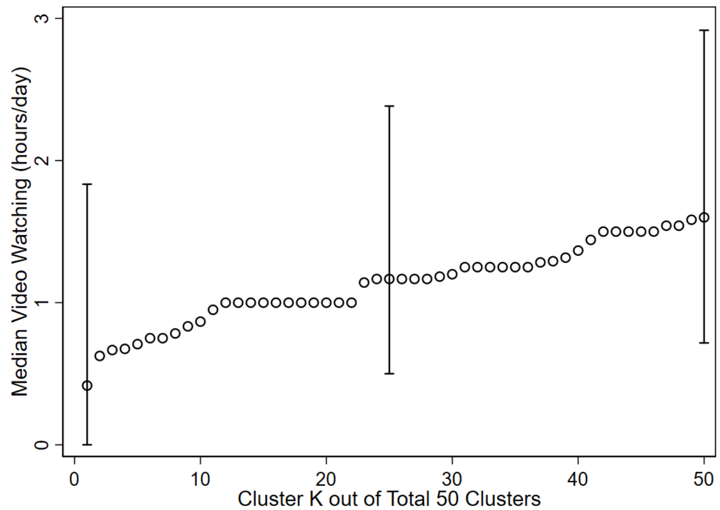


Notes: Estimates using the tail symmetry assumption and corresponding 95% confidence intervals shown for different numbers of clusters used in the analysis (500 bootstrap samples).

Figure B.2 shows the median of $H_i^* = H_i$ for each of the 50 clusters, sorted based on the median, as well as the interquartile distribution of H_i for three selected clusters.[33] As the cluster changes, not only the median changes, but other moments of the distribution change. For instance, the interquartile range is smaller for clusters 1 and 25 than for cluster 50. Thus, as we increase the number of clusters from $K = 1$ to $K = 50$, we are effectively absorbing potential nonlinearities and asymmetries in the selection component of the outcome, which violate Assumptions 1 and 2. And yet, the results from Figure B.1 tell us that the main estimates barely change when we do so. Thus, these results should be viewed as further evidence that our key results are not an artifact of these assumptions. Note that 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 itself. Thus, these results rule out the knife-edge case discussed in Footnote 24.

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

Figure B.2: 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.

C Understanding the Consequences of Measurement Error

In this section, we discuss why the measurement error case discussed in the paper leads to violations of Assumptions 1 and 2, and why 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 10).

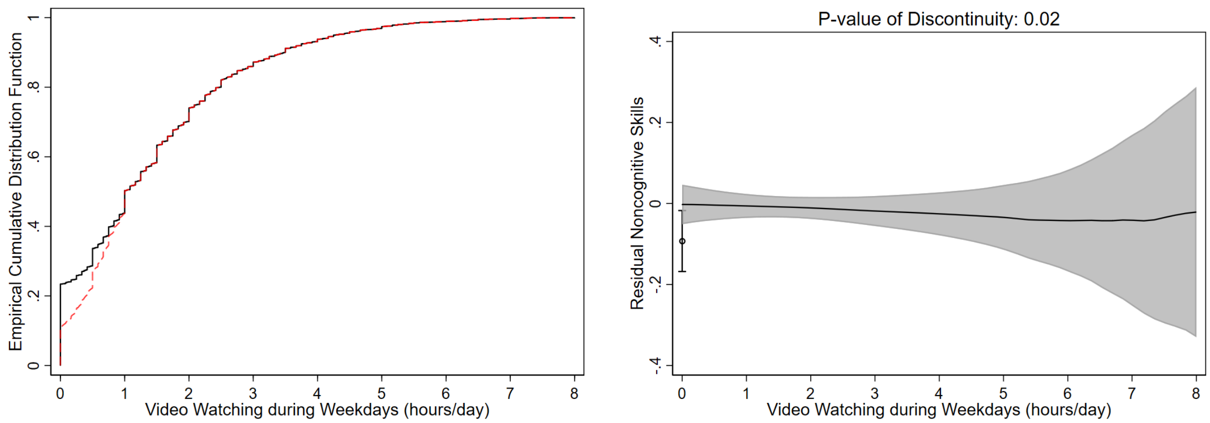
Figure C.1 illustrates the case where 50% of the observations at $H_i = 0$ are assumed to be mismeasured. In the left panel, the black solid plot shows the empirical CDF of H_i as observed in the data, and the red dashed plot represents the CDF of the true treatment variable $H_{i,\text{True}}$, according to this model of measurement error.

It is easy to see how the tail symmetric assumption under the mismeasured H_i will give us an expectation ($\mathbb{E}[H_i^*|H_i = 0]$) that is too negative relative to the true expectation under $H_{i,\text{True}}$. This happens because the only source of bias in the expectation will stem from the fact that we will wrongly assume that the bunching rate is a little over 20%, when in fact it is a little over 10%.[34] As shown in our results in Figure 7, if our expectation estimate turns

[34] Note that according to the tail symmetry assumption, we use only the upper tail of the black solid distribution from the left panel to estimate the lower tail of the distribution. This upper tail is identical to the upper tail of the true, red dashed distribution in the left panel, since this measurement error only applies to small values of H_i.

out to be farther from zero than the true value of the expectation, then our mismeasured H_i will lead our estimate of β to be closer to zero.

Figure C.1: Understanding Measurement Error (50% of Mismeasured Observations at $H_i = 0$)



Notes: The left panel shows the empirical CDF of H_i (solid black line) and the empirical CDF of $H_{i,True}$ (red dashed line) under the assumption that 50% of observations at $H_i = 0$ are mismeasured, with $H_{i,True}$ randomly drawn from the interval $(0, 1)$ in that case. The right panel is analogous to the right panel of Figure 6 but uses $H_{i,True}$ as the treatment variable.

The linearity assumption is violated in the presence of measurement error because the redistribution of observations from $H_i = 0$ to $H_{i,\text{True}}$, where $0 < H_{i,\text{True}} < 1$, affects the slope of $\mathbb{E}[S_i|H_{i,\text{True}} = h]$ as h approaches 0. This can be seen by comparing the right panel of Figure C.1 for noncognitive skills, which uses $H_{i,True}$ as the treatment variable, 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, as some of the observations at $H_i = 0$ (which have a lower residualized outcome, on average, than the observations at $H_i > 0$) end up randomly redistributed to the positive side near zero. This redistribution should not affect the average of the residualized outcome at $H_i = 0$, but it should reduce the average of the residualized outcome at $H_i > 0$ near zero, as shown in the right panel of Figure C.1. Because of this, the discontinuity $\Delta_{S_{\text{True}}}$ under the true treatment variable is smaller than the discontinuity under the mismeasured treatment variable, Δ_S. Thus, our mismeasured H_i will lead us to estimate a larger discontinuity in the outcome, which will lead us to estimate a larger β in magnitude.

The offsetting of the biases generated by the violations of the two assumptions can be seen clearly 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, $H_{i,\text{True}}$. The fact that the discontinuity estimates of Δ_S and $\Delta_{H^*} = -\mathbb{E}[H_i^*|H_i = 0]$ tend to be larger in magnitude than they would be under the true treatment variable means that δ and δ_{True} tend to be similar to each other, which explains why the estimates of β are fairly stable when measurement error is introduced.

D Additional Summary Statistics

Table D.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 D.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.