# Online Appendix

"The Economic Value of *Breaking Bad*: Misbehavior, Schooling and the Labor Market"

Nicholas W. Papageorge, Victor Ronda and Yu Zheng

## Appendix A Additional Descriptive Statistics and Factor Analysis

Descriptive statistics for the full sample of observed 11-year-olds are found in Tables S1 and S2. Also, descriptive statistics for the BSAG by SES used in Section 5.2 are found in Table S3 and the corresponding factor loadings are found in Table S4. The remainder of this Section of the appendix explains how factor analysis is used in this paper (see Figure S1 and Tables S5-S6).

In the National Child Development Survey (NCDS), childhood misbehavior and maladjustment are measured as follows. Teachers read a number of phrases and then report whether each phrase applies to the child in question. These measures are then aggregated into 10 variables capturing childhood maladjustment, known as the BSAG maladjustment variables.[1] We use factor analysis, a statistical technique used for data reduction to assess whether classroom misbehavior can be represented using fewer than the ten dimensions available from the NCDS. These techniques reduce the number of dimensions by uncovering linear combinations of the original variables that contain most of the information in the data and that also have meaningful interpretations. Specifically, the analysis determines the number of dimensions needed to adequately describe observed variation in classroom behavior and which of the BSAG variables are related to which dimensions of classroom misbehavior. From here on, we refer to the original BSAG variables as *measurements* and each dimension of classroom misbehavior as *factors*.

We begin by writing the measurements of classroom behavior as a linear function of unobserved factors. If there are $k$ underlying factors, we can write our original 10 BSAG variables as:

$$BSAG_{ji} = l_{j1}f_{1i} + ... + l_{jk}f_{ki} + \epsilon_{ji}, \text{ for } j = 1, \cdots, m, \text{ and } i = 1, \cdots, n \tag{1}$$

where $BSAG_{ji}$ is the value of the $j$th BSAG variable for individual $i$, $f_{qi}$ is the unobserved value of the $q$th factor for individual $i$, $l_{jq}$ represents the coefficient relating factors to measurements (usually referred to as the *factor loading*) and $\epsilon_{ji}$ is a residual, capturing

[1] There are actually 12 BSAG variables available. We exclude two of the original variables from this study in order to maintain consistency with recent research using the same data set (see Shepherd (2013)). The two omitted variables are called Miscellaneous Symptoms and Miscellaneous Nervous Symptoms. Our main results do not change if these variables are included in the analysis. Results from this robustness test are available upon request from the authors.

measurement error. For each individual, we can rewrite equation (1) in matrix form as follows:

$$BSAG = LF + \epsilon \tag{2}$$

where $BSAG$ is the $(m \times n)$ measurement matrix, $L$ is the $(m \times k)$ matrix of factor loadings and $F$ is the $(k \times n)$ matrix of unobserved factors, where $m$ is the number of measurements, $n$ represents the number of observations for each measurement and $k$ is the number of factors.

The first goal of factor analysis is to find a small number of unobserved factors $(k < m)$ that sufficiently explain the variation in the measurements. We are interested in the variation within and across the measurements instead of the measurements *per se*. In other words, we are interested in explaining the measurement covariance matrix $(V)$, which has size $m \times m$ and is given by:

$$V = BSAG\ BSAG^T = LL^T + \Omega \tag{3}$$

The above expression is valid under the assumptions that (i) $F$ and $\epsilon$ are independent and (ii) $\mathbb{E}[F] = 0$, with $FF^T = I$, implying that the factors are uncorrelated and $\Omega$ is the diagonal error variance matrix. Then, the factor analysis problem amounts to understanding the symmetric positive semidefinite matrix $LL^T$. We can decompose this matrix using an eigen-decomposition, so that:

$$\begin{aligned} V - \Omega = LL^T &= CDC^T \\ &\approx L_k L_k^T = C_k D_k C_k^T \end{aligned} \tag{4}$$

where $C$ is the $(m \times m)$ matrix whose columns are the eigenvectors of $L^T L$ and $D$ is the $(m \times m)$ diagonal matrix whose entries are the eigenvalues. The key 'trick' of factor analysis is that we can reduce the dimensionality of $F$ (or $L$) by picking the $k$ eigenvalues and eigenvectors that explain a lot of the measurement variance $U \equiv (V - \Omega)$. To accomplish this, we use $C_k$, the $(m \times k)$ matrix whose columns are the eigenvectors associated with the $k$ largest eigenvalues, and define as $D_k$ the diagonal matrix of the eigenvalues. The resulting $(m \times m)$ matrix $C_k D_k C_k^T$ is not equal to $U$, but will converge to $U$ as $k$ gets closer to $m$ and will be closer to $U$ than any other matrix with rank $k$.

Now that we understand the idea behind the factor analysis, we need to decide on $k$, the number of factors to be used. There are three widely used criteria in psychometrics. The first method is commonly known as Kaiser's criterion or Kaiser's stopping rule. It stipulates that only the number of latent independent factors with eigenvalues greater than

1 should be considered in the analysis. Recall, for a given factor, the eigenvalue measures the variance in all measures that is accounted for by that factor. A low eigenvalue means that the factor contributes little to explaining variance and may be treated as redundant and therefore ignored. The second method is known as the scree plot method. In this method, the researcher plots the relationship between the relative magnitude of the eigenvalues and the number of factors. The researcher then examines the scree plot and decides where the line stops descending precipitously and levels out. The number of points along the precipitously dropping part of the line, excluding the transition point, gives the number of factors that should be used in the analysis. The third method is known as parallel analysis. In this method, we create a dataset with random numbers and the same number of observations and variables as in the original data set. Then we compute the eigenvalues for each factor as we did for the original data using factor analysis. The researcher should keep only the number of factors where the eigenvalues from the random data are smaller than the eigenvalues from the factor analysis using the original data as the remaining factors are effectively capturing random noise.

We use all three methods on the BSAG maladjustment variables in our data and all suggest we should use exactly two factors in our analysis. Our findings match those of Ghodsian (1977) and Shepherd (2013), who also studied childhood misbehavior using NCDS data. To perform the test using Kaiser's criterion, we compute the eigenvalues of the correlation matrix together with the eigenvectors corresponding to each factor. We plot the eigenvalues of the factors in descending order in Figure S1. Factors 1 and 2 have eigenvalues of 3.69 and 1.78 respectively, whereas factors 3 to 10 have eigenvalues between 0.9 and 0.3. The Kaiser's stopping rule suggests keeping only the factors with eigenvalues equal or higher than 1, or the first two factors in our analysis. In the same figure, we also notice that the line connecting the eigenvalues stops descending and levels out after the second factor. The scree plot method then suggests that we should keep and use only the first two factors. Lastly, in the same figure, we plot the eigenvalues from the random data created by the parallel analysis (dashed line). The two lines intersect before the third factor, suggesting that factors 3 to 10 capture a level of variation in the data that is generated by random noise. Hence, only the first two factors should be used.

Now that we have shown that two independent random variables or "factors" explain the ten BSAG maladjustment variables, it remains to be determined which set of BSAG variables are related to which factor. Deciding on which measurement is related to which factor is less straightforward than deciding on the number of factors. Since the two eigenvectors can be rotated in an infinite number of ways, we rotate the original eigenvectors in order to maximize the variance accounted for by the first two factors using the *quartimin* method. This produces the rotated factor loadings shown in Table S5. Clearly, there is an association between the first 6 BSAG variables and the first factor. Coefficients range from 0.53 for anxiety of acceptance by adults to 0.80 for inconsequential behavior, which are fairly high. Moreover, intuitively, all these variables seem to be measures of outwardly expressed behaviors. Similarly, there is a clear association between the 7-9th measurements and the second factor. Again, these variables all represent inwardly expressed behavior. The last measurement is less clear. It seems to be statistically related to both factors and it is less clear intuitively if this behavior is inwardly or outwardly expressed. As a result, we permit the 10th behavior to be related to both latent factors.

It turns out this particular mapping between measurements and factors is the same as the one proposed by Ghodsian (1977) and used in other research (see, e.g., Shepherd (2013)). The first six measurements, taken together with the last measure, capture externalizing behavior and the last four capture internalizing behavior. This leads to the mapping from the BSAG measures to the two factors presented in Table 1. Finally, once we have decided on the mapping, it remains to estimate the two unobserved factors for each individual conditional on each individual's observed measurements, to be used in our reduced-form analysis. There are many methods available to compute the unobserved factor matrix $F$. One widely used method to estimate the unobserved factor is the "regression" or "Thompson" method, which constructs the weighting matrix $b$ that minimizes the mean squared error in Equation (2). The resulting formula for $F$ is given by:

$$F = L^T V^{-1} BSAG \tag{5}$$

where the weighting matrix, also known as the factor score matrix, is given by $b = L^T V^{-1}$. Table S6 presents the factor scores, or weights, used to construct each latent factor.

In the econometric model in the main paper we use a similar approach to estimate the unobserved factors. There we take both the number of factors and the mapping between the measurements and factors as given. We also re-estimate the factor loadings using the restrictions imposed by the mapping and estimate the resulting distribution of each factor separately by gender. The key differences is that we estimate the factor loadings together with the other relevant outcomes. In a sense, the econometric model allows us to use the outcome equations as additional measurements for the unobserved factors. Moreover, the measurement error model by-passes the estimation of the unobserved factors directly as we use the estimated distribution to integrate out the unobserved factors in the estimation. Both modifications are beneficial. The first modification allows us to use additional variation to identify the latent factors. The second modification allows us to reduce the measurement error in the construction of the latent factors.

## Appendix B Additional Reduced-Form Evidence

This appendix contains results from additional reduced-form specifications relating externalizing behavior to earnings. Similar to what we did in Section 2.4, in all results presented here, we construct skills by summing up corresponding observable BSAG measurements and test scores.

In Tables S7 and S8, we explore this relationship once we have controlled for selection into education. In Figures S2 and S3 and Tables S9 and S10, we explore possible non-linearities, non-monotonicites and interaction effects in the relationship between the unobserved factors and earnings. The general conclusion is the following: the positive relationship between externalizing and adult earnings holds under all specifications and, moreover, there is little evidence of non-linearities in the relationship. Lastly, In Tables S11 and S12 we explore if the positive relationship between externalizing behavior and earnings still holds as individuals age.

In the measurement error model in the main text, we jointly model how latent factors affect sorting into education along with other decisions and outcomes. Here, we ask whether the positive reduced-form relationship between externalizing and earnings discussed in Section 2.4 holds once we more formally account for selection into schooling. We employ the widely-used Lee (1983) and Dubin and McFadden (1984) methods for selection bias correction. These methods are akin to a two-stage least squares approach when selection is specified as a multinomial logit model.[2] As in the measurement error model, excluded variables include: class size, average class preparation, number of children in the household, mother's education and father's education.

The first stage estimates can be found in Table S7 and are not very different from what has been shown in the rest of the paper. Cognition is the most important variable for the schooling decision and both other factors describing classroom behavior are negatively associated with educational attainment. The second stage estimates (in addition to OLS estimates for comparison) can be found in Table S8. The effect of externalizing is positive among all educational groups albeit less so for individuals in higher education groups. More-

[2]The method proposed in Lee (1983) is a generalization of the two-step selection bias correction introduced by Heckman (1979) that allows for any parameterized error distribution. The method proposed in Dubin and McFadden (1984) is also a generalization of the method proposed by Heckman (1979) with the further advantage in comparison to the method in Lee (1983) that it does not make any assumption on covariances between the error term in the outcome and selection equations. These and other selection methods based on the multinomial logit model have been reviewed by Bourguignon, Fournier, and Gurgand (2007).

over, controlling for selection into schooling does not alter the results in any important way. If anything, the relationship between externalizing behaviors and earnings becomes stronger after we control for selection into schooling.[3]

We do not allow for non-linearities and interactions between the unobserved factors in the measurement error model estimated in the main text. Here, we run additional regressions allowing for non-linearities and interactions. Figures S2 and S3 plot the impact of the externalizing behavior on log-earnings after we have controlled for additional regressors. In order to control for the other variables we regress log-earnings on the other explanatory variables and use the residuals from that regression as the dependent variable in the non-parametric regression. The impact of externalizing on earnings appears linear, with a decline in the effect for the few individuals with externalizing behaviors above 4 standard deviations from the mean. We also explore non-linearities in Table S9, where we separately regress earnings on those individuals with externalizing behaviors that are one standard deviation above the mean versus individuals below that level. We do not find any difference in the relationship between the two groups. Last, in Table S10 we allow for a quadratic term and interactions between the unobserved factors when regressing those on log-earnings. Again, we find no evidence that interactions or non-linearities are relevant for the factors capturing classroom behavior.

All labor market outcomes in the main paper were constructed when individuals were 33 years old. In this section of the appendix we explore if the relationship between childhood behaviors and earnings change as individuals age. In Tables S11 and S12 we explore this relationship when individuals are 42 and 50 years old respectively. The main patterns remain as individuals age. That is, externalizing behaviors are positively related to earnings even as individuals age. The relationship between externalizing and earnings seems to peak at age 42 and then decrease when individuals reach age 50. It is possible the relationship falls due to changes in the control variables since we use control variables measured at age 33. Nonetheless, these results show that nothing special seems to be happening at age 33. We could have used labor market outcomes from any survey and still obtain similar results.

[3]Note that we use collapse education into three educational groups to increase each group's sample size. The reason is that the selection models we use automatically estimate different earnings equations for each education level. Using six educational groups therefore leads to many more coefficients and larger standard errors.

## Appendix Tables and Figures

**Appendix Table S1:** Summary Statistics - Full Sample

| | Both | Males | Females | |
|---|---|---|---|---|
| No Formal Education | 0.126 | 0.114 | 0.138 | *** |
| | (0.332) | (0.317) | (0.345) | |
| CSE | 0.124 | 0.111 | 0.137 | *** |
| | (0.330) | (0.315) | (0.344) | |
| O Level | 0.341 | 0.306 | 0.375 | *** |
| | (0.474) | (0.461) | (0.484) | |
| A Level | 0.141 | 0.184 | 0.0997 | *** |
| | (0.348) | (0.387) | (0.300) | |
| Higher Education | 0.142 | 0.144 | 0.139 | |
| | (0.349) | (0.351) | (0.346) | |
| Higher Degree | 0.126 | 0.141 | 0.111 | *** |
| | (0.332) | (0.348) | (0.314) | |
| Hourly Wage | 6.749 | 7.645 | 5.666 | *** |
| | (3.063) | (2.969) | (2.815) | |
| Weekly Hours Worked | 36.71 | 43.54 | 28.71 | *** |
| | (12.54) | (7.917) | (12.23) | |
| Weekly Earnings | 259.8 | 329.2 | 175.9 | *** |
| | (152.3) | (135.0) | (127.8) | |
| Experience | 140.4 | 158.9 | 122.6 | *** |
| | (55.69) | (50.80) | (54.35) | |
| In Paid Work | 0.792 | 0.902 | 0.685 | *** |
| | (0.406) | (0.297) | (0.464) | |
| Self Employed | 0.142 | 0.176 | 0.0987 | *** |
| | (0.349) | (0.380) | (0.298) | |
| Has a Partner | 0.794 | 0.783 | 0.804 | ** |
| | (0.405) | (0.412) | (0.397) | |
| Number of Children | 1.512 | 1.347 | 1.666 | *** |
| | (1.147) | (1.148) | (1.125) | |
| London | | 0.302 | 0.305 | |
| | | (0.459) | (0.460) | |
| Observations | 15,356 | 7,899 | 7,457 | 15,356 |

*Notes*: Summary statistics for the full sample of 15,356 individuals observed at age 11. Statistics are reported separately for both genders (Column [1]), for males (Column [2]) and for females (Column [3]). For education categories, employment and partnership, entries are in the form of percentages divided by 100. Experience is measured in months and wages and weekly earnings are in 1992 Great British pounds.

**Appendix Table S2:** Summary Statistics - BSAG Variables - Full Sample

| | Both | Males | Females | |
|---|---|---|---|---|
| Hostility Towards Adults | 0.904 | 1.079 | 0.719 | *** |
| | (1.946) | (2.088) | (1.766) | |
| Hostility Towards Children | 0.288 | 0.336 | 0.237 | *** |
| | (0.805) | (0.892) | (0.699) | |
| Anxiety for Acceptance by Adults | 0.559 | 0.545 | 0.573 | |
| | (1.212) | (1.188) | (1.237) | |
| Anxiety for Acceptance by Children | 0.334 | 0.464 | 0.197 | *** |
| | (0.803) | (0.953) | (0.575) | |
| Restlessness | 0.229 | 0.286 | 0.169 | *** |
| | (0.568) | (0.633) | (0.484) | |
| Inconsequential Behavior | 1.433 | 1.887 | 0.953 | *** |
| | (1.999) | (2.278) | (1.513) | |
| Depression | 1.049 | 1.196 | 0.893 | *** |
| | (1.546) | (1.614) | (1.454) | |
| Withdrawal | 0.347 | 0.410 | 0.279 | *** |
| | (0.826) | (0.910) | (0.720) | |
| Unforthcomingness | 1.606 | 1.630 | 1.582 | |
| | (2.137) | (2.059) | (2.216) | |
| Writing Off of Adults and Adult Standards | 1.019 | 1.263 | 0.760 | *** |
| | (1.703) | (1.911) | (1.406) | |
| Observations | 15,356 | 7,899 | 7,457 | 15,356 |

*Notes*: Summary statistics for maladjustment syndrome scores for the full sample of 14,158 individuals observed at age 11. Measures constructed using teachers' reports of misbehavior or misconduct in school. Statistics are reported separately for both genders (Column [1]), for males (Column [2]) and for females (Column [3]). For each maladjustment syndrome, a child receives a score, which is an integer between 0 and 15, with 15 indicating persistent display of behavior described by the maladjustment syndrome. In the table, entries are averages for each syndrome for the analysis sample.

**Appendix Table S3:** Summary Statistics - BSAG Variables, Subsamples by SES

| | Both | High SES | Low SES | Diff |
|---|---|---|---|---|
| Hostility Towards Adults | 0.765 | 0.700 | 1.108 | *** |
| | (1.756) | (1.647) | (2.210) | |
| Hostility Towards Children | 0.240 | 0.217 | 0.360 | *** |
| | (0.719) | (0.676) | (0.901) | |
| Anxiety for Acceptance by Adults | 0.515 | 0.483 | 0.686 | *** |
| | (1.152) | (1.098) | (1.386) | |
| Anxiety for Acceptance by Children | 0.298 | 0.285 | 0.369 | *** |
| | (0.762) | (0.749) | (0.820) | |
| Restlessness | 0.194 | 0.178 | 0.279 | *** |
| | (0.521) | (0.497) | (0.625) | |
| Inconsequential Behavior | 1.263 | 1.166 | 1.769 | *** |
| | (1.868) | (1.774) | (2.231) | |
| Depression | 0.932 | 0.857 | 1.324 | *** |
| | (1.452) | (1.380) | (1.728) | |
| Withdrawal | 0.307 | 0.292 | 0.387 | *** |
| | (0.771) | (0.743) | (0.902) | |
| Unforthcomingness | 1.477 | 1.414 | 1.810 | *** |
| | (2.035) | (1.992) | (2.221) | |
| Writing Off of Adults and Adult Standards | 0.908 | 0.855 | 1.183 | *** |
| | (1.586) | (1.523) | (1.859) | |
| Observations | 7296 | 6125 | 1171 | 7296 |

*Notes*: Summary statistics for maladjustment syndrome scores for our sample of 7,296 individuals. Measures constructed using teachers' reports of misbehavior or misconduct in school. Statistics are reported separately for all individuals (Column [1]), for individual that did not experience financial difficulties growing up (Column [2]) and for those that did (Column [3]. For each maladjustment syndrome, a child receives a score, which is an integer between 0 and 15, with 15 indicating a persistent display of behavior described by the maladjustment syndrome. In the table, entries are averages for each syndrome for the analysis sample. In Column [4], *, ** and *** mean that differences between males and females are significant at the 10, 5 and 1 percent levels, respectively.

**Appendix Table S4:** Measurement Error Model: Factor Loadings, by SES

| Latent Skill | Measures | [High SES] | [Low SES] |
|---|---|---|---|
| Externalizing Behavior | Inconsequential Behavior | 1.000 | 1.000 |
| | Hostility Towards Adults | 1.448 | 1.807 |
| | Hostility Towards Children | 1.989 | 2.005 |
| | Anxiety for Acceptance by Adults | 0.934 | 0.993 |
| | Anxiety for Acceptance by Children | 1.570 | 1.665 |
| | Restlessness | 1.697 | 1.577 |
| | Writing Off of Adults and Adult Standards | 0.357 | 0.443 |
| Internalizing Behavior | Withdrawal | 1.000 | 1.000 |
| | Depression | 1.057 | 0.917 |
| | Unforthcomingness | 1.757 | 1.877 |
| | Writing Off of Adults and Adult Standards | 0.692 | 0.722 |
| Cognition | Verbal Score on General Ability Test | 1.000 | 1.000 |
| | Reading Comprehension Test Score | 0.584 | 0.612 |
| | Mathematics Test Score | 1.079 | 1.049 |
| | Non Verbal Score on General Ability Test | 0.740 | 0.789 |

*Notes*: This table lists the factor loadings that express the relationship between each observed measure and the underlying factor it identifies by SES groups.

**Appendix Table S5:** FA: Rotated Factor Loadings

| | Factor 1 | Factor 2 | Uniqueness |
|---|---|---|---|
| Hostility Towards Adults | 0.72 | 0.19 | 0.45 |
| Hostility Towards Children | 0.73 | 0.09 | 0.45 |
| Anxiety for Acceptance by Adults | 0.53 | -0.23 | 0.66 |
| Anxiety for Acceptance by Children | 0.75 | -0.12 | 0.42 |
| Restlessness | 0.61 | 0.04 | 0.62 |
| Inconsequential Behavior | 0.80 | 0.15 | 0.32 |
| Depression | 0.39 | 0.67 | 0.39 |
| Withdrawal | 0.12 | 0.79 | 0.35 |
| Unforthcomingness | -0.06 | 0.79 | 0.37 |
| Writing Off of Adults and Adult Standards | 0.53 | 0.54 | 0.42 |

*Notes*: This table contains the factor loadings to the two factors that we retain using the *quartimin* rotation.

**Appendix Table S6:** Reduced Form: Factor Scoring Coefficients

| Cognition | | Externalizing | | Internalizing | |
|---|---|---|---|---|---|
| Measurement | Coefficient | Measurement | Coefficient | Measurement | Coefficient |
| Verbal Score | 0.27648 | Hostility To Children | 0.22625 | Depression | 0.34172 |
| Non Verbal Score | 0.26356 | Hostility To Adults | 0.22810 | Withdrawal | 0.35678 |
| Reading Comp. | 0.25578 | Anxiety Adults | 0.13640 | Unforthcomingness | 0.32127 |
| Mathematics Score | 0.27106 | Anxiety Children | 0.21961 | Writing Off | 0.30815 |
| Copying Designs | 0.13239 | Restlessness | 0.19218 | | |
| | | Inconsequential | 0.25119 | | |
| | | Writing Off | 0.19477 | | |

*Notes*: This table contains the scoring coefficients from the regression scoring method, which are used as weights in the construction of proxies for the three unobserved factors in the reduced-form analysis in Section 2.4.

**Appendix Table S7:** Reduced Form: Educational Attainment - First Stage

| | O Level or A Level | Higher Ed. or Higher Deg. |
|---|---|---|
| Cognition | 1.238*** | 2.064*** |
| Externalizing | -0.070* | -0.174*** |
| Internalizing | -0.155*** | -0.225*** |

*Notes*: This table contains parameter estimates for the first stage IV regressions in Table S8. The coefficients are estimated by multinomial logit, where the base group includes individuals with no formal education or with a CSE degree. Instruments include mother's education, father's education, class size, class preparation and number of children in the household measured at age 11. We also include a gender dummy.

**Appendix Table S8:** Reduced Form: Externalizing and Log Weekly Earnings

| | OLS | IV-DMF | IV-Lee | IV-Lee [M] | IV-Lee [F] |
|---|---|---|---|---|---|
| No Formal Ed. or CSE | 0.039* | 0.041** | 0.042** | 0.017 | 0.048 |
| O Level or A Level | 0.059*** | 0.062*** | 0.057*** | 0.042*** | 0.069** |
| Higher Ed. or Higher Deg. | 0.010 | 0.017 | 0.019 | 0.070* | -0.052 |

*Notes*: This table contains parameter estimates for the externalizing variable from regressions used to link non-cognitive skills to earnings. We regress log earnings of workers on a set of observable variables along with proxies for unobserved skills. To construct proxies for unobserved skills, we apply principal components factor analysis to all the variables used to measure that skill. Column 1 displays the coefficients obtained by OLS. In Columns 2, we perform the Dubin-McFadden (1984) correction method in order to control for selection on schooling. In columns 3-5, we perform the Lee (1983) correction method for selection in schooling. Column 3 displays the results for the whole sample, while column 4 and 5 displays the results for males and females separately.

**Appendix Table S9:** Extremes

| Variable | [1] | [2] | [3] | [4] |
|---|---|---|---|---|
| Cognition | .273*** | .091*** | .253*** | .078* |
| Externalizing | .156*** | .082*** | .173*** | .052* |
| Internalizing | -.006 | -.051*** | -.061* | -.071*** |
| CSE | . | .008 | . | .123 |
| O Level | . | .092** | . | .134 |
| A Level | . | .204*** | . | .235** |
| Higher Education | . | .421*** | . | .230** |
| Higher Degree | . | .664*** | . | .689*** |
| london | . | .174*** | . | .334*** |
| Female | . | -.816*** | . | -.961*** |
| Has a Partner | . | .145*** | . | .224*** |
| Number of Children | . | -.152*** | . | -.053* |
| Experience | . | .003*** | . | .002*** |
| Const. | 5.295*** | 4.951*** | 5.094*** | 4.909*** |
| Obs. | 3552 | 3552 | 378 | 378 |

*Notes*: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to earnings. This table examines the possibility of non-linearities in the returns to externalizing behaviors. Models [3] and [4] include all individuals with externalizing behavior 1 standard deviation above the mean. Models [1] and [2] include the rest of the sample.

**Appendix Table S10:** Interactions

| Variable | [1] | [2] | [3] | [4] |
|---|---|---|---|---|
| Cognition | .090*** | .086*** | .089*** | .086*** |
| Cognition$^2$ | . | .012 | . | .012 |
| Externalizing | .058*** | .069*** | .057*** | .069*** |
| Externalizing$^2$ | . | -.005 | . | -.006 |
| Internalizing | -.053*** | -.063*** | -.052*** | -.064*** |
| Internalizing$^2$ | . | .004 | . | .006 |
| Ext.$\times$ Cog. | . | . | -.008 | -.007 |
| Int.$\times$ Cog. | . | . | .002 | .009 |
| Ext.$\times$ Int. | . | . | -.003 | .001 |
| Const. | 4.940*** | 4.917*** | 4.937*** | 4.915*** |
| Obs. | 3930 | 3930 | 3930 | 3930 |

*Notes*: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to earnings. To construct proxies for unobserved skills, we apply principal components factor analysis to all the variables used to measure that skill. All models include both male and female individuals and a gender dummy.

**Appendix Table S11:** Log Weekly Earnings at 42

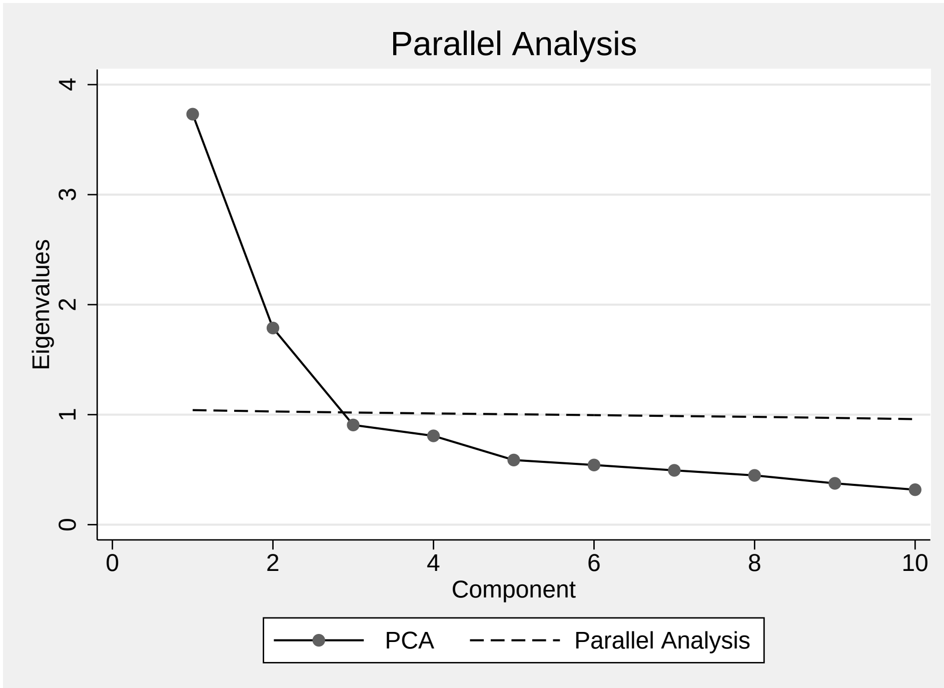
| Variable | [1] | [2] | [3] | [4] | [5] | [6] | [7] |
|---|---|---|---|---|---|---|---|
| Misbehavior | -.098*** | -.028* | . | . | . | . | . |
| Externalizing | . | . | .036** | .049*** | .054*** | .039** | .072** |
| Internalizing | . | . | -.064*** | -.051*** | -.037*** | -.046*** | -.026 |
| Cognition | . | .184*** | .186*** | .054*** | .040** | .067*** | .016 |
| CSE | . | . | . | .184*** | .131** | .152** | .153* |
| O Level | . | . | . | .281*** | .189*** | .154*** | .238*** |
| A Level | . | . | . | .424*** | .297*** | .272*** | .289*** |
| Higher Education | . | . | . | .623*** | .385*** | .271*** | .503*** |
| Higher Degree | . | . | . | .779*** | .606*** | .475*** | .708*** |
| Has a Partner | . | . | . | . | .018 | .202*** | -.107* |
| Number of Children | . | . | . | . | -.007 | .013 | -.044** |
| Experience | . | . | . | . | .002*** | .001*** | .002*** |
| Skilled Manual Occu. | . | . | . | . | .084** | .094** | -.010 |
| Skilled Non-manual Occu. | . | . | . | . | .108*** | .180*** | .094* |
| Managerial Occupation | . | . | . | . | .397*** | .331*** | .431*** |
| Female | -.985*** | -.968*** | -.964*** | -.907*** | -.795*** | . | . |
| London | .155*** | .133*** | .134*** | .122*** | .090*** | .197*** | -.020 |
| Const. | 6.105*** | 6.080*** | 6.080*** | 5.696*** | 5.231*** | 5.220*** | 4.626*** |
| Obs. | 4452 | 4452 | 4452 | 4452 | 4452 | 2226 | 2226 |

*Notes*: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to earnings at age 42. We regress log earnings of workers at age 42 on a set of observable variables at age 33 along with proxies for unobserved skills. The controls are all constructed for individuals when they were 33 years old. To construct proxies for unobserved skills, we sum up all variables used to measure that skill in subsequent analysis and then normalize each unobserved skill. Models [1]-[5] include all individuals and a gender dummy, Model [6] includes only males and Model [7] only females.

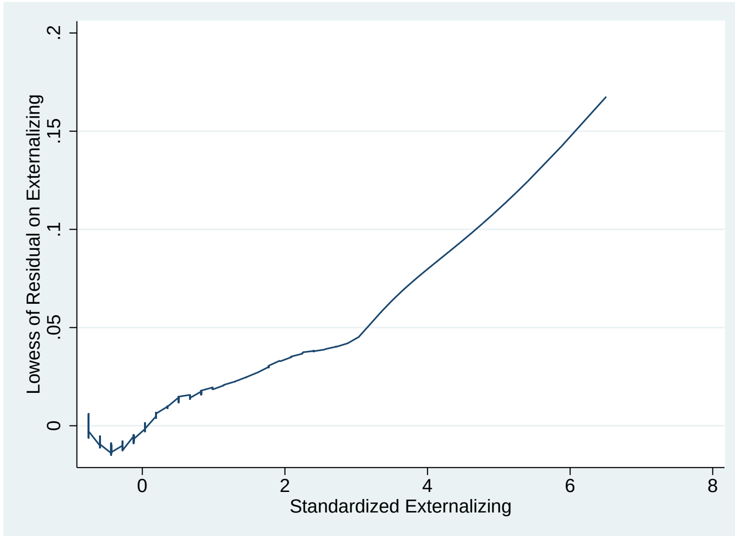
**Appendix Table S12:** Log Weekly Earnings at 50

| Variable | [1] | [2] | [3] | [4] | [5] | [6] | [7] |
|---|---|---|---|---|---|---|---|
| Misbehavior | -.101*** | -.033*** | . | . | . | . | . |
| Externalizing | . | . | .019 | .029** | .031** | .020 | .040* |
| Internalizing | . | . | -.052*** | -.045*** | -.032*** | -.045*** | -.017 |
| Cognition | . | .187*** | .188*** | .079*** | .067*** | .056*** | .079*** |
| CSE | . | . | . | .045 | .012 | .050 | -.001 |
| O Level | . | . | . | .135*** | .079** | .028 | .115** |
| A Level | . | . | . | .302*** | .213*** | .210*** | .195*** |
| Higher Education | . | . | . | .457*** | .286*** | .252*** | .317*** |
| Higher Degree | . | . | . | .582*** | .452*** | .413*** | .472*** |
| Has a Partner | . | . | . | . | .026 | .154*** | -.067 |
| Number of Children | . | . | . | . | .025** | .017 | .031* |
| Experience | . | . | . | . | .002*** | .0008** | .002*** |
| Skilled Manual Occu. | . | . | . | . | .071** | .081** | .064 |
| Skilled Non-manual Occu. | . | . | . | . | .089*** | .106** | .079** |
| Managerial Occupation | . | . | . | . | .329*** | .317*** | .322*** |
| Female | -.745*** | -.735*** | -.733*** | -.681*** | -.609*** | . | . |
| London | .191*** | .169*** | .168*** | .156*** | .142*** | .220*** | .070** |
| Const. | 6.393*** | 6.361*** | 6.362*** | 6.111*** | 5.705*** | 5.729*** | 5.130*** |
| Obs. | 3639 | 3639 | 3639 | 3639 | 3639 | 1723 | 1916 |

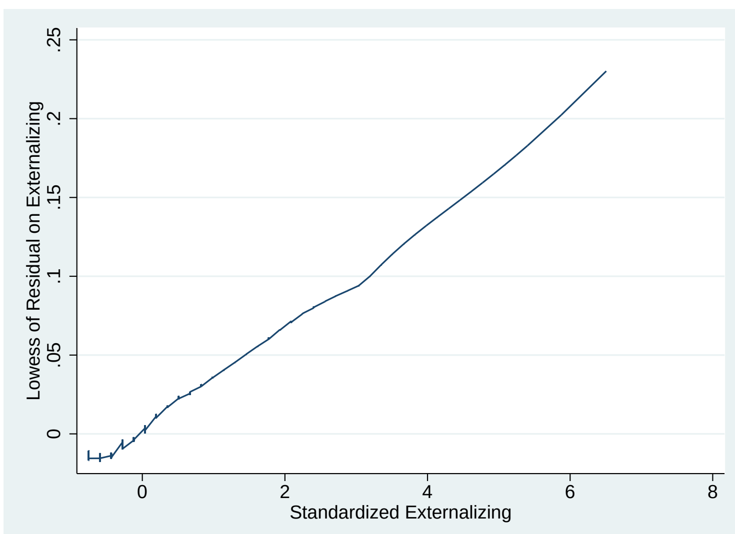
*Notes*: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to earnings at age 50. We regress log earnings of workers at age 42 on a set of observable variables at age 33 along with proxies for unobserved skills. The controls are all constructed for individuals when they were 33 years old. To construct proxies for unobserved skills, we sum up all variables used to measure that skill in subsequent analysis and then normalize each unobserved skill. Models [1]-[5] include all individuals and a gender dummy, Model [6] includes only males and Model [7] only females.




**Appendix Figure S1:** Factor Analysis. The solid line depicts the eigenvalues associated with each factor in descending order in the principal component analysis. The dashed line depicts the eigenvalues computed from the random data created by the parallel analysis. From the principal component analysis, both the Kaiser's criterion and the scree plot test suggest that we should only keep the first two factors. Moreover, since the two lines intersect before the third factor, the parallel analysis suggests that only the first two factors are informative for our analysis and factors 3 to 10 are mostly random noise.




**Appendix Figure S2:** Non-parametric regression. Here we use a non-parametric regression method (lowess) to plot the impact of the externalizing behavior on log-earnings controlled for cognition, internalizing behavior and a gender dummy. In order to control for the other variables we regress log-earnings on the other explanatory variables and use the residual of that regression as the explanatory variable in the non-parametric regression graphed here.




**Appendix Figure S3:** Non-parametric regression with controls. Here we use a non-parametric regression method (lowess) to plot the impact of the externalizing behavior on log-earnings controlled for cognition, internalizing behavior, a gender dummy, educational choices, partnership status, fertility and experience. In order to control for the other variables we regress log-earnings on the other explanatory variables and use the residual of that regression as the explanatory variable in the non-parametric regression graphed here.

## References


Bourguignon, François, Martin Fournier, and Marc Gurgand. 2007. "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons." *Journal of Economic Surveys* 21 (1):174–205.

Dubin, Jeffrey A and Daniel L McFadden. 1984. "An econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* :345–362.

Ghodsian, M. 1977. "Children's Behaviour and the BSAG: Some Theoretical and Statistical Considerations." *British Journal of Social and Clinical Psychology* 16 (1):23–28.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103 (6):2052–86.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* :153–161.

Lee, Lung-Fei. 1983. "Generalized Econometric Models with Selectivity." *Econometrica* :507–512.

Lundberg, Shelly. 2012. "Personality and Marital Surplus." *IZA Journal of Labor Economics* 1 (1):1–21.

Shepherd, Peter. 2013. "Bristol Social Adjustment Guides at 7 and 11 years." Centre for Longitudinal Studies.