NBER WORKING PAPER SERIES

THE ECONOMIC VALUE OF BREAKING BAD: MISBEHAVIOR, SCHOOLING AND THE LABOR MARKET

Nicholas W. Papageorge
Victor Ronda
Yu Zheng

Working Paper 25602
http://www.nber.org/papers/w25602

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2019, Revised August 2022

First draft: January 31, 2014. Current draft: August 16, 2022. We gratefully acknowledge helpful comments from: Robert Barbera, Pedro Carneiro, Sarah Cattan, Flavio Cunha, Seth Gershenson, Donna Gilleskie, Barton Hamilton, Hans von Kippersluis, Patrick McAlvanah, Robert Moffitt, Albert Park, Richard Spady, Sergio Urzua and Benjamin Williams along with seminar participants at the City University of Hong Kong, Tinbergen Institute, the Brookings Institution, Georgetown University, University of Western Ontario, Aarhus University, DePaul University, European University Institute, Stony Brook University, University of Cyprus, the Econometric Society World Congress, SOLE 2016, Barcelona GSE Summer Forum (Structural Microeconometrics). We acknowledge excellent research assistance from Mikkel Houmark. The usual caveats apply. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Economic Value of Breaking Bad: Misbehavior, Schooling and the Labor Market
Nicholas W. Papageorge, Victor Ronda, and Yu Zheng
NBER Working Paper No. 25602
February 2019, Revised August 2022
JEL No. I20,J10,J20

ABSTRACT

Prevailing research argues that childhood misbehavior in the classroom is bad for schooling and, presumably, bad overall. We examine lifecycle impacts of a widely studied socio-emotional skill, externalizing behavior, which captures childhood misbehavior in school and is linked to aggression, hyperactivity and lower educational attainment. Externalizing behavior has been the focus of hundreds of papers across several fields and its negative impact on educational attainment has justified multitudes of policies to address or discourage it. Aligned with prior work, we find that externalizing behavior lowers educational attainment for males. However, we also provide novel evidence that it increases earnings for males and females. The earnings premium holds across genders and occupations, is replicated in several data sets and is robust to alternative modeling assumptions. That a skill can be both helpful and harmful raises concerns about policies surrounding skill acquisition, especially if the skill has opposite effects across crucial or largely unavoidable phases of the lifecycle, such as schooling and work. For example, well-meaning policies to improve schooling can have negative repercussions over the lifecycle. More broadly, our results illustrate the need to measure skill prices across contexts and challenge the widespread, implicit and largely untested assumption that the same skills that are valuable in childhood are also valuable in adulthood.

Nicholas W. Papageorge
Department of Economics
Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218
and IZA
and also NBER
papageorge@jhu.edu

Victor Ronda
The University of Chicago
Center for the Economics of
Human Development
1126 E. 59th Street
Chicago, IL 60637
vronda@uchicago.edu

Yu Zheng
School of Economics and Finance
Queen Mary University of London
Mile End, London E1 4NS
Great Britain
and CEPR
yu.zheng@qmul.ac.uk

An online appendix is available at http://www.nber.org/data-appendix/w25602

1 Introduction

Economic research has shown that human capital consists of multiple factors that drive lifecycle behavior and outcomes and that the returns to these factors can vary across sectors and contexts. For example, to explain career choices, Willis and Rosen (1979) emphasize differences across occupations in returns to manual versus academic skill.[1] However, this research has generally ignored the possibility that a component of human capital that is productive in some sectors of the economy can be counter-productive in others. For some factors constituting human capital, such as cognition, it is indeed difficult to imagine cases in which more is not better; and while a factor such as mechanical skill may not be particularly useful in some occupations, there is little reason to think that it should make someone less productive since it can remain idle. For socio-emotional skills, including personality traits—now widely viewed as crucial components of human capital (Heckman and Rubinstein, 2001)—it is not conceptually obvious if more is always better (or even harmless). The possibility that socio-emotional skills can be both helpful and harmful raises concerns about policies designed to either curb or encourage them. This is especially worrisome if a skill has opposite effects across unavoidable phases of the lifecycle, such as schooling and work. In this case, an intervention designed to improve school performance that targets children's socio-emotional skills (which are relatively malleable (Heckman and Kautz, 2014)) could have negative repercussions over the lifecycle.

In this paper, we examine a widely-studied pair of socio-emotional skills measured among schoolchildren to capture misbehavior: *externalizing behavior* and *internalizing behavior*.[2] Externalizing behavior is linked to aggression and hyperactivity, while internalizing behavior captures anxiety, depression, shyness, unassertiveness and fearfulness. The conceptual development of externalizing and internalizing behaviors dates back to Achenbach (1978). Since then, they have been measured in dozens of data sets and have been the subject of literally hundreds of studies.[3] Many studies focus on their negative impacts on educational outcomes, which has led to a multitude of interventions and programs in classrooms and

[1] This point has its origins in Roy (1951) and Mandelbrot (1962), which are later developed into a model of comparative advantage and self-selection in the labor market by Willis and Rosen (1979), Heckman and Sedlacek (1985), and many papers thereafter.

[2] Regarding the nomenclature: "externalizing behavior" and "internalizing behavior" describe the two socio-emotional skills (sometimes called noncognitive skills) that are measured using teachers' reports of childhood maladjustment or misbehavior.

[3] When we compare the number of publications in the PubMed database that mention internalizing or externalizing to those that mention the "Big Five" personality traits, we find that since 1980 there are strong increasing trends in the number of both and that the number of publications related to internalizing and externalizing is roughly double the number of those related to the "Big Five" (see Appendix A).

schools, mostly designed to address, discourage or curb externalizing behavior.[4] The key empirical fact we establish is that externalizing behavior increases earnings for men and women and also lowers educational attainment for boys. Our findings are evident in preliminary analyses of the data, are robust to alternative modeling assumptions and can be replicated across data sets. In other words, we provide robust evidence that a well-known and well-studied socio-emotional skill can have mixed effects in school versus at work. In short, *Breaking Bad* can be good—or at least lucrative.[5]

Our main analysis uses a longitudinal dataset from Britain, the National Child Development Survey (NCDS), to estimate an econometric model relating childhood misbehavior to educational attainment and labor market outcomes. We approximate schooling, hours of work and wages using linear-in-parameters equations, and we model correlation across equations as unobserved heterogeneity in the form of three latent factors identified using a measurement system. The first two latent factors capture the socio-emotional skills described above and are measured using multiple teachers' reports of children's misbehavior or maladjustment in school. The third factor captures cognition and is measured using math and reading test scores. Given earlier work showing that boys are more externalizing than girls (see, e.g., Bertrand and Pan (2013)), we estimate the model separately for males and females to ensure we can make within-gender comparisons to identify the impacts of externalizing behavior. Gender interactions extend to how we measure skills (including, e.g., possible gender differences in how teachers report misbehavior).

We investigate whether our results are driven by selection, especially for boys, for whom externalizing has opposite effects in school versus work. One possibility is that externalizing behavior does not make schooling more difficult for boys, but instead leads to lower relative returns to education, incentivizing less education (Prada and Urzúa, 2017). We reject this hypothesis by showing that externalizing behavior does not interact with the returns to education in earnings equations. Another possibility is that, since externalizing behavior makes schooling more difficult for boys, it generates positive selection on other productive

[4]The constructs of externalizing and internalizing behaviors are well established in the developmental psychology literature (Ghodsian, 1977; Campbell, Shaw, and Gilliom, 2000; Eisenberg et al., 2001; Duncan and Magnuson, 2011). Stage and Quiroz (1997) and O'Connor and Hayes (2020) conduct meta-analysis of 99 and 17 journal articles, respectively, about interventions to target and address disruptive behavioral problems in public education, mostly in the US. Similar interventions have also been introduced in Europe (e.g., Humphrey et al. (2010) and Närhi, Kiiski, and Savolainen (2017) in the UK and Sørlie and Ogden (2015) in Norway). Externalizing and internalizing behaviors have also been studied in economic research (see, e.g., Neidell and Waldfogel (2010); Duncan and Dunifon (2012); Bertrand and Pan (2013); Gertler et al. (2014); Heckman, Pinto, and Savelyev (2013); Doyle (2020)).

[5]According to www.urbandictionary.com the definition of the term breaking bad is to "challenge conventions" or to "defy authority." Breaking Bad is also the title of a television show in which the protagonist is an unsuccessful chemist with a talent for producing illicit drugs, illustrating how certain skills or behaviors may lead to low productivity in one sector and high productivity in another.

factors, such as cognition, that are captured as an externalizing premium on the labor market. We account for this possibility by including cognition in our model. In a robustness check, we also allow for additional unobserved heterogeneity in the form of a fourth latent factor to capture additional potential sources of positive selection, and find that our main results remain intact. We also examine the role of selection into additional outcomes, including marriage and fertility. High-externalizing people may be less likely to have a family and thus focus on their careers. Instead, estimates show that externalizing behavior increases marriage and fertility both for men and women. While this has little bearing on men's earnings, high-externalizing women are less likely to work since they are more likely to have children. Thus, for females, conditioning on fertility increases the externalizing premium. This evidence (along with additional analyses and robustness checks) is consistent with the idea that externalizing behavior is a productive skill on the labor market that, for boys, is also unproductive at school.

To explore channels that could explain our results, we investigate whether the externalizing premium is limited to certain occupations or job tasks. If so, any debate about policies curbing externalizing behaviors might be inconsequential since the risk of inadvertently lowering earnings would be limited to small sectors of the labor market. We reject this hypothesis. There are very few job tasks and virtually no occupations in which the returns to externalizing are negative. Another possibility is that the returns to externalizing behavior fail to materialize in work environments that value social niceties in ways not captured by occupational or task categories. Lacking a direct test, we speculate that such positions are likely filled by individuals from advantaged backgrounds, who would thus exhibit lower returns to externalizing behavior. Evidence suggests otherwise: returns to externalizing are statistically equivalent by childhood socioeconomic status. Yet, estimated coefficients are smaller for individuals from disadvantaged families. Wealthier children may be given the benefit of the doubt when acting aggressively in the workplace. Alternatively, they may learn to channel their childhood misbehavior productively, which might be captured through personality traits measured in adulthood. Yet we find only weak relationships between externalizing behavior and the "Big 5" personality traits that cannot explain the main patterns we estimate. An examination of stress hormones yields similar results. The bottom line is that, while we are able to rule out a number of possible explanations, conclusive evidence on specific channels is elusive.

Nevertheless, our findings on externalizing behavior challenge a large literature on child development focused on skills measured during childhood and their impact on schooling. Externalizing behavior in particular has been examined in hundreds of papers (see Malti and Rubin (2018) for a review). The implicit assumption in much of this literature appears to be that whichever skills, characteristics or traits are bad for schooling must be bad overall. Based on this assumption, researchers have generally failed to directly examine long-run impacts of skills shown to be detrimental for schooling, which forestalls exploration of the potential downsides of policies aiming to curb supposedly negative traits. This is not just a theoretical problem. School districts, states and regions across the world have adopted programs that are meant to develop socio-emotional skills that promote behaviors deemed good for school (O'Connor and Hayes, 2020; Nangle, Erdley, and Schwartz-Mette, 2020).[6] Since the long-run costs are poorly understood and could be substantial, it is not clear if such interventions are efficient or, indeed, even ethical. For example, if an intervention aims to improve behavior in school, would parents agree to it if they were made aware that it might lead their children to earn less?

More broadly, our main findings on externalizing behavior suggest a need to extend how we conceptualize and measure human capital to incorporate the possibility that some skills can be both helpful and harmful depending on the context. A vast and active literature in economics has continuously expanded our understanding of what constitutes human capital beyond traditional factors such as cognition, education and work experience to include, for example, health (Grossman, 2000) and socio-emotional skills (Heckman and Rubinstein, 2001).[7] This literature continues to generate new insights, including differences in the returns to schooling across sociodemographic groups (Lundberg, 2013), the role of social norms in investing in human capital (Lundberg and Pollak, 2007), the impact of biased beliefs about returns to investments (Cunha, Elo, and Culhane, 2020), and the importance of early childhood investments given dynamic complementarities (Cunha and Heckman, 2008), to name a few. Nevertheless, the idea that a skill can have both positive and negative returns—and the consequences of policies surrounding skill acquisition—has generally been overlooked.

A notable exception is Levine and Rubinstein (2017), who show that individuals who engaged in illicit behaviors (e.g., drug use or petty crime) as teenagers earn more as incorporated entrepreneurs compared to those who did not engage in such behaviors. The idea, similar to ours, is that whatever factors drive illicit behavior during adolescence could prove helpful in adulthood. Our research shows that this type of sign reversal in skill prices is not limited to a small sliver of the labor market (3.4% of their sample consists of incorporated

[6]Some well-known interventions include the Check-In/Check-Out targeted intervention, a.k.a the Behavior Education Program (Crone, Hawken, and Horner, 2003; Campbell and Anderson, 2011), and the Fast Track prevention program (CPPRG, 1992). Some element of the intervention is designed to curb misbehavior, e.g., setting a goal of "not leaving my seat once without permission" under the Check-In/Check-Out system, while others focus on encouraging prosocial behaviors, e.g., providing social skills training under the Fast Track prevention program. However when it comes to the evaluation of interventions, reduction in occurrences of misbehavior is invariably the most common measure.

[7]Excellent summaries of this research are found in Borghans et al. (2008) and Almlund et al. (2011).

entrepreneurs). This is an important distinction since workers without the proper skills could presumably avoid a small sector of the economy. In contrast, we find that a broadly distributed skill has opposite effects on two crucial and often unavoidable phases of the lifecycle: schooling and work—and that this holds across occupations.[8] Another important distinction is that, while we empirically investigate a host of possible channels, Levine and Rubinstein (2017) do not. Instead, they speculate that the link between illicit behavior and business success is due to creative thought. This explanation seems plausible, and suggests that what underlies illicit behaviors could be socially beneficial. Yet, absent any evidence to support their claim, one can imagine other possibilities.[9] An alternative interpretation is that illicit teenagers are dishonest in adulthood and engage in illegal business activities or accounting practices, which are lucrative, but not socially beneficial. This example illustrates the importance of additional empirical analysis when drawing conclusions about skills.

Thus, the key implications of our findings are caution and further research. Caution is needed because earlier research, including vast amounts of research on externalizing behavior, has failed to recognize the possibility that promoting a skill that is useful in one context can be counterproductive in another. Policies to promote good behavior in school, based on the largely untested and faulty assumption that doing so can only help, might actually harm children when they grow up. However, inconclusive evidence on channels, i.e., the specific actions and choices that underlie positive returns to externalizing behavior in the labor market, further complicates policy evaluation. Some specific manifestations could be socially beneficial (e.g., a willingness to challenge unproductive social norms), while others may not (e.g., bullying or cheating). Broadly, any policy conclusions related to complex skills or traits—especially one that can be helpful or harmful depending on the context, such as externalizing behavior among males—require a more thorough understanding of why they are priced so differently over the lifecycle.

Section 2 introduces the NCDS dataset, discusses measurements of misbehavior that identify externalizing and internalizing behavior, and presents a preliminary data analysis. Section 3 describes the econometric model we estimate and main results along with sensitivity analyses and robustness tests. Section 4 explores channels. Section 5 concludes.

[8]An additional difference is that externalizing behavior is productive even if not accompanied by high cognition.

[9]In results available upon request (and using the same data set as Levine and Rubinstein (2017)), we are unable to find evidence that illicit behavior is correlated to two measures of creativity: self-reports of being unconventional and interests in arts and music.

2 Data and Preliminary Analysis

2.1 The National Childhood Development Study

This project uses data from the National Childhood Development Study (henceforth, NCDS), which is an ongoing longitudinal survey following the universe of individuals born in the same week in 1958 in Great Britain. It is particularly well-suited for our study since it collects teachers' reports of classroom misbehavior for a large sample of children and then follows these children through adulthood, meaning that we can relate misbehavior in elementary school to educational attainment and labor outcomes. To date, there have been eleven surveys to trace all the members of the cohort still living in Great Britain. Surveys occurred when subjects were born and when they were aged 7 (1965), 11, 16, 23, 33, 42, 44, 46, 50 and 55 (2013).

We focus on information gathered at birth and in the first five sweeps, covering ages 7 to 33. The NCDS initially contained information on 18,555 births. In constructing our analytic sample, we keep respondents with valid information on test scores and classroom misbehavior at age 11 and educational attainment and labor outcomes at age 33. We drop individuals with missing information on variables treated in some of our analyses as intermediate outcomes, such as relationship status, fertility, employment status and employment history. We also drop individuals who are reported as employed but have missing information on earnings at age 33. We impute data for individuals with missing information on variables used in some specifications as controls, such as parents' education and occupation. The resulting analytic sample has information on 7,241 individuals, of whom 3,573 are males and 3,668 are females.[10] We present summary statistics for the key variables used in the paper in Appendix B.

2.2 Key Variables and Summary Statistics

2.2.1 Education and Labor Outcomes

In the UK, schooling is compulsory until age 16. Thereafter, students can leave school without any qualifications (no certificate), study for an exam to obtain a Certificate of Secondary Education (CSE) or study towards obtaining the Ordinary Levels (O-Levels),

[10] Most of the drop in observations is due to attrition at the fifth survey, where only 11,364 out of the original 18,555 births were surveyed in 1991 at age 33. To assess whether sample attrition drives our main results, we compare our analytic sample to the sample of all individuals observed at age 11, which we call the "full sample" (see Appendix B). Compared to the full sample, our analytic sample is slightly more educated, less likely to be self-employed, receives slightly lower wages and works fewer hours. However, none of these differences are statistically significant.

where the latter are more academically demanding.[11] Individuals aiming to attain a higher degree take another set of examinations, the Advanced Levels (A-Levels). Students who are successful in their A-Levels are able to continue to attain either a higher-education diploma (after two years of study) or a bachelor's degree (after three years of study). At the postgraduate level, students can obtain a higher degree: Master of Philosophy (MPhil) or Doctor of Philosophy (PhD). In summary, individuals in our sample can sort into six mutually exclusive schooling levels: no certificate, CSE, O-Levels, A-Levels, higher education (including diploma and bachelors) or higher degree (including MPhil and PhD).

2.2.2 Socio-Emotional Skills and Cognition

Next, we discuss variables used to construct measures of unobserved skills, including the two socio-emotional skills that are the focus of our analysis, along with cognition. Following earlier work (see e.g., Cunha, Heckman, and Schennach, 2010), we measure cognitive skill using a set of math and reading test scores. These tests are administered when children are 11 years old. We measure socio-emotional skills using variables describing classroom misbehavior. When a child in the sample was 11 years old, the child's teacher was asked to complete an inventory listing the child's behaviors in the classroom. The teacher was given a list of roughly 250 descriptions of specific behaviors and asked to underline the items which best describe the child. These descriptions include statements such as: "too timid to be naughty," "brags to other children," "normally honest with school work," "adopts extreme youth fashions," and "has stolen money."

Completed inventories were then used to compute scores on a set of ten summary variables known as the Bristol Social Adjustment Guide or *BSAG* maladjustment syndromes.[12] The ten syndromes are: hostility towards adults, hostility towards children, anxiety for acceptance by adults, anxiety for acceptance by children, restlessness, inconsequential behavior, writing off adults and adults standards, depression, withdrawal, and unforthcomingness. The syndromes have been used since their introduction in Stott, Sykes, and Marston (1974) to assess the psychological development of children. The maladjustment syndrome scores range from 0 to 15 but most individuals have a score near 0 (see Table B2 in Appendix B). While most individuals score at or near zero on most maladjustment syndromes, the median student has a score of 4 and few students (15.5%) score zero on all of them. This means that

[11] CSEs and O-Levels were replaced by the General Certificates of Secondary Education (GCSE) in 1986 after individuals in our sample had finished their schooling.

[12] The full list of the behavioral inventory is available in the Appendix of Shepherd (2013). In particular, each item on the inventory was assigned to one of 10 syndromes and the variables are the sum of these items from the teacher inventories. Unfortunately, the original teacher inventory data are not available. If they were, one could use them directly to identify latent skills.

our results are not driven by a small percentage of very poorly behaved students.

2.3 Preliminary Analysis

Before presenting the formal econometric model used in our main analysis, we explore the basic patterns in our data. We follow previous work by Ghodsian (1977) to construct crude measures of socio-emotional skill. We divide the BSAG syndromes into two groups based on apparent differences among what behaviors the syndromes capture.[13] The first variable, *externalizing behavior*, is constructed from summing over maladjustment syndromes such as "hostility towards adults" and "restlessness" among others, and expresses anxious, aggressive, and outwardly-expressed behavior. The second variable, or *internalizing behavior*, is constructed by summing over maladjustment syndromes such as "depression" and "withdrawal" among others, and expresses withdrawn and inhibited behavior. In addition, we construct a generic measure of misbehavior by simply summing up all ten syndromes. This variable better resembles previous work on misbehavior and will help illustrate how findings change once we recognize that misbehavior captures two separate socio-emotional skills. Similarly, we obtain a measure of cognitive ability by summing test scores. Table B5 in Appendix B summarizes the approach.

We relate the crude measures of misbehavior, externalizing behavior, internalizing behavior, and cognition to schooling and labor market outcomes in Table 1. In Columns (1) to (4) we link unobserved skills to educational attainment using an ordered logit model. Column (1) shows that the aggregated measure of misbehavior is strongly negatively correlated with educational attainment, while a higher level of cognition leads to higher educational attainment. Disaggregating misbehavior, in Column (2), yields similar results. It shows that externalizing and internalizing behaviors independently lower schooling attainment. Columns (3) and (4) repeat the analysis but separately by gender. The externalizing schooling penalty is more pronounced for boys than for girls. So far, these results are not surprising and mirror what has been found in the literature using the NCDS (e.g., Farmer, 1995) and other data (e.g., Duncan and Magnuson, 2011). In Columns (5) to (9), we regress log weekly earnings at age 33, conditional on being employed, onto measures of socio-emotional and cognitive skills. In line with previous research (e.g., Segal, 2013), Column (5) shows that the aggregated measure of misbehavior is associated with both lower schooling and lower earnings.

[13]This division proposed in Ghodsian (1977) is also motivated by a principle components factor analysis, which suggests there are two underlying latent factors measured by the BSAG syndromes. The same two stable factors, representing the externalizing and internalizing taxonomy, are found in other similar analysis (Achenbach, 1978). We replicate this factor analysis in work available upon request. In addition, the results from our benchmark model support this division.

Results change dramatically when we view childhood misbehavior as reflecting two distinct factors in Column (6). While internalizing behaviors lower earnings, externalizing behaviors seem to lead to higher earnings. The positive association between externalizing behavior and earnings is a new finding in the literature on the economic consequences of childhood misbehavior. This positive association remains, and becomes stronger, once we control for educational attainment in Column (7) and is present for both males in Column (8) and females in Column (9). These results provide initial evidence that a socio-emotional skill that is productive in the labor market is not productive in school and highlight the importance of considering the two dimensions of misbehavior separately. We build on this preliminary work in the remainder of the paper. We develop a formal latent factor model, which treats observed maladjustment syndromes and test scores as measures with error of underlying skills. The formal model accounts for measurement error and correlation across measures, and allows us to identify the different underlying skills and their effects on adult outcomes. We use the formal model to explore gender differences and decompose the impact of skills on earnings into the separate impacts on wages and hours worked. We also test and show our findings' robustness across various assumptions and explore mechanisms.

3 Model and Main Estimates

This section presents the benchmark econometric model and explains how we identify and estimate model parameters (Sections 3.1 and 3.2). Next we present model estimates, including the finding that externalizing has opposite effects on schooling versus earnings (Section 3.3). Finally, we note that the model requires parametric and identifying assumptions, some of which are testable. Where possible, we assess sensitivity of the main results to alternative modeling assumptions; we also provide evidence that main results are robust across several data sets (Section 3.4).

3.1 Model

3.1.1 Measurement System for Unobserved Skills f

Let the vector of unobserved non-cognitive skills and cognition be denoted $\mathbf{f}$, which can nevertheless be proxied by a set of observable measurements such as the ten BSAG maladjustment syndromes and the four aptitude test scores measured at the age of 11. Specifically, let M be a vector of $K = 14$ measurements of the three latent skills $\mathbf{f} = (f_1, f_2, f_3)$, where

f_1 is externalizing behavior, f_2 is internalizing behavior and f_3 is cognition.

$$M = \begin{pmatrix} M_1 \\ \vdots \\ M_K \end{pmatrix} = \begin{pmatrix} m_1 + \sum_{j=1}^{3} \lambda_{1j} f_j + \ W\delta_1 + \varepsilon_1 \\ \vdots \\ m_K + \sum_{j=1}^{3} \lambda_{Kj} f_j + W\delta_K + \varepsilon_K \end{pmatrix}, \tag{1}$$

where m_k is the mean of the measurement k, and λ_{kj} is the factor loading of latent skill j on the kth measurement. Given many zeros on BSAG maladjustment scores, we use the logarithm of each BSAG score plus one as the relevant measurement in the measurement system. The latent skills follow a joint normal distribution, with mean μ and variance-covariance matrix $\mathbf{\Sigma}$:

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \sim N(\mu, \mathbf{\Sigma}) = N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}\right). \tag{2}$$

In the measurement system described by equations (1) and (2), we also include a vector of additional observables denoted W (and associated coefficients δ): class size, the percentage of students in the respondent's school taking GCE exams, a dummy for school managed by the local educational authority (henceforth LEA and similar to a school district), and the number of full-time teachers in the school. These additional variables are included since we may otherwise misattribute variation in outcomes to variation in skills that is actually due to differences in schooling attributes or systematic variation in how teachers report misbehavior that could relate to teacher or school quality.

As suggested by Williams (2020), to identify the measurement system, we assume three "dedicated measures," one for each skill. That is, for each skill we choose one measure that is only affected by that skill. We choose "hostility towards children" (M_1) as the dedicated measurement for externalizing behavior (f_1), "depression" (M_2) as the dedicated measurement for internalizing behavior (f_2), and "verbal ability" (M_3) for cognition (f_3), in which case the relevant measurement equations are reduced to

$$\begin{aligned} M_1 &= m_1 + 1 \cdot f_1 + 0 \cdot f_2 + 0 \cdot f_3 + W\delta_1 + \varepsilon_1 \\ M_2 &= m_2 + 0 \cdot f_1 + 1 \cdot f_2 + 0 \cdot f_3 + W\delta_2 + \varepsilon_2 \\ M_3 &= m_3 + 0 \cdot f_1 + 0 \cdot f_2 + 1 \cdot f_3 + W\delta_3 + \varepsilon_3 \end{aligned} \tag{3}$$

For the remaining 11 measurements, we allow all three skills to load on them.

The choice of dedicated measures is a matter of judgement and is motivated by how we interpret each of the factors. Literature in psychology and medicine posits that externalizing behavior is closely associated with disruptive disorders, which motivates our choice of "hostility towards children" as the dedicated measurement (Duncan and Magnuson, 2011; Kendler and Myers, 2014). Internalizing behavior is commonly associated with depressive disorders, which motivates our choice of "depression" as the dedicated measurement (Regier, Kuhl, and Kupfer, 2013; Kendler and Myers, 2014). Finally, as factors do not have a natural scale, we normalize the coefficients of the dedicated measurements to unity as is commonly done in this literature.

3.1.2 Parameterizations of the Schooling Decision Rule and Labor Market Outcomes

We approximate the schooling decision with a linear-in-parameters multinomial logit model, with six schooling levels from no certificate to higher degree: $s \in \{0, 1, ..., 5\}$. Taking schooling level 0 as the base state, let the log-odds of schooling level s be

$$I_s = \log \frac{Pr(S = s)}{Pr(S = 0)} = \underbrace{Z}_{\substack{\text{observed by} \\ \text{econometrician}}} \cdot \beta_s + \underbrace{\eta_s}_{\substack{\text{unobserved by} \\ \text{econometrician}}}, \quad s = \{1, ..., 5\}, \tag{4}$$

where Z is a vector of variables observed by the econometrician that affect the schooling decision. It includes two basic controls, a dummy for the child experiencing financial difficulty at home and a London dummy, which we will include in all outcome equations along with parents' education and occupation status. β_s is a vector of parameters mapping variables in Z to schooling outcomes and η_s is a set of school-level-specific shocks that are unobserved by the econometrician. We impose separability between the observed and unobserved variables in the representation of the schooling decision rule.

We focus on two labor market outcomes in the benchmark model: the hourly wage and the weekly hours worked for individuals who are employed at age 33.[14] More specifically, the log hourly wage, y, and the log weekly working hours, h, are represented by the following

[14] To avoid biases due to non-random attrition, we focus on age 33 earnings. In descriptive analyses (results of which are available upon request), we find similar returns to externalizing behavior among individuals when they are age 42 and age 50, suggesting that the externalizing premium extends over the lifecycle.

two equations:

$$
\begin{aligned}
y &= \underbrace{X}_{\substack{\text{observed by}\\ \text{econometrician}}} \cdot\beta_Y + \sum_{s=1}^{5}\gamma_{s,Y}\cdot \underbrace{\mathbf{1}[s]}_{\substack{\text{observed by}\\ \text{econometrician}}} + \underbrace{U_Y}_{\substack{\text{unobserved by}\\ \text{econometrician}}} &(5)\\
h &= \underbrace{X}_{\substack{\text{observed by}\\ \text{econometrician}}} \cdot\beta_H + \sum_{s=1}^{5}\gamma_{s,H}\cdot \underbrace{\mathbf{1}[s]}_{\substack{\text{observed by}\\ \text{econometrician}}} + \underbrace{U_H}_{\substack{\text{unobserved by}\\ \text{econometrician}}}. &(6)
\end{aligned}
$$

X is the set of basic controls and the β's are vectors of associated coefficients. $\mathbf{1}[s]$ is an indicator function indicating the observed schooling level with associated coefficients γ. U_Y and U_H are unobserved determinants of wages and hours worked. We summarize the observable controls, W, Z and X, in the measurement, schooling and labor market outcome equations in Table 2.

We assume that all dependences across the unobserved components in the schooling and outcome equations, η_s, U_Y, and U_H, are generated by the vector of skills, $\mathbf{f}$, unobserved by the econometrician. More specifically, suppose

$$
\begin{aligned}
\eta_s &= \mathbf{f}'\alpha_{\mathbf{S}} + \nu_s, &(7)\\
U_Y &= \mathbf{f}'\alpha_{\mathbf{Y}} + \omega_Y, &(8)\\
U_H &= \mathbf{f}'\alpha_{\mathbf{H}} + \omega_H, &(9)
\end{aligned}
$$

where the α's are equation-specific vectors of coefficients attached to latent skills $\mathbf{f}$, ν_s is a normal idiosyncratic error term for the schooling choice, and ω_Y and ω_H are normal idiosyncratic error terms for the two labor market outcomes, the log hourly wage and the log weekly hours worked.

3.1.3 Identifying Assumptions

The key identifying assumption is that conditional on $\mathbf{f}$, Z, and X, choices and outcomes are statistically independent. Formally, we array the ν_s, $s \in \{1, ..., 5\}$ into a vector $\nu = (\nu_1, \nu_2, \nu_3, \nu_4, \nu_5)$ and array ω_Y and ω_H into a vector $\omega = (\omega_Y, \omega_H)$. We assume that,

$$
\begin{aligned}
\nu_s &\perp\!\!\!\perp \nu_{s'}, \forall s \neq s', &(10)\\
\omega_Y &\perp\!\!\!\perp \omega_H, &(11)\\
\omega &\perp\!\!\!\perp \nu. &(12)
\end{aligned}
$$

Assumptions (10), (11) and (12) maintain independence of the shocks over schooling categories, and across schooling and labor market outcomes. In addition, we array the measurement errors, ε_k, $k \in \{1, ..., K\}$ into a vector $\varepsilon = (\varepsilon_1, ..., \varepsilon_K)$ and assume that,

$$\begin{aligned} \varepsilon_k &\perp\!\!\!\perp \varepsilon_{k'}, \forall k \neq k', \quad (13) \\ (\omega, \nu) &\perp\!\!\!\perp \varepsilon. \quad (14) \end{aligned}$$

Assumptions (13) and (14) maintain that the measurement errors are independent from each other, and independent from the shocks.[15] Last, we assume that,

$$\begin{aligned} (\nu, \omega, \varepsilon) &\perp\!\!\!\perp (\mathbf{f}, Z, X, W), \quad (15) \\ \mathbf{f} &\perp\!\!\!\perp (X, Z, W). \quad (16) \end{aligned}$$

Assumption (15) assumes independence of all the shocks and measurement errors with respect to factors and observables, and Assumption (16) assumes independence of factors with respect to observables.[16]

3.2 Estimation

We summarize the parameters to be estimated by a vector denoted Φ:

$$\Phi = (\beta, \gamma, \alpha, \Xi) \quad (17)$$

where β denotes the set of coefficients on the vectors of observables absent the schooling level in equations (4)-(6), γ is the set of coefficients governing the returns to schooling, α is the set of coefficients governing the returns to unobserved skills and Ξ are coefficients of the measurement system described in equations (1) and (2).

We estimate the model by simulated maximum likelihood in two stages and allow all parameters to differ by gender. In the first stage, we estimate the measurement system for unobserved skills. For each suggestion for parameters in the measurement system indexed by g_1 and denoted $\Xi^{(g_1)}$, and for each individual i, we simulate a vector of unobserved factors $T = 500$ times and, for each draw of the factors, compute the probability of observing each

[15] In a robustness check, we allow for correlation among some of the error terms in our measurement system. We allow the errors for anxiety towards children and anxiety towards adults to be correlated, and for hostility towards children and hostility towards adults to also be correlated. In both cases and for both genders, the estimated correlation is zero. These results are available upon request.

[16] Williams (2020) discusses these assumptions in more detail. In particular,Williams (2020) describes conditions under which Assumption (16) can be relaxed.

measurement. More specifically, given a parameter suggestion, we draw a block matrix of size $T \times I \times J$ from a standard normal distribution, where J is the number of latent factors (i.e., 3), and I is the number of individuals. Then, for each individual i and draw t, we construct a vector of latent factors $(f_{i1t}^{(g_1)}, f_{i2t}^{(g_1)}, f_{i3t}^{(g_1)})$ and compute $\mathbf{f}_{it}^{\mathbf{M},(g_1)}(M_i)$, the probability of observing the classroom misbehavior measurements and test scores, for individual i, draw t and parameter suggestion (g_1). The simulated log likelihood function is computed as the sum of the log of each individual's average likelihood contribution taken over the T draws:

$$\mathcal{L}_1^{(g_1)} = \sum_{i=1}^{I} \log \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{f}_{it}^{\mathbf{M},(g_1)} (M_i) \right) \tag{18}$$

Using both simplex and gradient methods, we evaluate $\mathcal{L}_1^{(g_1)}$ at different values in the parameter space, indexing these suggestions by (g_1), and continue until a maximum is found.

In the second stage, given the parameter estimates $\widehat{\Xi}$ found in the first step, we estimate the remaining structural parameters, (β, γ, α).[17] Taking $\widehat{\Xi}$ as given, we follow a similar procedure to compute the density functions corresponding to each outcome: the probability of individual i reaching a schooling level s, $\left(\mathbf{f}_{it}^{\mathbf{S},(g_2)}(s)\right)$, the probability of observing wage y_i, $\left(\mathbf{f}_{it}^{\mathbf{Y},(g_2)}(y_i)\right)$, and hours worked h_i, $\left(\mathbf{f}_{it}^{\mathbf{H},(g_2)}(h_i)\right)$, for individual i, draw t and parameter suggestion (g_2). The simulated log likelihood in the second stage is given by:

$$\mathcal{L}_2^{(g_2)} = \sum_{i=1}^{I} \log \Bigg(\frac{1}{T} \sum_{t=1}^{T} \mathbf{f}_{it}^{\mathbf{M},(\widehat{\Xi})} (M_i) \times \prod_{s=0}^{5} \mathbf{f}_{it}^{\mathbf{S},(g_2)}(s)^{\mathbf{1}[s=s_i]} \\ \times \mathbf{f}_{it}^{\mathbf{H},(g_2)} (h_i)^{\mathbf{1}(e_i=1)} \times \mathbf{f}_{it}^{\mathbf{Y},(g_2)} (y_i)^{\mathbf{1}(e_i=1)} \Bigg) \tag{19}$$

where s_i represents the observed schooling choice and e_i the observed employment status (with employed taking the value 1) in the data.[18]

3.3 Empirical Results

For clarity, we report the estimation results regarding the key variables in the measurement system, the multinomial logit model for education attainment, and the linear model for hours and wages in Tables 3 to 5 in this section and relegate the complete set of estimation results

[17] In Appendix C, we present estimates from an alternative specification where we estimate the measurement system jointly with outcomes. Results remain largely unchanged.

[18] Standard errors are computed by constructing the Hessian of the joint likelihood function using the outer product measure. To compute the outer product measure, we calculate two-sided numerical derivatives of the joint likelihood function for each estimated parameter. In each direction, the derivative is calculated by perturbing each parameter and then computing the likelihood.

of the benchmark model to Appendix C.

3.3.1 Mapping Unobserved Skills to Observed Misbehaviors

Starting with the joint distribution of unobserved skills, we find a positive correlation between externalizing and internalizing behavior along with a negative correlation between the two socio-emotional skills and cognition. These correlations could reflect the distribution of skill endowments at birth or early childhood investments if the same environments that promote externalizing and internalizing behaviors also slow cognitive development (Heckman and Cunha, 2007).[19] Accounting for correlation across factors means that we avoid mis-attributing returns to skills. For example, failing to account for the positive association between externalizing and internalizing behavior could lead us to over-estimate the degree to which each socio-emotional skill negatively affects schooling.

In Table 3, we report estimates of factor loadings mapping latent skills to BSAG maladjustment syndromes and aptitude test scores. Estimates are reported separately by gender. Consistent with the interpretation of the two socio-emotional skills discussed before, externalizing behavior loads heavily onto disruptive and impulsive syndromes such as hostility towards adults, anxiety towards children or adults, inconsequential behavior and restless behaviors, while internalizing behavior loads heavily onto inhibited syndromes such as withdrawal, unforthcomingness and writing off adults and standards. Cognition loads mostly onto the tests scores. These results are also broadly in line with how we grouped the measurements as reflecting the three skills in the preliminary analysis in Section 2. Across genders, there are some differences in the factor loadings, but they are generally small and insignificant.

Most of the coefficients on the variables related to school characteristics have the expected signs. A higher percentage of students in the school taking GCE O-Levels qualification exams is negatively associated with misbehaviors and positively associated with test scores. Being in a public school (i.e., managed by LEA) tends to reduce girls' test performances, but not boys'. The number of teachers is an indicator of the size of the school, with bigger schools associated with lower test scores. A larger class size tends to reduce measurements closely related to externalizing behavior and increase test scores, for both boys and girls, which is in contrast with previous research (Fredriksson, Öckert, and Oosterbeek, 2012).[20]

[19]An example would be childhood poverty, which we investigate in Section 4.2. The positive correlation between externalizing and internalizing behavior is well-documented in the child development literature. Children under stress as a result of poverty or a family disruption tend to develop both aggressive and depressive symptoms (Wolfson, Fields, and Rose, 1987).

[20]It is possible that class size captures omitted school-level variables which positively affect student out-

3.3.2 Education

The marginal effect estimates of the multinomial logit model for educational attainment are reported in Table 4. There is a significant negative relationship between externalizing behavior and educational attainment for boys in the sense that externalizing significantly increases the chance of having no or low educational qualifications. A difference from the estimates in the preliminary analysis is that the negative relationship between externalizing and schooling for females is no longer present. The marginal effects are small and the sign of the relationship is unclear. In other words, high-externalizing females are better able to finish school in comparison to high-externalizing males. This finding is generally consistent with earlier literature showing that the negative impact of externalizing behavior on schooling is more salient for boys than for girls (Bertrand and Pan, 2013). Indeed, teachers are more likely to punish male versus female children for the same level of aggression (Gregory, 1977). On the other hand, we find that internalizing behavior is negatively associated with educational attainment for females, but less strongly so for males. This is also in line with research that finds stronger effects of conduct disorders and weaker effects of anxiety and depressive symptoms for the educational attainment of males in comparison to females (Kessler et al., 1995).

Effect sizes for socio-emotional skills in the schooling model are much smaller than those for cognition, which predicts schooling at similar magnitudes across genders, but are of similar magnitudes as the effects of family characteristics, which all have the expected signs. Having parents with more education and who work in more lucrative occupational categories is related to higher educational attainment for the child. Moreover, individuals living in poverty during their childhood, suggesting relatively few family resources available to invest in children, are less likely to attain higher levels of education.

Estimates for the schooling model are broadly consistent with literature that studies the impact of emotional problems in school. A possible interpretation is that high externalizing boys face relatively low returns to higher levels of education which would disincentivize education if students are aware of this difference. This type of pattern is discussed in Prada and Urzúa (2017) to explain the negative relationship between mechanical skill and education. However, we find no evidence that externalizing behavior leads to heterogeneous returns to education when exploring potential mechanisms in Appendix D. The alternative that is consistent with our estimates is that externalizing children face higher costs, such as effort costs, to finish school. There is a basis for this interpretation in earlier literature. McLeod and

comes, such as teacher quality if better teachers are assigned to larger classes. This type of bias would be more concerning if these variables were the focus of our analyses rather than controls to address potential mis-reporting.

Kaiser (2004) argue that children with internalizing and externalizing behaviors withdraw from social relationships in school, including those with teachers, in order to minimize their exposure to negative interactions.

3.3.3 Labor Market Outcomes

Literature studying the consequences of externalizing behavior has generally limited attention to educational attainment. In contrast, we assess the relationship between childhood misbehavior and labor market outcomes. Estimates of hours and wage equations conditional on employment are reported in Table 5. The benchmark model results, where we control for educational attainment, are presented in Column (2). For males, a one-standard-deviation increase in externalizing behavior predicts a statistically significant 6.4% increase in hourly wages, but does not significantly affect weekly hours worked. For females, a one-standard-deviation increase in externalizing behavior predicts a marginally significant 4.7% increase in hours worked per week, but does not significantly affect hourly wages.

One explanation for this finding relates to our interpretation of the negative impact of externalizing on schooling. To compensate for difficulties attaining high levels of education, such as high effort costs, high-externalizing students could be positively selected on other dimensions that lead to higher earnings. This dynamic would pertain to males, who face the schooling penalty. A leading contender would be positive selection in the form of higher cognition, which leads to higher earnings. Since our model explicitly controls for cognition, this does not drive our results. As part of our sensitivity analyses (summarized in Section 3.4), we allow for an additional (fourth) unobserved factor, which would capture additional sources of positive selection. Results remain largely unchanged.

We also investigate whether the "net effect" of externalizing behavior is lucrative. Our benchmark model includes educational attainment in the labor market outcome equations, which captures the negative impact of externalizing on earnings through lower educational attainment. If we omit education, the estimated coefficient mapping externalizing to hours and wages includes both the direct effect of externalizing on earnings and the indirect effect working through schooling, the net of which could be negative. Net effects are reported in Column (1). Compared to the benchmark model (Column (2)), omitting education increases the point estimates of the effect of cognition on hourly wages for both males and females, but it also reduces the point estimate of the effect of externalizing on hourly wages for males, though only slightly. Notably, the coefficient is still positive whether or not we include schooling, suggesting that more externalizing males earn higher wages despite the negative impact of externalizing on schooling. The story for females is more complex, as externalizing influences marriage and fertility decisions. We explore these pathways in Section 3.3.4.

Internalizing behavior is negatively related to both wages and hours worked. For males, a one-standard-deviation increase in internalizing behavior predicts a very significant 9.6% decrease in hourly wage and a marginally significant 1.8% decrease in weekly hours worked. For females, the counterpart coefficients in both the wage and hours worked equations are negative, but neither is significant. We also find that cognition significantly increases hourly wages (by 2.5% for males and 4.4% for females), but does not influence the hours decision for either gender. The remaining parameters follow conventional wisdom. For example, higher educational attainment increases worker productivity, but has little effect on the number of hours worked for those already employed. Also, individuals living in or around London earn significantly higher hourly wages, while individuals who experience financial difficulties in childhood receive lower hourly wages.

Our findings demonstrate a more nuanced relationship between childhood misbehavior and labor market outcomes than has been recognized in previous literature. They also illustrate how socio-emotional skills can have mixed effects on economic outcomes.

3.3.4 Additional Lifecycle Outcomes

Differences in effects of externalizing behavior on hourly wages and weekly hours worked for males and females suggest that externalizing might work through different channels across genders. To further probe our findings, we examine two additional lifecycle outcomes that are especially relevant for the age group under study: marriage and fertility. For example, it is possible that high-externalizing individuals are less likely to be in relationships or to have children, which could free up time to work longer hours or to focus on working more productively.

We assess how estimated coefficients change when we add endogenous intermediate outcome variables to the wage and hours equations, while keeping the measurement system mapping latent skills to observed measurements of misbehavior as in the benchmark model. We find that while having a partner has a strong positive effect on wages for both males and females, having children lowers wages and weekly hours worked for females only. Controlling for partnership and fertility does not change the coefficients on externalizing in any significant way for males, but it roughly doubles the point estimates of the impact of externalizing on wages and hours worked as well as increases their statistical significance for females.

To understand the gender difference in how fertility affects the externalizing earnings premium, we estimate a linear regression of the number of children by age 33 on the three skills from the previously estimated measurement system and find that externalizing males and females are both more likely to have a larger number of children by age 33. However since fertility only lowers wages and hours for females, this channel operates to counteract the direct positive effects of externalizing in the labor market for females but not males. It is noteworthy that these patterns remain after we control for further outcomes such as months of experience and occupations.

To visualize these results, we plot earnings against different levels of externalizing separately for men and women in Figure 1. The slope of the curve represents the impact of externalizing behavior on earnings. To generate the figure, we simulate weekly earnings, which is the product of hourly wages and weekly hours worked, as we vary the externalizing behavior from the 5th percentile to the 95th percentile, keeping other latent skills and covariates at the population median. We repeat this exercise conditioning on different sets of intermediate outcome variables. For males, conditioning on intermediate outcomes does not change the slope very much (Panel (a)). For females, the slope increases noticeably when we condition on the number of children by age 33 (Panel (b)), reflecting the positive relationship between externalizing and fertility along with the negative relationship for females between number of children and earnings. An interpretation of this result is that there are large labor market returns to high-externalizing women who do not have children.

3.4 Sensitivity and Robustness

The positive impact of externalizing behavior on earnings and its mixed effects on schooling versus earnings for males are novel findings that have not been explored in earlier literature. We constructed our main econometric model to account for obvious sources of selection. Nevertheless, securing our estimates requires a number of assumptions and pertains to a particular cohort in one time and place. Our next step is to assess whether findings are robust to alternative modeling assumptions or when we use alternative samples from other countries or time periods. In this section, we summarize our findings from the various sensitivity and robustness analyses we painstakingly conducted.

3.4.1 Sensitivity

The benchmark model requires a number of assumptions. At each point where we made a choice and there are reasonable alternative assumptions, we examined them, leading us to estimate dozens of alternative models. We report our full investigation in Appendix C. More specifically, we considered: alternative functional form assumptions for the schooling outcome equation (Appendix C.2.1); alternative sets of controls in the labor market outcomes equations in the benchmark model, including the exclusion of all controls (Appendix

C.2.2); the addition of a fourth latent factor in the measurement system of the benchmark model (Appendix C.2.3); independence across the three factors (Appendix C.2.4); alternative dedicated measurements (Appendix C.2.5); and joint estimation of the parameters in the measurement system and the outcome equations (Appendix C.2.6). The main result that externalizing reduces educational attainment for boys but raises earnings for boys and girls is remarkably robust to all these variations in modeling assumptions and estimation procedures. For brevity, we have relegated details of each analysis to appendices. However, we elaborate here on one sensitivity analysis in which we vary assumptions on dedicated measurements. The assumption of dedicated measurements is key to the identification of the model and the interpretation of the results.

In the benchmark model, we assume there are three unobservable skills, externalizing behavior, internalizing behavior and cognition, which are identified from measures of childhood classroom misbehavior and test scores. Identification of measurement system parameters requires that, for each skill, we designate one particular measurement (the dedicated measurement) that is not a measurement of the two other skills. Which measurement to dedicate to each skill is ultimately a choice. A benefit of our approach is that we can re-estimate the model iterating over all possible candidates for the dedicated measurements of the two socio-emotional skills to assess sensitivity.

We plot the effect on weekly earnings from a one-standard-deviation increase in externalizing behavior for each different choice of dedicated measurements in Figure 2. The dashed bars indicate the results from the benchmark econometric model. While different dedicated measurement choices imply different magnitudes of the effects on earnings, in the majority of cases externalizing behavior has a significantly positive earnings premium for both genders, and the benchmark specification does not deliver the largest point estimate for either gender. In no specification do we find significant evidence against our main result. The earnings premium loses significance when we choose withdrawal or unforthcomingness to be the dedicated measure for internalizing behavior and depression loads heavily on the "externalizing" factor. In this case, we identify an "externalizing" factor that is a mixture of what we typically regard as outwardly expressed externalizing behavior and inwardly expressed internalizing behavior, and the impact of this factor on earnings is muted given the negative correlation between depression and productivity. Alternatively, we can construct an externalizing factor that does not map to depression, loads heavily onto outwardly expressed aggressive behaviors, and which has a positive impact on earnings. The benchmark model imposes the latter assumption.

Broadly, this analysis illustrates the fundamental identification problem in measuring underlying traits as discussed in Almlund et al. (2011). The researcher faces a trade-off between letting the data guide the analysis versus imposing just enough structure to identify economically meaningful objects.

3.4.2 Generalizability Across Data Sets

A possible concern is that our findings are specific to Great Britain in the 1950s. We thus explore other data sets in more contemporary settings and in different social contexts. We replicate our main analysis in the 1970 British Cohort Study, the National Education Longitudinal Study of 1988, the Panel Study of Income Dynamics, and the National Longitudinal Survey of Youth 1979 Children and Young Adults. The latter three are U.S. data sets. These are the major longitudinal studies that follow individuals over the lifecycle with measurements of both behavior during childhood in school and labor market outcomes for the same individuals. Detailed descriptions of the data sets and the full empirical results are found in Appendix C.3.

In each dataset, we construct crude measures of skills and link these to schooling and earnings. Results are summarized in Table 6. We show that, in all data sets, externalizing behavior is associated with fewer years of schooling. This negative effect is strongly significant, with the exception of the PSID where the negative coefficient is significant at the 10% level. Compared to the NCDS sample, the point estimates of the correlation between externalizing behavior and years of schooling in the samples of younger cohorts tend to be bigger, suggesting an externalizing penalty in school that persists across cohorts. We also show that externalizing behavior is significantly associated with higher earnings in the two British data sets, the 1958 and the 1970 cohort, and two U.S. data sets, NELS and PSID.[21] The point estimate of the impact of externalizing on earnings from the NCDS lies between estimates obtained from other data sets. These results suggest that the documented externalizing earnings premium does not vary systematically across countries or over time.

[21] The CNLSY is the only dataset where we do not find a significant relationship between externalizing behavior and earnings. This can be due to two reasons. First, the CNLSY is the only dataset where we rely on parents' report of children's behaviors and previous research has highlighted important differences in parental and teachers' reports of children's behavior (Achenbach, McConaughy, and Howell, 1987; De Los Reyes and Kazdin, 2005), and evidence of bias in maternal reports (Boyle and Pickles, 1997; Najman et al., 2000). Second, the CNLSY sample with observed earnings is a selected sample born from young mothers. It is thus possible that our findings using the CNLSY arise from sample selection towards children born into poorer households, which aligns with the lack of evidence of an externalizing premium among low-SES families from the NCDS (see Section 4.2).

4 Additional Analyses and Interpretation of Results

We have demonstrated a robust empirical regularity that a well-studied socio-emotional skill, externalizing behavior, lowers educational attainment for males and raises earnings for males and females. These patterns, including the sign reversal for males, are robust to alternative modeling assumptions and hold across datasets from different places and times.

In this section, we provide additional analyses to further investigate this empirical pattern. The goal is to gain a better understanding of specific behaviors or choices that can explain these patterns. Understanding these patterns should be a part of any policy discussions since what drives them could be socially beneficial or costly. While we can rule out some possible hypotheses, evidence on specific channels is inconclusive. More specifically, we consider whether occupational sorting (Section 4.1), childhood economic disadvantage (Section 4.2) and other personality traits or stress hormones (Section 4.3) could drive our main findings. For brevity, the complete empirical results from these three sub sections are found in Appendix D.

4.1 Occupational Sorting

Some dimensions of human capital are more productive for the completion of certain tasks, as different tasks in life require different skills in different degrees (see, e.g., Roy, 1951; Mandelbrot, 1962; Willis and Rosen, 1979; Heckman and Sedlacek, 1985; Heckman, Stixrud, and Urzúa, 2006). Building on this idea, a reasonable hypothesis is that externalizing behavior is productive for a limited set of tasks and is thus lucrative in a subset of possible occupations. This would have policy implications. For example, if results are driven by a very small number of tasks, it might be that low-externalizing people could be trained on just these tasks at relatively low cost (or could avoid occupations that require such tasks) and that the general thrust of policies aiming to reduce externalizing behavior are not necessarily the wrong approach.

We extend our labor market model to allow for the returns to skills to vary with occupational tasks using the O*NET[22] task-intensity scales as in Acemoglu and Autor (2011). We focus on two well-studied measures: the abstract/social task intensity and the routine/manual task intensity. The task intensities are composite measures of O*NET Work Activities and Work Context Importance scales.[23] The two composite scales were constructed

[22]The O*NET is an American classification system, and the NCDS collected detailed information on individual occupations in the ISCO-88 classification system. We rely on the methodology in Hardy, Keister, and Lewandowski (2018) to link the NCDS individuals' occupations to the O*NET classification.

[23] The abstract/social task measure is a normalized composite scale of six O*NET subscales: "analyzing

using factor analysis and are standardized to have mean zero and variance equal to one. The results are found in Table 7.

While we find some heterogeneity in the returns to externalizing behavior across tasks, the externalizing behavior labor market premium is predominantly positive. Mainly, for males, we find that the returns to externalizing behavior are smaller in occupations that are intensive in abstract and social tasks and larger in occupations that are intensive in manual and routine tasks. This heterogeneity is, however, very small. Since the task intensities in Table 7 are standardized, the estimates suggest that only for jobs with routine tasks below the 2.5th percentile (2 standard deviations below the mean) or with abstract tasks in the 97.5th percentile (2 standard deviations above the mean) or both (as measured in the NCDS) would we expect to find an overall negative return to externalizing in the labor market. Roughly 5% of individuals in our sample have an occupation meeting one of these criteria. Individuals in our sample in occupations requiring sufficiently high levels of abstract tasks to meet this threshold include senior government officials and managers of personnel departments. Those requiring sufficiently low levels of manual tasks include fashion models.

This points to the generality of our main message to the vast majority of working adults. This is in contrast to the finding in Levine and Rubinstein (2017), who show that the combination of being smart and illicit during youth appears to be productive in a very small sector of the economy, incorporated entrepreneurship. One could argue that there is no reason to foster or accommodate illicit behavior since it only seems to benefit a small sliver of the labor market. Our findings on externalizing behavior suggest a much starker social trade-off.

4.2 Childhood Disadvantage

Studying a sample of disadvantaged Black children in the U.S., Heckman, Pinto, and Savelyev (2013) find that an early childhood education program increased earnings in part by reducing externalizing behavior. In contrast, we show that externalizing behavior can be valuable in the labor market. In this section, we explore whether differences in findings are explained by differences in the socioeconomic status of the group being analyzed and evaluate the possibility that children born into poorer families face a higher likelihood of criminality or police involvement for the same level of externalizing behavior.

data/information," "thinking creatively," "interpreting information for others," "establishing and maintaining personal relationships," and "guiding, directing, and motivating subordinates and coaching and developing others." The routine/manual task measure is a normalized composite scale of six O*NET subscales: "importance of repeating the same tasks," "importance of being exact or accurate," "structured versus unstructured work," "controlling machines and processes," "keeping a pace set by machinery or equipment," and "time spent making repetitive motions."

We construct a subsample of our analytic sample consisting of subjects who faced financial difficulties during childhood to resemble the family characteristics of the sample studied in Heckman, Pinto, and Savelyev (2013), which we refer to as the "Low SES" subsample.[24] We then expand the benchmark econometric model by including a measure of police involvement at age 16 as an additional outcome equation and as an additional explanatory variable in the schooling, wage and hours equations. We estimate this expanded model for the "Low SES" subsample and for all other subjects in our analytic sample, which we call the "High SES" subsample, separately.[25]

Estimates from this expanded model show patterns that are similar to the main results. However, we find suggestive evidence of differences by childhood SES. First, we estimate a larger schooling penalty for externalizing behavior among individuals that grew up in low-SES households. This finding is broadly consistent with results in Ramey (2018), who shows that high-externalizing Black students in the U.S. face a higher likelihood of punishment by suspension in comparison to similarly externalizing whites. This could arise because schools that serve low-SES children have fewer resources to address externalizing behavior and therefore react to it through suspensions or expulsions. Perhaps most importantly, we find suggestive evidence that the labor market returns to externalizing behavior may not extend to the "Low SES" subsample (Table 8). First, individuals that grew up with financial difficulties experience less than a third of the externalizing wage premium compared to individuals that did not, while wage returns to other skills are similar across the two groups. Second, externalizing behavior has insignificant if not negative effects on hours worked for the low-SES group, while for the high-SES group its effect is significantly positive. An important caveat to these results is that we cannot statistically differentiate the returns to externalizing behavior for the two socioeconomic groups because the standard errors in the estimates for the "Low SES" sample are large, possibly due to the relatively small sample size of the low-SES group.

We examine whether, consistent with the story in Heckman, Pinto, and Savelyev (2013), criminality explains the differential labor market returns to externalizing by group. It does not. While externalizing behavior predicts higher police involvement, police involvement does not appear to derail labor market prospects among individuals in the British sample we

[24] An individual is coded as experiencing financial difficulties during childhood if the interviewer reported that the household appeared to be experiencing poverty in 1965 or if a member of the household self-reported having financial difficulties in the 12 months prior to being observed in either 1969 or 1974. Summary statistics for the sub-samples and estimation results are found in Appendix D.

[25] Note that we estimate the model including the measurement system separately by group, since it is possible that underlying skills map to observed behaviors differently by group. Similarly, to study black-white differentials in labor market outcomes in the U.S., Urzúa (2008) allows the distribution and impact of underlying skills to vary by race.

study. These results raise the possibility that the returns to externalizing behavior might be negative in a context where police involvement is highly penalized in the labor market such as that studied in Heckman, Pinto, and Savelyev (2013), but not necessarily so in a different context such as in the British sample we study.

If aggressive behavior is deemed unacceptable in jobs limited to advantaged people with social connections, we might expect the externalizing premium to be larger for people from less advantaged backgrounds. We find that returns are, if anything, larger for more advantaged groups. This analysis helps to reconcile results with those in Heckman, Pinto, and Savelyev (2013). However, it raises additional questions about skill prices across socioeconomic groups (Lundberg, 2013). We are therefore left with at least two distinct, but related, possibilities. The first is that there are true differences in the productivity of externalizing behavior across groups. For example, children born into wealthier families may be better able to channel their hyperactivity into productive activities. Instead, it appears that the same skill is rewarded differently for individuals from different groups. For example, managers or co-workers may view high-externalizing individuals from high-SES families as ambitious leaders and be willing to hire them in high-wage positions or to promote them. In contrast, high-externalizing individuals from lower SES families may find their advancement thwarted if they are viewed as disruptive, aggressive or impolite. Seen another way, these results suggest the concerning possibility that children from poorer families are unable to unleash the potential of skills that are valuable and lucrative for children born into wealthier families.

4.3 Personality Traits and Stress Hormones

Given that the externalizing premium in the labor market appears to be widespread across different occupations, countries and cohorts, we investigate if this is a genuinely new result in the literature which emerges after we disaggregate childhood misbehavior into different underlying factors, or is simply a re-labeling of a well-studied personality trait which we already know about.

Several studies have examined the relationship between externalizing and internalizing behaviors and the "Big 5" personality traits (openness, conscientiousness, extraversion, agreeableness and neuroticism). Evidence suggests that externalizing behavior is negatively associated with conscientiousness, agreeableness, and openness to new experience, while internalizing behavior is mostly related to neuroticism (Ehrler, Evans, and McGhee, 1999; Almlund et al., 2011). Moreover, agreeableness predicts lower earnings (Judge, Livingston, and Hurst, 2012).[26] Could it be the externalizing individuals are simply less agreeable peo-

[26]To explain why, Barry and Friedman (1998) show that individuals with higher levels of agreeableness

ple? We test for this possibility by adding the "Big 5" traits to the descriptive earnings equation in Section 2.3 as controls. Controlling for the "Big 5" traits reduces the effect of externalizing behavior on earnings by about 20% and increases the negative effect on education by about 15%. However, our main findings remain after we control for the "Big 5" personality traits, suggesting that, despite correlations, the skills we study are distinct factors with independent impacts on economic outcomes.[27]

Similarly, our main findings could be explained by externalizing individuals having higher levels of stress hormones (cortisol), which recent research has shown to relate to risk-taking behaviors that could potentially make schooling difficult but be productive at work (Shirtcliff et al., 2005). Again, we do not find evidence that this is the case. While externalizing is positively correlated with cortisol levels later in life, controlling for cortisol does not change significantly the relationship between externalizing behaviors and schooling and earnings in our sample. In general, our results are not explained by correlations between externalizing behavior and well-known constructs that have been examined in earlier literature.

5 Conclusion

Few would disagree that better health and stronger cognition improve outcomes on most any conceivable dimension. Other components of human capital, such as mechanical skill, may not be especially productive in certain contexts, but there is little evidence suggesting they reduce productivity and instead may simply remain idle. There has been little focus on the idea that some skills could be both helpful and harmful, depending on the context.

We examine the lifecycle impacts of a widely studied socio-emotional skill, externalizing behavior, which underlies childhood misbehavior in school. This skill has been the focus of hundreds of papers and its negative impact on educational attainment has been used to justify multitudes of policies to curb or discourage it. Aligned to prior literature, we find that externalizing behavior lowers educational attainment for boys. However, we also find that

are worse negotiators as they are susceptible to being anchored by early offers in the negotiation process. Relatedly, Spurk and Abele (2011) show that less agreeable individuals are more competitive in the workplace and place a higher emphasis on career advancement.

[27] One important caveat to our results on personality using the NCDS is that the "Big 5" personality traits are measured at age 50, after educational and labor market outcomes are realized. Thus, estimates could be biased due to simultaneity, if labor market shocks influence how individuals respond to the personality questions. We therefore address the question of adjusting for additional unobserved skills using the British Cohort Study (BCS), which we describe in more detail in Section 3.4.2. Using the BCS, we construct socio-emotional skills from a larger set of behavioral questions. The larger number of measurements allows us to identify as many as 8 distinct factors, three of them capturing externalizing behavior, internalizing behavior and cognition. We find that the key patterns described in our benchmark model still hold when we identify externalizing behavior using this larger set of measurements, and also when we include additional factors capturing additional socio-emotional skills in schooling and labor outcome equations.

externalizing behavior increases earnings for men and women. These empirical patterns are not explained by selection into marriage or fertility. Moreover, the externalizing premium is not limited to specific tasks or occupations; there are virtually no occupations in which externalizing behavior is unproductive. Our findings are robust to a host of alternative modeling assumptions and hold across data sets from different times and places.

Our finding challenges a large literature on child development that implicitly assumes that the skills, characteristics or traits that are good for schoolchildren are good overall. Moreover, while an active literature has expanded the set of factors we now understand to constitute human capital, the idea that some of these factors could be both helpful and harmful depending on the context is novel. This is particularly striking in the case of externalizing behavior since opposite signs for males occur across two key phases of the lifecycle: schooling and work, both of which are unavoidable for most individuals. Our findings also underline the importance of taking account of contexts when measuring the returns to skills. Doing so in a narrow setting can be misleading and inappropriately extrapolated.

More broadly, the mixed impacts of externalizing behavior we estimate for males suggest models of human capital accumulation should be extended to accommodate sign reversals in skill prices. Doing so would suggest we revisit policies aiming to develop socio-emotional skills among schoolchildren. The concern is that well-meaning interventions might promote education, but prove harmful in the long run. Alternatively, policies previously deemed undesirable because they encouraged externalizing behavior among schoolchildren may have overlooked benefits. For example, Baker, Gruber, and Milligan (2008) examine the negative consequences of universal daycare access, among which they count an increase in aggressive behaviors. As the children in their sample enter the labor market, future research could examine whether this was indeed a downside or, alternatively, whether there are long-term benefits to the skills they accumulated due to the policy change.

Our investigation into explanations and channels allows us to rule out some hypotheses, but is ultimately inconclusive. We do not yet understand the specific behaviors, actions or choices that make externalizing behavior a valuable skill across occupations and tasks in the labor market. If they are socially beneficial, perhaps externalizing behavior should be promoted. If not, perhaps it should be curbed. Future research could fill this gap. In particular, a useful path forward would be to measure differences in how high- and low-externalizing individuals act in the workplace, including the choices they make, their interactions with colleagues and their work decisions. Only then can we begin to understand the price reversal we document and consider how it could inform policy.

References

Acemoglu, Daron and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, vol. 4. Elsevier, 1043–1171.

Achenbach, Thomas M. 1978. "The Child Behavior Profile: I. Boys Aged 6-11." *Journal of Consulting and Clinical Psychology* 46:478–488.

Achenbach, Thomas M., Stephanie H. McConaughy, and Catherine T. Howell. 1987. "Child/Adolescent Behavioral and Emotional Problems: Implications of Cross-Information Correlations for Situational Specificity." *Psychological Bulletin* 101 (2):213–232.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." *Handbook of the Economics of Education* 4 (1).

Autor, David H and Michael J Handel. 2013. "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages." *Journal of Labor Economics* 31 (S1):S59–S96.

Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2008. "Universal Child Care, Maternal Labor Supply, and Family Well-Being." *Journal of Political Economy* 116 (4):709–745.

Barry, Bruce and Raymond A Friedman. 1998. "Bargainer Characteristics in Distributive and Integrative Negotiation." *Journal of Personality and Social Psychology* 74 (2):345.

Bertrand, Marianne and Jessica Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5 (1):32–64.

Borghans, Lex, Angela Lee Duckworth, James J Heckman, and Bas Ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43 (4):972–1059.

Boyle, Michael H and Andrew R Pickles. 1997. "Influence of Maternal Depressive Symptoms on Ratings of Childhood Behavior." *Journal of Abnormal Child Psychology* 25 (5):399–412.

Campbell, Amy and Cynthia M. Anderson. 2011. "Check-In/Check-Out: A Systematic Evaluation and Component Analysis." *Journal of Applied Behavior Analysis* 44 (2):315–326.

Campbell, Susan B, Daniel S Shaw, and Miles Gilliom. 2000. "Early Externalizing Behavior Problems: Toddlers and Preschoolers at Risk for Later Maladjustment." *Development and Psychopathology* 12 (3):467–488.

CPPRG. 1992. "A Developmental and Clinical Model for the Prevention of Conduct Disorders: The FAST Track Program." *Development and Psychopathology* 4:505–527.

Crone, Deanne A., Leanne S. Hawken, and Robert H. Horner. 2003. *Responding to Problem Behavior in Schools: The Behavior Education Program. Second Edition.* Guilford Press.

Cunha, F. and J. J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Non-Cognitive Skill Formation." *Journal of Human Resources* 43:738–782.

Cunha, Flávio, Irma Elo, and Jennifer Culhane. 2020. "Maternal Subjective Expectations about the Technology of Skill Formation Predict Investments in Children One Year Later." *Journal of Econometrics* .

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3):883–931.

De Los Reyes, Andres and Alan E Kazdin. 2005. "Informant Discrepancies in the Assessment of Childhood Psychopathology: a Critical Review, Theoretical Framework, and Recommendations for Further Study." *Psychological Bulletin* 131 (4):483.

Doyle, Orla. 2020. "The First 2,00 Days and Child Skills." *Journal of Political Economy* 128 (6):2067–2122.

Duncan, Greg J and Rachel Dunifon. 2012. "'Soft-Skills' and Long-Run Labor Market Success." In *5th Anniversary Retrospective (Research in Labor Economics, Volume 35)*, edited by Solomon W. Polachek and Konstantinos Tatsiramos. Emerald Group Publishing Limited, 313–339.

Duncan, Greg J and Katherine Magnuson. 2011. "The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. Russell Sage Foundation, 47–69.

Ehrler, David J, J Gary Evans, and Ron L McGhee. 1999. "Extending Big-Five Theory into Childhood: A Preliminary Investigation into the Relationship Between Big-Five Personality Traits and Behavior Problems in Children." *Psychology in the Schools* 36 (6):451–458.

Eisenberg, Nancy, Amanda Cumberland, Tracy L. Spinrad, Richard A. Fabes, Stephanie A. Shepard, Mark Reiser, Bridget C. Murphy, Sandra H. Losoya, and Ivanna K. Guthrie. 2001. "The Relations of Regulation and Emotionality to Children's Externalizing and Internalizing Problem Behavior." *Child Development* 72 (4):1112–1134.

Farmer, Elizabeth MZ. 1995. "Extremity of Externalizing Behavior and Young Adult Outcomes." *Journal of Child Psychology and Psychiatry* 36 (4):617–632.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2012. "Long-term Effects of Class Size." *Quarterly Journal of Economics* 128 (1):249–285.

Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M Chang, and Sally Grantham-McGregor. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344 (6187):998–1001.

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.

Gregory, Mary K. 1977. "Sex Bias in School Referrals." *Journal of School Psychology* 15 (1):5–8.

Grossman, Michael. 2000. "The Human Capital Model of the Demand for Health." In *Handbook of Health Economics*, vol. 1A. Elsevier, 347–408.

Hardy, Wojciech, Roma Keister, and Piotr Lewandowski. 2018. "Educational Upgrading, Structural Change and the Task Composition of Jobs in Europe." *Economics of Transition* 26 (2):201–231.

Heckman, James J and Flavio Cunha. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2):31–47.

Heckman, James J and Tim Kautz. 2014. "Fostering and Measuring Skills: Interventions that Improve Character and Cognition." In *The Myth of Achievement Tests*, edited by James J. Heckman, John Eric Humphries, and Tim Kautz. University of Chicago Press, 341–430.

Heckman, James J, 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 and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review* 91 (2):145–149.

Heckman, James J and Guilherme Sedlacek. 1985. "Heterogeneity, aggregation, and Market Wage Functions: an Empirical Model of Self-selection in the Labor Market." *Journal of Political Economy* 93 (6):1077–1125.

Heckman, James J, Jora Stixrud, and Sergio Urzúa. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *The Journal of Labor Economics* 24 (3):411–482.

Humphrey, Neil, Afroditi Kalambouka, Michael Wigelsworth, Ann Lendrum, Clare Lennie, and Peter Farrell. 2010. "New Beginnings: Evaluation of a Short Social–Emotional Intervention for Primary-Aged Children." *Educational Psychology* 30 (5):513–532.

Judge, Timothy A, Beth A Livingston, and Charlice Hurst. 2012. "Do Nice Guys and Gals Really Finish Last? The Joint Effects of Sex and Agreeableness on Income." *Journal of Personality and Social Psychology* 102 (2):390–407.

Kendler, Kenneth S. and John Myers. 2014. "The Boundaries of the Internalizing and Externalizing Genetic Spectra in Men and Women." *Pschological Medicine* 44 (3):647–655.

Kessler, Ronald C, Cindy L Foster, William B Saunders, and Paul E Stang. 1995. "Social Consequences of Psychiatric Disorders, I: Educational Attainment." *American Journal of Psychiatry* 152 (7):1026–1032.

Levine, Ross and Yona Rubinstein. 2017. "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?" *Quarterly Journal of Economics* 132 (2):963–1018.

Lundberg, Shelly. 2013. "The College Type: Personality and Educational Inequality." *Journal of Labor Economics* 31 (3):421–441.

Lundberg, Shelly and Robert A Pollak. 2007. "The American Family and Family Economics." *Journal of Economic Perspectives* 21 (2):3–26.

Malti, Tina and Kenneth H. Rubin. 2018. *Handbook of Child and Adolescent Aggression.* Guilford Press.

Mandelbrot, Benoit. 1962. "Paretian Distributions and Income Maximization." *The Quarterly Journal of Economics* 76 (1):57–85.

McLeod, Jane D and Karen Kaiser. 2004. "Childhood Emotional and Behavioral Problems and Educational Attainment." *American Sociological Review* 69 (5):636–658.

Najman, Jake M, Gail M Williams, Jane Nikles, SUE Spence, William Bor, Michael O'callaghan, Robyne Le Brocque, and Margaret J Andersen. 2000. "Mothers' Mental Illness and Child Behavior Problems: Cause-Effect Association or Observation Bias?" *Journal of the American Academy of Child & Adolescent Psychiatry* 39 (5):592–602.

Nangle, Douglas W., Cynthia A. Erdley, and Rebecca A. Schwartz-Mette. 2020. *Social Skills Across the Life Span: Theory, Assessment, and Intervention.* Academic Press.

Närhi, Vesa, Tiina Kiiski, and Hannu Savolainen. 2017. "Reducing Disruptive Behaviours and Improving Classroom Behavioural Climate with Class-Wide Positive Behaviour Support in Middle Schools." *British Educational Research Journal* 43 (6):1186–1205.

Neidell, Matthew and Jane Waldfogel. 2010. "Cognitive and Noncognitive Peer Effects in Early Education." *The Review of Economics and Statistics* 92 (3):562–576.

O'Connor, Kim M. and Ben Hayes. 2020. "How Effective are Targeted Interventions for Externalizing Behavior When Delivered in Primary Schools?" *International Journal of School and Educational Psychology* 8 (3):161–173.

Prada, María F and Sergio Urzúa. 2017. "One Size Does Not Fit All: Multiple Dimensions of Ability, College Attendance, and Rarnings." *Journal of Labor Economics* 35 (4):953–991.

Ramey, David M. 2018. "The Social Construction of Child Social Control via Criminalization or Medicalization: Why Race Matters." *Sociological Forum* 33 (1):139–164.

Regier, Darrel A., Emily A. Kuhl, and David J. Kupfer. 2013. "The DSM-5: Classification and Criteria Changes." *World Psychiatry* 12 (2):92–98.

Roy, Andrew Donald. 1951. “Some Thoughts on the Distribution of Earnings.” *Oxford Economic Papers* 3 (2):135–146.

Segal, Carmit. 2013. “Misbehavior, Education, and Labor Market Outcomes.” *Journal of the European Economic Association* 11 (4):743–779.

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

Shirtcliff, Elizabeth A, Douglas A Granger, Alan Booth, and David Johnson. 2005. “Low salivary cortisol levels and externalizing behavior problems in youth.” *Development and Psychopathology* 17 (1):167–184.

Spurk, Daniel and Andrea E Abele. 2011. “Who Earns More and Why? A Multiple Mediation Model From Personality to Salary.” *Journal of Business and Psychology* 26 (1):87–103.

Stage, Scott A. and David R. Quiroz. 1997. “A Meta-Analysis of Interventions to Decrease Disruptive Classroom Behavior in Public Education Settings.” *School Psychology Review* 26 (3):333–368.

Stott, Denis Herbert, Emily G Sykes, and NC Marston. 1974. *The Social Adjustment of Children: Manual to the Bristol Social Adjustment Guides.* Hodder & Stoughton.

Sørlie, Mari-Anne and Terje Ogden. 2015. “School-Wide Positive Behavior Support – Norway: Impacts on Problem Behavior and Classroom Climate.” *International Journal of School and Educational Psychology* 3 (3):202–217.

Urzúa, Sergio. 2008. “Racial Labor Market Gaps: The Role of Abilities and Schooling Choices.” *Journal of Human Resources* 43 (4):919–971.

Williams, Benjamin. 2020. “Identification of the Linear Factor Model.” *Econometric Reviews* 39 (1):92–109.

Willis, Robert J. and Sherwin Rosen. 1979. “Education and Self-Selection.” *Journal of Political Economy* 87 (5):S7–S36.

Wolfson, Johanna, Joel H Fields, and Susan A Rose. 1987. “Symptoms, Temperament, Resiliency, and Control in Anxiety-Disordered Preschool Children.” *Journal of the American Academy of Child & Adolescent Psychiatry* 26 (1):16–22.

6 Tables and Figures

Table 1: PRELIMINARY ANALYSIS

Outcome	Education				Log Earnings				
	M&F		M	F	M&F			M	F
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Misbehavior	-0.234				-0.026				
	(0.026)				(0.009)				
Externalizing Behavior		-0.154	-0.197	-0.088		0.027	0.035	0.022	0.045
		(0.030)	(0.037)	(0.051)		(0.010)	(0.010)	(0.008)	(0.022)
Internalizing Behavior		-0.116	-0.063	-0.185		-0.058	-0.050	-0.058	-0.033
		(0.029)	(0.038)	(0.044)		(0.010)	(0.010)	(0.008)	(0.021)
Cognition	1.219	1.220	1.178	1.289	0.199	0.199	0.083	0.070	0.108
	(0.030)	(0.030)	(0.041)	(0.044)	(0.009)	(0.009)	(0.011)	(0.009)	(0.022)
Educational Attainment					( )	( )	(X)	(X)	(X)
Obs.	7241	7241	3573	3668	4888	4888	4888	2643	2245

Notes: This table presents descriptive evidence linking early skills to educational attainment and earnings. Columns (1) to (4) contain parameter estimates from an ordered logit model used to link unobserved skills to educational attainment. Columns (5) to (9) present estimates from a linear regression of log-earnings on crude measures of unobserved skills. We present results separately by gender in Columns (3), (4), (8) and (9). To construct the crude measures of the three unobserved skills, we sum up all variables used to measure that skill according to Table B5 and then normalize each unobserved skill to have mean zero and standard deviation one. Misbehavior is a normalized aggregated measure, where we sum of variables used to measure both externalizing and internalizing behaviors. We report standard errors in parentheses.

Table 2: Additional Control Variables Used in the Analysis

Variables	Measurement System	Schooling Choices	Labor Outcomes
Class Size	x		
Percentage of Students Taking GCE exams	x		
Local Education Authority Dummy	x		
Number of Full-Time Teachers	x		
Financial Difficulties		x	x
London Dummy		x	x
Mother Education		x	
Father Education		x	
No Father Info.		x	
Father in Skilled Oc.		x	
Father in Managerial Oc.		x	
Working Mother		x	

Notes: This table summarizes the additional control variables we use in the measurement equations, the schooling choice equations and the labor outcome equations.

Table 3: MEASUREMENT SYSTEM: FROM SKILLS TO MISBEHAVIORS AND TEST SCORES

	Males			Females		
	Exter.	Inter.	Cog.	Exter.	Inter.	Cog.
Hostile Towards Children	1.000	0.000	0.000	1.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Hostile Towards Adults	1.639	0.206	0.026	1.672	0.166	0.009
	(0.082)	(0.043)	(0.014)	(0.080)	(0.042)	(0.012)
Anxiety Towards Children	1.614	-0.335	-0.053	1.288	-0.283	-0.033
	(0.074)	(0.036)	(0.010)	(0.066)	(0.035)	(0.008)
Anxiety Towards Adults	1.079	-0.328	-0.067	1.466	-0.380	-0.085
	(0.068)	(0.045)	(0.012)	(0.086)	(0.054)	(0.013)
Inconsequential Behavior	2.100	0.074	-0.145	1.509	0.150	-0.107
	(0.091)	(0.047)	(0.014)	(0.079)	(0.044)	(0.013)
Restless Behavior	0.778	-0.053	-0.058	0.554	-0.005	-0.055
	(0.042)	(0.023)	(0.008)	(0.036)	(0.022)	(0.007)
Depression	0.000	1.000	0.000	0.000	1.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Withdrawal	-0.983	1.132	0.137	-0.957	0.935	0.099
	(0.078)	(0.050)	(0.010)	(0.072)	(0.045)	(0.009)
Unforthcomingness	-2.117	1.899	0.195	-2.750	2.161	0.187
	(0.143)	(0.088)	(0.018)	(0.178)	(0.106)	(0.019)
Write Off Adults and Standards	0.082	1.077	0.069	-0.341	1.086	0.069
	(0.072)	(0.052)	(0.013)	(0.077)	(0.051)	(0.012)
Verbal Ability	0.000	0.000	1.000	0.000	0.000	1.000
	(.)	(.)	(.)	(.)	(.)	(.)
Reading Ability	-0.116	-0.070	0.832	-0.101	-0.040	0.820
	(0.108)	(0.065)	(0.021)	(0.130)	(0.075)	(0.021)
Non-Verbal Ability	-0.149	0.068	0.898	-0.103	0.008	0.928
	(0.099)	(0.062)	(0.020)	(0.122)	(0.071)	(0.021)
Math Ability	-0.087	-0.107	0.889	0.051	-0.189	0.879
	(0.087)	(0.054)	(0.019)	(0.113)	(0.065)	(0.020)

Notes: This table lists the parameter estimates of the measurement system (equation 1) for the male and female subsamples. Standard errors in parentheses. For the full set of parameter estimates, see Tables C2 and C3 in Appendix C.

Table 4: Education Attainment, Marginal Effects

	Males					
	No Qual.	CSE	O-lvl	A-lvl	H.Edu	H.Deg
Externalizing Behavior	0.014	0.021	0.004	-0.006	-0.022	-0.011
	(0.003)	(0.010)	(0.019)	(0.017)	(0.015)	(0.007))
Internalizing Behavior	0.003	0.004	0.009	-0.005	-0.008	-0.003
	(0.004)	(0.011)	(0.021)	(0.019)	(0.017)	(0.008)
Cognition	-0.063	-0.075	-0.064	0.051	0.069	0.081
	(0.005)	(0.008)	(0.014)	(0.012)	(0.010)	(0.007)
	Females					
	No Qual.	CSE	O-lvl	A-lvl	H.Edu	H.Deg
Externalizing Behavior	0.003	-0.025	0.013	0.006	0.009	-0.005
	(0.003)	(0.012)	(0.021)	(0.013)	(0.016)	(0.006)
Internalizing Behavior	0.014	0.034	-0.015	-0.015	-0.019	0.002
	(0.004)	(0.012)	(0.021)	(0.014)	(0.017)	(0.007)
Cognition	-0.071	-0.102	0.003	0.055	0.057	0.058
	(0.006)	(0.009)	(0.015)	(0.009)	(0.011)	(0.006)

Notes: This table lists marginal effects estimates from a multinomial logit model used to link socio-emotional and cognitive skills to educational attainment. We estimate educational attainment on a set of observable variables along with the unobserved skills. The coefficients on the three skills have been standardized to represent a 1 standard deviation effect. Standard errors in parentheses. Standard errors for the marginal effects are calculated using the delta method. For the full set of parameter estimates, see Table C4 in Appendix C.

Table 5: Labor Market Outcomes

	Log Hourly Wages				Log Hours Worked			
	Males		Females		Males		Females	
	[1]	[2]	[1]	[2]	[1]	[2]	[1]	[2]
Externalizing Behavior	0.055	0.064	0.023	0.013	0.012	0.015	0.047	0.047
	(0.018)	(0.017)	(0.024)	(0.020)	(0.008)	(0.008)	(0.024)	(0.025)
Internalizing Behavior	-0.099	-0.096	-0.043	-0.021	-0.014	-0.018	-0.023	-0.020
	(0.020)	(0.018)	(0.027)	(0.022)	(0.010)	(0.009)	(0.027)	(0.026)
Cognition	0.106	0.025	0.163	0.044	-0.015	-0.007	0.078	0.021
	(0.011)	(0.011)	(0.015)	(0.013)	(0.005)	(0.006)	(0.016)	(0.017)
Educational Attainment	( )	(X)	( )	(X)	( )	(X)	( )	(X)

Notes: This table lists parameter estimates from a linear model used to link socio-emotional and cognitive skills to hourly wages and hours worked. We regress log hourly wages and log hours worked on a set of observable variables along with the unobserved skills. The coefficients on the three skills have been standardized to represent a 1 standard deviation effect. Standard errors in parentheses. For the full set of parameter estimates, see Tables C5 and C6 in Appendix C.

Table 6: Robustness Across Datasets

	NCDS	BCS	NELS	PSID	CNLSY
			Years of Education		
Externalizing	-0.074	-0.122	-0.161	-0.176	-0.136
	(0.027)	(0.026)	(0.023)	(0.110)	(0.030)
Internalizing	-0.069	0.019	-0.165	-0.037	0.015
	(0.028)	(0.027)	(0.021)	(0.105)	(0.028)
Cognition	1.088	0.587	0.637	0.770	0.220
	(0.031)	(0.027)	(0.026)	(0.092)	(0.023)
N	7241	5789	5052	468	1597
			Log-Earnings		
Externalizing	0.032	0.020	0.028	0.068	0.002
	(0.009)	(0.011)	(0.009)	(0.034)	(0.024)
Internalizing	-0.047	-0.033	-0.040	-0.090	-0.066
	(0.009)	(0.011)	(0.009)	(0.033)	(0.025)
Cognition	0.079	0.064	0.019	0.044	0.077
	(0.010)	(0.011)	(0.011)	(0.025)	(0.019)
N	4888	5140	5161	249	1269

Notes: This table compares estimates from OLS regressions used to link socio-emotional and cognitive skills to to years of education and log earnings in early adulthood across datasets. For each dataset, we regress years of education and log weekly earnings on education attainment along with proxies for the unobserved skills. Standard errors in parentheses. For details of data construction and estimation results, see Appendix C.3.

Table 7: Log Weekly Earnings and Occupational Tasks

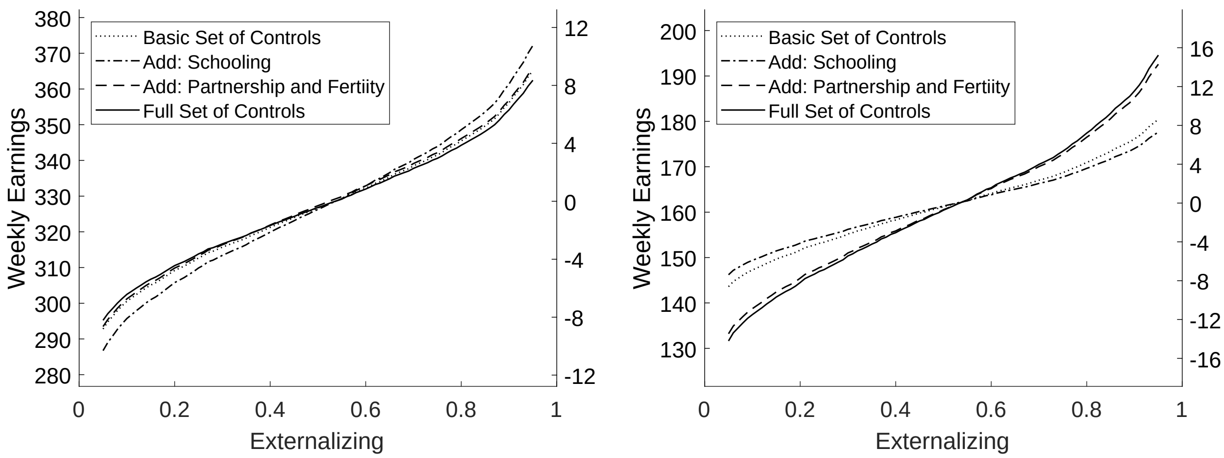
	Males		Females	
Externalizing Behavior	0.069	0.059	0.087	0.077
	(0.018)	(0.016)	(0.028)	(0.029)
Internalizing Behavior	-0.104	-0.090	-0.063	-0.049
	(0.020)	(0.018)	(0.032)	(0.035)
Cognition	0.019	0.015	0.028	0.050
	(0.012)	(0.011)	(0.019)	(0.020)
Ext. x Abstract		-0.026		0.003
		(0.015)		(0.038)
Int. x Abstract		0.011		-0.023
		(0.015)		(0.042)
Cog. x Abstract		-0.004		-0.017
		(0.010)		(0.023)
Abstract Intensity		0.045		0.114
		(0.027)		(0.060)
Ext. x Routine	0.036		-0.012	
	(0.015)		(0.031)	
Int. x Routine	-0.033		0.018	
	(0.015)		(0.034)	
Cog. x Routine	-0.006		0.012	
	(0.010)		(0.021)	
Routine Intensity	0.041		0.196	
	(0.020)		(0.049)	

Notes: This table lists parameter estimates from a linear model used to link socio-emotional and cognitive skills to hourly wages and hours worked across occupational tasks. We regress log hourly wages and log hours worked on a set of observable variables along with the unobserved skills and their interaction with the occupational task intensities. Task intensities are standardized composite measures of O*NET Work Activities and Work Context Importance scales, as in Acemoglu and Autor (2011) and Autor and Handel (2013). The abstract/social task measure is a normalized composite scale of six O*NET subscales: "analyzing data/information," "thinking creatively," "interpreting information for others," "establishing and maintaining personal relationships," and "guiding, directing, and motivating subordinates and coaching and developing others." The routine/manual task measure is a normalized composite scale of six O*NET subscales: "importance of repeating the same tasks," "importance of being exact or accurate," "structured versus unstructured work," "controlling machines and processes," "keeping a pace set by machinery or equipment," and "time spent making repetitive motions." The coefficients on the three skills have been standardized to represent a 1 standard deviation effect. Standard errors in parentheses. For the full set of parameter estimates, see Table D2 in Appendix D.

Table 8: Labor Market Outcomes, High SES vs. Low SES

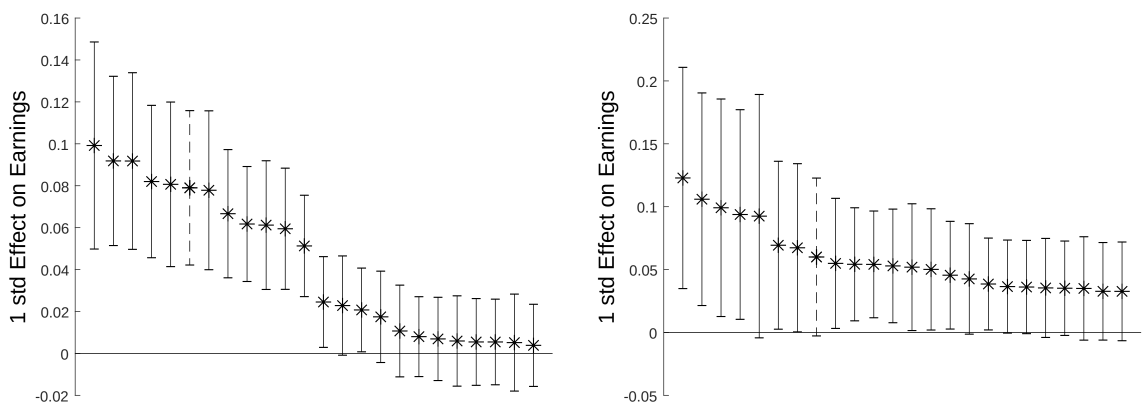
	Log Hourly Wages		Log Hours Worked	
	High SES	Low SES	High SES	Low SES
Externalizing Behavior	0.052	0.015	0.047	-0.030
	(0.013)	(0.029)	(0.013)	(0.032)
Internalizing Behavior	-0.074	-0.066	-0.040	0.043
	(0.015)	(0.032)	(0.015)	(0.036)
Cognition	0.052	0.046	0.002	0.050
	(0.009)	(0.019)	(0.010)	(0.023)
Police Involvement	-0.018	-0.012	0.061	0.041
	(0.028)	(0.057)	(0.042)	(0.074)
Police Info. Missing	-0.007	0.011	0.011	0.017
	(0.013)	(0.032)	(0.014)	(0.034)
Educational Attainment	(X)	(X)	(X)	(X)

Notes: This table lists parameter estimates from a linear model used to link socio-emotional and cognitive skills to hourly wages and hours worked, by high-SES and low-SES subsamples. We regress log hourly wages on a set of observable variables along with the unobserved skills. The coefficients on the three skills have been standardized to represent a 1 standard deviation effect. Standard errors are in parentheses. For the full set of parameter estimates, see Tables D6 and D7 in Appendix D.



(a) Males (b) Females

Figure 1: Decomposition of Effects of Externalizing on Weekly Earnings: Figure 1 visualizes the results from regressing weekly earnings on a varying set of controls presented in Tables C7 and C8. It illustrates how the predicted weekly earnings in regression models with different sets of controls vary, when we increase externalizing behavior from the lowest 5th percentile to the highest 95th percentile, keeping other latent skills and covariates at the population median.



(a) Males (b) Females

Figure 2: Distribution of Effects of Externalizing on Earnings: Figure 2 visualizes the effects on weekly earnings from 1 standard deviation increase in externalizing behavior from specifications that span all possible combinations of the dedicated measurements for externalizing and internalizing behaviors. It summarizes the results reported in Tables C32 and C33. The dashed bars indicate results from our benchmark model.