

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

Nicholas W. Papageorge[†]

Department of Economics, Johns Hopkins University, IZA and NBER

Victor Ronda[‡]

Department of Economics and Business Economics
and TrygFonden's Centre for Child Research, Aarhus University

Yu Zheng[§]

School of Economics and Finance, Queen Mary University of London and CEPR

ABSTRACT: Prevailing research argues that childhood misbehavior in the classroom is bad for schooling and, presumably, bad for labor market outcomes. In contrast, we argue that some childhood misbehavior represents underlying socio-emotional skills that are valuable in the labor market. We follow work from psychology and categorize observed classroom misbehavior into two underlying latent factors. We then estimate a model of educational attainment and earnings outcomes, allowing the impact of each of the two factors to vary by outcome. We find that one of the factors, labeled in the psychological literature as externalizing behavior (and linked, for example, to aggression), reduces educational attainment yet increases earnings. For men, it increases wages, while for women it increases hours. Unlike most models where skills that increase human capital through education also increase earnings, our findings illustrate how some socio-emotional skills can be productive in some economic contexts and not only unproductive, but counter-productive in others. Using a task model, we extend our results to show heterogeneity in returns for males, but not for females. We also find that different kinds of secondary schools exhibit different externalizing penalties, suggesting the tasks schools emphasize can affect how externalizing behavior interacts with education.

KEYWORDS: Labor, Education, Socio-Emotional Skills

JEL CLASSIFICATION: J10 J20 I20

*First draft: January 31, 2014. Current draft: June 16, 2020. 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.

[†][Corresponding author] papageorge@jhu.edu.

[‡]ronda@econ.au.dk.

[§]yu.zheng@qmul.ac.uk

1 Introduction

Economists generally recognize that human capital consists of multiple skills that drive educational and labor market outcomes. An early contribution is Willis and Rosen (1979), who distinguish between academic and manual skill. More recently, a burgeoning literature in economics has extended the concept of human capital to incorporate socio-emotional skills such as perseverance and grit (Heckman and Rubinstein, 2001).¹ It is not controversial that returns to skills can differ across sectors and that some skills are more productive in schooling than in work or in one occupation than in another. For example, to explain career choices, Willis and Rosen (1979) emphasize variation in the returns across occupations to manual versus academic skill.²

Despite potential differences in returns, the skills that constitute human capital are all typically seen as enhancing productivity — both in school and in the labor market. However, this view overlooks how some components of human capital could be productive in some economic contexts but could actually be counterproductive in others. If so, then policies designed to promote human capital accumulation in one context could have negative economic consequences in another. This is especially the case for policies that target socio-emotional skill formation aimed at children or adolescents, for whom socio-emotional skills have been shown to be relatively malleable and to influence a variety of outcomes (Heckman and Kautz, 2014).

In this paper, we demonstrate that some components of childhood misbehavior predict higher earnings even though they are associated with lower educational attainment. We examine a widely-studied pair of socio-emotional skills known as *externalizing behavior* and *internalizing behavior*.³ Externalizing behavior is linked to aggression and hyperactivity, while internalizing behavior captures anxiety, depression, shyness, unassertiveness and fearfulness (Ghodsian, 1977; Duncan and Magnuson, 2011; Duncan and Dunifon, 2012).⁴ Using

¹Excellent summaries of this research are found in Borghans et al. (2008) and Almlund et al. (2011).

²This 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 in many papers thereafter.

³Regarding 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.

⁴These two constructs are well established in the developmental psychology literature (Ghodsian, 1977; Campbell, Shaw, and Gilliom, 2000; Duncan and Magnuson, 2011; Duncan and Dunifon, 2012). They have also been widely used in economic research (see, e.g., Neidell and Waldfogel, 2010; Bertrand and Pan, 2013; Gertler et al., 2013; Heckman, Pinto, and Savelyev, 2013; Doyle, 2020, to cite a few). Moreover, these skills have been shown to be malleable (Gertler et al., 2013; Heckman, Pinto, and Savelyev, 2013) and to be predictive of a variety of educational outcomes (Campbell, Shaw, and Gilliom, 2000; Duncan and Magnuson, 2011; Bertrand and Pan, 2013). We study how these constructs relate to both schooling and labor outcomes,

a longitudinal dataset from Britain, the National Child Development Survey (NCDS), we 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. We also estimate the model separately for males and females. The key empirical fact we establish is that, for both genders, one of the factors underlying observed classroom misbehavior, externalizing behavior, lowers educational attainment, but is also associated with higher earnings. In other words, we provide evidence demonstrating that a penchant for *breaking bad* can be good.⁵ For males, the externalizing behavior labor market premium is driven mainly by an increase in wages, while for females, it is primarily driven by an increase in hours worked.

We go on to conduct a series of sensitivity analyses and extensions. We show that the mixed effects of externalizing behavior are not driven by selection into employment, occupation, marriage and fertility, though it is related to these lifecycle choices and outcomes. We also show that our results are robust to a host of alternative modeling assumptions and to the inclusion of different sets of control variables. Finally, a benefit of studying internalizing and externalizing behaviors is that they are measured in a variety of data sets, allowing researchers to compare findings across cohorts and countries. We are thus able to show that our findings on mixed returns extend to other data sets, including the 1970 British Cohort Study (BCS), the National Education Longitudinal Study of 1988 (NELS), the Panel Study of Income Dynamics (PSID), and the National Longitudinal Survey of Youth (NLSY) 1979: Children and Young Adults (CNLSY). This provides compelling evidence that our findings are not unique to one particular group or era, but instead reflect an empirical regularity found across cohorts and countries.

One interpretation of our results is that they reflect how a skill interacts with general features of schooling and the labor market. For example, high externalizing workers may tend to be energetic, which increases their productivity across occupations. Similarly, the externalizing schooling penalty may arise from difficulties related to a fundamental aspect of

which distinguishes our contribution to prior work. We discuss these differences in more detail in Section 2.3.

⁵According 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 an American television show in which the protagonist is an unsuccessful chemist who reveals a striking talent for producing illicit drugs. The show offers an extreme example of how certain skills or behaviors may lead to low productivity in one sector and high productivity in another.

education, such as learning new concepts. Alternatively, our main estimates could capture averages that obscure heterogeneity across schools and occupations. We conceptualize and test for this possibility using a task-based framework, wherein skills influence labor market productivity and educational attainment through their impact on the performance of occupational tasks (O*NET task-intensity scales as in Acemoglu and Autor, 2011).⁶ We find that the returns to skills are task-dependent: externalizing behavior decreases productivity in abstract/social tasks such as “establishing and maintaining personal relationships,” and increases work productivity in routine/manual tasks, such as “keeping a pace set by machinery or equipment.” Next, we apply the idea of task-dependent returns to schooling. However, lacking detailed data on the tasks performed in different schools we exploit differences in school types in Great Britain during the period we study, in particular, comprehensive versus private or grammar schools. Earlier work (see, e.g., Harmon and Walker, 2000) indicates that these schools have different climates, curricula, codes of conduct, activities and programs, suggesting there may be differences in the tasks they emphasize. We demonstrate that the externalizing penalty for educational attainment is significantly smaller for students attending comprehensive schools compared to grammar and private schools. This provides preliminary and suggestive evidence that which tasks are emphasized in different types of schools can help to explain the externalizing schooling penalty.

While we are cautious about using our findings to draw specific policy conclusions, we discuss several implications of our findings. First, and most directly related to the contexts we study, identifying a skill that raises earnings but lowers educational attainment runs counter to the typical view of ability bias in estimates of the returns to education (Becker, 1967). Often, the presumption is that the unobserved skills leading to success in education also promote earnings.⁷ Further, while heterogeneity in returns across tasks allows individuals to sort away from occupations where their skills are less valuable (Willis and Rosen, 1979) or the degree programs these occupations require (Prada and Urzúa, 2017), avoiding primary and secondary schooling is exceedingly costly (and in many places forbidden). For most occupations, formal education is a crucial and lucrative labor market input, which puts high-externalizing individuals at a disadvantage. Our findings exploiting differences in school types suggest at least the possibility that schools could modify the tasks they require of students to minimize the negative impact of externalizing behavior on educational attain-

⁶The idea that some dimensions of human capital are specific to certain tasks is not new (see, e.g., Gibbons and Waldman, 2004). We extend this insight to the idea that returns to childhood skills might also be dependent on the tasks required at work and at school. This is intuitive since cognitive and socio-emotional skills are often measured from performance on tasks (see Kautz et al., 2014).

⁷There are a number of exceptions. For example, Card (2012) shows that IV estimates could lead to larger coefficients on education in wage equations. The argument is based on heterogeneity in treatment effects coupled with the particular group for whom the IV affects attendance.

ment. Yet, future research on school-based tasks would be needed to assess the benefits and drawbacks of any such policies.

More broadly, our findings provide evidence that a skill can be both productive and not only less productive, but counterproductive, depending on the economic context. Earlier literature has explored the idea that the returns to skill can vary. For example, Lundberg (2013) finds evidence of demographic differences in the relationship between personality traits and high school completion. More closely related, Levine and Rubinstein (2017) show that when accompanied by high cognitive skill, individuals who engage in illicit behaviors during high school are more likely to become entrepreneurs and, among those who become incorporated (roughly 3.4% of their sample), earn more than individuals with low illicit behaviors. This illustrates the idea that a presumably bad set of behaviors can lead to success in some domains.⁸ We show that this type of pattern is not limited to a specific set of behaviors and a small sector of the labor market and need not be accompanied by high cognition. Rather, we show that a prevalent childhood socio-emotional skill (one that has been measured in several data sets and studied for decades in several disciplines) has opposite effects on two of the most critical phases of life which virtually everyone experiences: schooling and work. This leads to a more general point: the skills that are valuable during childhood are not necessarily valuable in adulthood.

Finally, the notion that a skill can have mixed effects suggests we should reimagine how best to measure and evaluate skills. The task-based approach offers one way forward. In particular, understanding which tasks are emphasized in the contexts in which returns to skills are measured provides information on the degree to which estimated returns generalize to other economic contexts. To that end, the task model could be extended beyond occupations to understand why certain skills are more or less productive at different kinds of schools. It could be further extended to other economic outcomes in which skills interact with tasks to drive productivity and outcomes, including parenting, marriage and interpersonal relationships. In short, our findings suggest it is not generally meaningful to think of the various skills comprising human capital as good or bad *per se*. Different phases of life require different tasks and thus different skills. The value of each particular skill depends on the economic context in which it is measured, distinguished by the predominant sets of tasks that characterize it.

The paper is organized as follows. In Section 2, we introduce the NCDS dataset, discuss measurements of misbehavior that identify externalizing and internalizing behavior, and

⁸Levine and Rubinstein (2018) reiterate the point that illicit behaviors predict success in entrepreneurship while also examining how skills and wealth drive selection into entrepreneurship, other self-employment and salaried employment using a three-sector Roy model.

conduct a preliminary data analysis. In Section 3, we describe the main “benchmark” econometric model we estimate, including the measurement system used to identify latent skills, along with estimation, and present the main results. Section 4 reports results on a host of sensitivity analyses, including on alternative assumptions to identify our econometric model and replication of our main empirical results in a variety of additional data sets. In Section 5, we explore heterogeneity in the returns to skills in the labor market and schooling. In Section 6, we discuss the implications of our findings for our understanding of human capital. Section 7 offers brief concluding remarks focusing on limitations to policy implications along with directions for future research.

2 Data and Preliminary Analysis

In this section, we introduce the NCDS dataset, describe key variables used in our analysis and provide estimates from a preliminary econometric model relating childhood misbehavior with schooling and earnings. We demonstrate that once we treat externalizing and internalizing behaviors separately, externalizing behavior is associated with higher earnings even though it also predicts lower educational attainment.

2.1 The National Child Development Study

The NCDS is an ongoing longitudinal survey that follows 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. Therefore, the dataset allows us to relate misbehavior in elementary school to educational attainment along with labor market 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 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.⁹

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), where the latter are more academically demanding.¹⁰ 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).

Summary statistics on education, labor market outcomes and a basic set of controls are found in Table A1 in Appendix A. 51% of our sample is female. Females in our sample are less educated compared to males. On average, employed females' wages are 29% lower — and hours are 51% lower — than those reported by employed males. Males are also significantly more likely to be employed and, conditional on employment, to be self-employed. In general, large gender differences in schooling and labor market outcomes suggest that we should allow the parameters of our econometric model to vary by gender.

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. We measure socio-emotional skills using variables describing classroom misbehavior. When a child

⁹Most of the drop in observations is due to attrition at the fifth survey. Out of the original 18,555 births, only 11,364 individuals 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." 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 is statistically significant. Summary statistics for the full sample are reported in Tables A2 and A3 in Appendix A, where we provide additional summary statistics for variables and samples used throughout this study.

¹⁰CSEs and O-Levels were replaced by the General Certificates of Secondary Education (GCSE) in 1986 after individuals in our sample had finished their schooling.

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.¹¹ 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.

In Table A4 in Appendix A , we present summary statistics for each of the BSAG maladjustment syndromes, separately by gender. Values range from 0 to 15, with a higher value indicating a higher prevalence of a particular maladjustment syndrome. The means are usually low, and most of the variation comes from individuals with low values of maladjustment for each measurement. To avoid disproportionate effects from outliers, we use the logarithm of each BSAG score plus one as the relevant measurement in the benchmark model in Section 3. Our results are robust to different specifications. Overall, females appear to misbehave less frequently than males. Specifically, males exhibit higher scores for all of the BSAG variables except for “anxiety for acceptance by adults.” Gender differences in misbehavior are consistent with earlier findings documented for Great Britain (Duncan and Magnuson, 2011; Duncan and Dunifon, 2012) and the U.S. (Bertrand and Pan, 2013).

Following earlier work (see e.g., Cunha, Heckman, and Schennach (2010)), we measure cognitive skill using a set of math and reading test scores. Test score averages are found in Table A4 in Appendix A . These tests are administered when children are 11 years old. According to the table, girls score marginally higher than boys on tests of verbal and non-verbal ability, where non-verbal ability measures identification of shapes and symbols. In contrast, average math scores for boys are marginally higher.

The benchmark econometric model used in our main analysis, described in Section 3.1, includes a measurement system that uses these observed maladjustment syndromes and test scores as measurements to identify unobserved skills. In contrast, for the preliminary analysis, we use the variables described above to construct crude measures of the unobserved

¹¹In 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.

skills. To construct these measures of socio-emotional skill, we follow Ghodsian (1977), who proposed dividing up the BSAG syndromes into two groups based on apparent differences among what behaviors the syndromes capture.¹² Variables assigned to each group are then summed to create two new variables. 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. Similarly, we obtain a measure of cognitive ability by summing test scores. How we assign measurements to each of the three skills is summarized in Table 1.¹³ In addition, we construct a generic measure of misbehavior by simply summing up all ten syndromes. This variable is used to illustrate how findings change once we recognize that misbehavior captures two separate socio-emotional skills. Finally, we normalize these newly constructed crude measures of externalizing behavior, internalizing behavior, cognition, and misbehavior, so that each variable has mean equal to zero and variance equal to one for the full sample. Summary statistics for these measures are reported in Table A4 in Appendix A separately by gender. According to the table, boys exhibit significantly higher externalizing and internalizing behaviors compared to girls. Boys are roughly 0.3 standard deviations higher on average. We also find that average cognition for girls is about 0.06 standard deviations higher than it is for boys.

2.2.3 Additional Control Variables

There are three sets of additional control variables that we use in our subsequent analyses. Table 2 summarizes which additional variables are included in which equations. Conditioning on these variables helps to mitigate concerns related to omitted variables bias, but it is important to state at the outset that their inclusion does not drive our key findings. In both our preliminary analysis using crude measures of skills, as well as our benchmark econometric model that features a formal measurement system, we obtain our main results

¹²This 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. We replicate this analysis in Appendix B.

¹³These measures have been externally validated in the sense that they are positively correlated with a range of other measurements of social maladjustment from teachers, professional observers, parents and peers (Achenbach, McConaughy, and Howell, 1987). Moreover, they have been studied extensively by psychologists researching child development and, of late, by some economists (Blanden, Gregg, and Macmillan, 2007; Aizer, 2009; Agan, 2011; Heckman, Pinto, and Savelyev, 2013). Both Aizer (2009) and Agan (2011) study how externalizing behavior is linked to anti-social and criminal activity. For general surveys of research on externalizing and internalizing behaviors, see Duncan and Magnuson (2011) and Duncan and Dunifon (2012).

once we include measures of cognition, externalizing behavior and internalizing behavior.¹⁴

The first set of additional variables are two basic controls, which are included in all schooling and outcome equations. The first is an indicator for childhood poverty. The variable we construct, “Financial Difficulty,” takes the value one if (i) the interviewer reported that the household appeared to be experiencing poverty in 1965 or (ii) a member of the household self-reported having financial difficulties in the 12 months prior to being observed in either 1969 or 1974, and zero otherwise.¹⁵ The second basic control is an indicator variable for living in London. Including this variable is common practice using the NCDS given possible London-specific differences in schooling or labor outcomes.¹⁶ Summary statistics for the financial difficulty and London dummy variables are found in Table A1 in Appendix A. 36% of the sample lives in or around London before age 16 versus 30% at age 33. 16% of our sample experienced financial difficulty in their childhood.

We include a second set of control variables in schooling equations, but not in other equations (Heckman, Humphries, and Veramendi, 2018). The reasoning is that externalizing behavior could capture a productive skill on the labor market, but could also relate to family backgrounds that lead to lower schooling, such as an absent father or low parental education. If so, an estimated negative impact on schooling may simply reflect omitted family background variables rather than mixed effects of a socio-emotional skill. To address this concern, we include a set of family background variables, which are excluded from the outcome equations: whether the mother studied beyond the minimum schooling age, whether the father studied beyond the minimum schooling age, whether the father’s information is missing, father’s occupation, and mother’s employment status, all observed when the child is age 11.

A third set of control variables related to school characteristics is included in our measurement system to address possible mis-reporting differences across teachers and schools. We postpone a discussion of this final set of control variables until we introduce the measurement system in Section 3.1.

¹⁴Summary statistics for additional control variables are found in Table A5 in Appendix A.

¹⁵We include this variable in all equations because it is a variable along which we stratify our sample in one of our subsequent analyses, briefly discussed in Section 6.

¹⁶In the NCDS, the definition of region of residence changed from the first 4 surveys (ages 0, 7, 11 and 16) to the fifth (age 33) survey. Before age 16, we say an individual lives in or around London if he or she lives in East, South East or South England. At age 33, we say an individual lives in or around London if he or she lives in South East England. The reason is that the categories change across surveys. 57%, 85% and 72% of individuals living in East, South East, or South England at age 11 are living in South East England at age 33. Individuals in these regions have higher earnings on average than individuals living in other regions. The results are not sensitive to changes in the classification or whether we include dummies for all the possible regions of residence.

2.3 Relating Misbehavior, Schooling and Earnings

Our preliminary analysis relates the crude measures of externalizing behavior, internalizing behavior, and cognition to schooling and labor market outcomes. An advantage of the preliminary analysis is that this approach has been taken in previous studies, which means we can directly compare our findings to those in earlier work. In particular, we can show that securing our key results — including the finding that externalizing behavior has mixed effects on schooling and earnings — does not require a more sophisticated measurement system, but emerges once we control for measures of internalizing behavior and cognition as they have been constructed in earlier work. Earlier work includes research using the NCDS dataset studying externalizing and internalizing behaviors (Farmer, 1993, 1995; Jackson, 2006). It also includes research using different samples since the division of misbehavior into these two socio-emotional skills extends to other data sets, including the CNLSY and the PSID (Yeung, Linver, and Brooks-Gunn, 2002; Agan, 2011; Bertrand and Pan, 2013). Finally, using crude measures facilitates a comparison of empirical patterns across data sets, which we perform in Section 4.2. The reason is that other data sets often contain summary measures of externalizing and internalizing behaviors, and therefore we cannot always apply the same type of measurement system used in our benchmark econometric model estimated from the NCDS data. As we discuss in Section 3.1 when introducing the benchmark econometric model, use of these crude measures imposes a number of unattractive assumptions that the formal measurement system allows us to relax.

For the preliminary analysis, we explain educational attainment using an ordered probit model. The outcome variable is one of the six possible schooling levels.¹⁷ Formally, defining s_i^* as a latent variable determining schooling, we estimate regressions of the following form:

$$s_i^* = E_i\psi^E + I_i\psi^I + C_i\psi^C + Z_i'\beta_s + e_i^S \quad (1)$$

where observed schooling $s_i = s$ if $\mu_L^s \leq s_i^* < \mu_H^s$ and μ_L^s and μ_H^s are the particular bounds for schooling level s . E_i and I_i are the crude measures of externalizing and internalizing behaviors and C_i is a crude measure of cognition, constructed according to the description in Section 2.2.2. Recall, we have normalized the measures of unobserved skills. Abusing notation somewhat, Z_i is a vector of control variables, which varies across specifications. Finally, e_i^S is a normally distributed disturbance.

Estimates of equation (1) are reported in Table 3. We start by regressing schooling on

¹⁷We use an ordered probit in our preliminary analysis to simplify exposition. However, results are robust to using a more flexible specification, such as a multinomial logit or probit model. In the benchmark econometric model used in our main analysis, we estimate a multinomial logit model.

the crude measure of generic misbehavior and estimate a negative relationship in Columns [1]-[2]. The magnitude declines when we include cognition, suggesting a negative correlation between the two variables. In Columns [3]-[6], we allow externalizing and internalizing behaviors to have separate effects on schooling. In Column [3], we start by only including externalizing behavior. In the subsequent three columns, we add our measures of cognition and of internalizing behavior, and then both. Results in Column [6], including all three measures of skills, show that both externalizing and internalizing behaviors independently lower schooling attainment, while a higher level of cognition leads to higher educational attainment. Moreover, the impact of cognition is roughly ten times larger than the impacts of either socio-emotional skill. While models in Columns [1]-[6] include indicators for London, financial difficulties and female, in Column [7] we also include family background variables (the second set of additional control variables described above). Most affect schooling in ways we would expect. For example, higher parental education has a positive impact on the respondent’s own education. However, including these variables has very little effect on the size of the coefficients on cognitive and socio-emotional skills.

In Columns [8] and [9], we use the same set of regressors as in Column [7], but stratify the sample by gender. Across genders, externalizing and internalizing behaviors have a negative impact on schooling, while cognition strongly raises educational attainment. Comparing genders, the negative coefficient on externalizing is larger for males, while the coefficient on internalizing is larger for females. Patterns are similar if we compute marginal effects, which are reported in Table A7 in Appendix A.

To explain earnings, we regress log weekly earnings at age 33, conditional on being employed, onto measures of socio-emotional and cognitive skills (Table 4).¹⁸ Defining y_i as log earnings at age 33 for individual i , we estimate OLS regressions of the following form:

$$y_i = E_i\phi^{\mathbf{E}} + I_i\phi^{\mathbf{I}} + C_i\phi^{\mathbf{C}} + X_i'\beta + e_i^Y \quad (2)$$

where X_i includes the basic set of controls (indicators for female, financial difficulties and living in London) and may or may not include schooling outcomes. Columns [1] and [2] contain estimates using the single measure of misbehavior, controlling for cognition or not. In line with previous research (e.g., Segal, 2013), we find this single measure of misbehavior is associated with both lower schooling and lower earnings.

Results change dramatically when we view childhood misbehavior as reflecting two distinct factors and control for cognition. In Columns [3]-[6] of Table 4, we regress log earnings

¹⁸As with schooling, our choice to use earnings as the outcome variable is for ease of exposition. In our main benchmark model, we allow externalizing to have separate effects on wages and hours.

onto E_i and I_i separately. The positive price of externalizing behavior emerges as soon as we control for internalizing behavior and cognition (Column [6]). Comparing Column [4] to [6], we deduce that if externalizing and cognition are negatively correlated while cognition is valuable for earnings, then omitting cognition will downwardly bias the estimated impact of externalizing. Similarly, if externalizing and internalizing are positively correlated while internalizing is bad for earnings, then omitting internalizing again generates a downward bias on estimated returns to externalizing behavior (Columns [5] and [6]). These results suggest that it is important to account for the correlation patterns across all three skills to produce estimates interpretable as skill prices. In our benchmark econometric model described in Section 3.1, we thus allow arbitrary correlations across the latent factors capturing underlying skills.

Column [6] presents strong initial evidence that externalizing behavior carries an earnings premium. When we control for schooling outcomes in the earnings equation (Column [7]), the positive coefficient on externalizing behavior becomes even larger. When we further separate the sample by gender (Columns [8] and [9]), we conclude that mixed effects of externalizing behavior hold for both males and females in our sample.

The results from the preliminary analysis presented in Tables 3 and 4 provide initial evidence that a socio-emotional skill that is productive on the labor market is not productive in school.¹⁹ It is also worth highlighting that, according to Table 3, the coefficient on externalizing is positive whether or not we control for schooling. An alternative possibility would be that externalizing behavior predicts higher earnings only after we have controlled for its negative impact on schooling. Such a finding would still support the idea that externalizing is potentially valuable in the labor market. However, it would also suggest that lower levels of externalizing behavior could have a positive net effect on labor market outcomes since the negative effect of externalizing through schooling on earnings would overwhelm the direct positive effect on earnings. In contrast, estimates suggest that externalizing behavior has a positive net effect on earnings despite having a negative impact on schooling.

¹⁹This preliminary evidence is robust to a host of alternative specifications, which are explored in Appendix A. One alternative is to measure earnings at age 42 or 50, which yields similar results (Appendix A.5). We continue to use labor market outcomes at age 33 since otherwise we lose a considerable number of observations due to sample attrition as the NCDS cohort ages. We also show explicitly that the positive relationship between externalizing behavior and earnings emerges as soon as we control for internalizing behavior and cognition, and does not require any additional controls (Appendix A.3). We also explore potential non-linear effects and complementarities between factors and find no evidence of either (Appendix A.4). Finally, we report estimates where additional socio-emotional skills are included, in particular, the Big 5 personality traits (Appendix A.6). The externalizing premium decreases by about 20% when we control for the Big 5 personality traits. However, the Big 5 were measured at age 50 after earnings and education were realized, which could introduce bias due to simultaneity. In Section 3.2.3, we discuss this point in greater detail.

The positive association between externalizing behavior during childhood and adult earnings has generally not been recognized in previous literature on the economic consequences of childhood misbehavior.²⁰ There are several reasons for this lack of recognition. First, most of the literature on the long run effects of childhood misbehavior takes for granted that externalizing is broadly unproductive, focusing instead on negative impacts on school-related outcomes (Bertrand and Pan, 2013). This may be a result of data limitations since linking childhood misbehavior to labor market outcomes requires a long panel spanning from childhood well into adulthood. However, even studies using the NCDS dataset have not linked externalizing behavior to earnings (Farmer, 1993, 1995; Jackson, 2006).

Second, many studies use a single aggregated measure of childhood misbehavior or maladjustment. We discuss two such studies which are otherwise similar to ours, highlighting the importance of recognizing that misbehavior reflects distinct socio-emotional skills with potentially different returns in the labor market. Similar to our paper, Fronstin, Greenberg, and Robins (2005) uses the NCDS to study the effect of childhood maladjustment on labor market outcomes. Importantly, to justify the use of a single aggregated measure of misbehavior, the authors refer to earlier work showing that externalizing and internalizing behaviors have a similar effects on mental health in early adulthood, which might suggest similar effects on other outcomes (Chase-Lansdale, Cherlin, and Kiernan, 1995). In contrast, we show that the two factors have opposite effects on earnings.

Another related paper, Segal (2013), uses the National Education Longitudinal Survey (NELS) to relate five different teacher-reported measures of childhood misbehavior to education and labor market outcomes. The author shows that a variable that summarizes five measures of “misbehavior” predicts lower earnings. However, when the five measures are included individually in the same regression, the coefficient for one of the five measures, “disruptiveness,” is positively related to earnings. Segal (2013) argues that the positive effect of disruptiveness on earnings is spurious since the association reverses when the other four measures are excluded from the regression (see Footnote 32 on p. 23 of the study). In contrast, we argue that these differences in estimates highlight the importance of including multiple measures of possibly correlated variables capturing misbehavior. We also show that summing multiple measures potentially obfuscates how each skill underlying misbehavior can have different effects on economic outcomes.

²⁰As discussed earlier, in a related contribution, Levine and Rubinstein (2017) show that illicit behavior during adolescence, when accompanied by high cognition, predicts success for the 3.4% of the sample that selects incorporated entrepreneurship. Our findings on externalizing apply more broadly, as we discuss in Section 5, where we explore heterogeneity in returns.

3 Model and Estimation Results

Summing the BSAG maladjustment syndromes and test scores to create crude measures of underlying skills is simple and straightforward, but also imposes a number of unattractive assumptions. For example, each measurement is assigned to only one underlying skill. One implication is that externalizing behavior is assumed to have no effect on cognitive test scores. Moreover, measurements assigned to each skill are given equal weights. In this section, we develop our benchmark econometric model, which relaxes some of these assumptions. The benchmark model features a formal measurement system, which treats observed maladjustment syndromes and test scores as measures with error of underlying skills. The model produces estimates of the joint distribution of latent skills and the mapping of such skills to observed measurements, which depends in part on the precision of each measure. The measurement system also allows each measure to provide information about more than one factor. For example, a maladjustment syndrome can be a measure of both socio-emotional skill and of cognition. Moreover, externalizing behavior can affect maladjustment syndromes along with cognitive test scores. Using this framework, we are able to secure identification of the impact of underlying skills imposing relatively few assumptions. Results using the benchmark model are discussed in Section 3.2. As the benchmark model still imposes some somewhat arbitrary assumptions, which we detail below, we assess the sensitivity of the results to alternative assumptions in Section 4.

3.1 Model

3.1.1 Parameterizations of the Schooling Decision Rule and Potential Outcomes

We approximate the schooling decision with a linear-in-parameters multinomial logit model with 6 schooling levels: $s \in \{0, 1, \dots, 5\}$. Taking schooling level 0 as the base state, let the log-odds of schooling level s be

$$I_s = \log \frac{Pr(S = s)}{Pr(S = 0)} = \underbrace{Z}_{\text{observed by econometrician}} \cdot \beta_s + \underbrace{\eta_s}_{\text{unobserved by econometrician}}, \quad s = \{1, \dots, 5\}, \quad (3)$$

where Z is a vector of variables observed by the econometrician that affect the schooling decision (see Table 2), 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. More specifically, the log hourly wage, y , and the log weekly working hours, h , are represented by the following two equations:

$$y = \underbrace{X}_{\text{observed by econometrician}} \cdot \beta_Y + \sum_{s=1}^5 \gamma_{s,Y} \cdot \underbrace{\mathbf{1}[s]}_{\text{observed by econometrician}} + \underbrace{U_Y}_{\text{unobserved by econometrician}} \quad (4)$$

$$h = \underbrace{X}_{\text{observed by econometrician}} \cdot \beta_H + \sum_{s=1}^5 \gamma_{s,H} \cdot \underbrace{\mathbf{1}[s]}_{\text{observed by econometrician}} + \underbrace{U_H}_{\text{unobserved by econometrician}} \quad (5)$$

X is the set of basic controls shown in Table 2 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 assume there exists a vector \mathbf{f} of skills that are unobserved by the econometrician and which generate all dependence across the η_s , U_Y , and U_H . More specifically, suppose

$$\eta_s = \mathbf{f}'\alpha_{\mathbf{S}} + \nu_s, \quad (6)$$

$$U_Y = \mathbf{f}'\alpha_{\mathbf{Y}} + \omega_Y, \quad (7)$$

$$U_H = \mathbf{f}'\alpha_{\mathbf{H}} + \omega_H, \quad (8)$$

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 outcomes, the log hourly wage and the log weekly hours worked.

3.1.2 Measurement System for Unobserved Skills f

The vector of skills \mathbf{f} is not directly observed, but it can be proxied by a set of observable measurements. We allow for the ten BSAG maladjustment syndromes and four aptitude test scores measured at the age of 11 (Table 1) to be proxies for three latent skills. 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. We propose a

linear measurement system:

$$M = \begin{pmatrix} M_1 \\ \vdots \\ M_K \end{pmatrix} = \begin{pmatrix} m_1 + \sum_{j=1}^3 \lambda_{1j} f_j + W\delta_1 + \varepsilon_1 \\ \vdots \\ m_K + \sum_{j=1}^3 \lambda_{Kj} f_j + W\delta_K + \varepsilon_K \end{pmatrix}, \quad (9)$$

where m_k is the mean of the measurement k , and λ_{kj} is the factor loading of latent skill j on the k th measurement.²¹

The latent skills follow a joint normal distribution, with mean μ and variance-covariance matrix Σ :

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \sim N(\mu, \Sigma) = N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \right) \quad (10)$$

Referring back to Table 2, notice we include a vector of additional observables denoted W and associated coefficients δ . W includes class size, the percentage of students in the respondent's school taking GCE exams, a dummy for the local educational authority (LEA, i.e. public schools), and the number of full-time teachers in the school. These additional variables are included to address the concern that school attributes simultaneously affect schooling and labor outcomes along with teacher mis-reporting. If we omit these variables, we may misattribute variation in outcomes to variation in skills that is actually due to differences in schooling attributes.²²

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 , $s \in \{1, \dots, 5\}$ 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,

$$\nu_s \perp \nu_{s'}, \forall s \neq s', \quad (11)$$

$$\omega_Y \perp \omega_H, \quad (12)$$

$$\omega \perp \nu. \quad (13)$$

²¹The BSAG maladjustment scores range from 0 to 15 but most individuals have a score near 0 (see Table A4 in Appendix A). To account for this feature of the data, we use the logarithm of each BSAG score plus one as the relevant measurement in the measurement system.

²²As with other control variables, results are not affected if these variables are omitted.

Assumptions (11), (12) and (13) maintain independence of the shocks over schooling categories, and across schooling and labor market outcomes. This assumption is testable and in Section 4.1.2 we provide evidence that \mathbf{f} adequately captures the unobserved covariation of the three outcomes.

In addition, we array the measurement errors, ε_k , $k \in \{1, \dots, K\}$ into a vector $\varepsilon = (\varepsilon_1, \dots, \varepsilon_K)$ and assume that,

$$\varepsilon_k \perp \varepsilon_{k'}, \forall k \neq k', \quad (14)$$

$$(\omega, \nu) \perp \varepsilon. \quad (15)$$

Assumptions (14) and (15) maintain that the measurement errors are independent from each other, and independent from the shocks.²³

Lastly, we assume that,

$$(\nu, \omega, \varepsilon) \perp (\mathbf{f}, Z, X, W), \quad (16)$$

$$\mathbf{f} \perp (X, Z, W). \quad (17)$$

Assumption (16) assumes independence of all the shocks and measurement errors with respect to factors and observables, and Assumption (17) assumes independence of factors with respect to observables.²⁴ The latter assumption might seem restrictive. In Section 4.1.6, we discuss alternative sets of models where we change the set of variables in Z and X , including a model where we allow Z and X to be empty vectors.

Identification of the measurement system requires further restrictions. One restriction that secures identification is to choose three “dedicated measures,” that is, for each skill we choose one measure that is only affected by that skill (Williams, 2018). 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). We allow all three skills to load on the remaining 11 measurements. The choice of dedicated measures is somewhat arbitrary, yet 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

²³In 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.

²⁴Williams (2018) discusses these assumptions in more detail. In particular, Williams (2018) describes conditions under which Assumption (17) can be relaxed.

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). In Section 4, we discuss changes to results when we rely on alternative restrictions, including different choices of dedicated measurements. 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. These identifying restrictions amount to

$$\begin{aligned}
M_1 &= m_1 + 1 \cdot f_1 + 0 \cdot f_2 + 0 \cdot f_3 + W\delta_1 + \varepsilon_1 \\
M_2 &= m_2 + 0 \cdot f_1 + 1 \cdot f_2 + 0 \cdot f_3 + W\delta_2 + \varepsilon_2 \\
M_3 &= m_3 + 0 \cdot f_1 + 0 \cdot f_2 + 1 \cdot f_3 + W\delta_3 + \varepsilon_3
\end{aligned}
\tag{18}$$

3.1.4 Likelihood and Estimation Procedure

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

$$\Phi = (\beta, \gamma, \alpha, \Xi)
\tag{19}$$

where β denotes the set of coefficients on the vectors of observables absent the schooling level in equations (3)-(5), 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 (9) and (10).

We estimate the model by simulated maximum likelihood in two stages. In the first stage, we estimate the measurement system for unobserved skills. In the second stage, given the parameter estimates $\hat{\Xi}$ found in the first step, we estimate the remaining structural parameters, (β, γ, α) .²⁵ We implement the estimation for boys and girls separately; that is, we allow all parameters to differ by gender.

In the first stage, 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 times and, for each draw of the factors, compute the probability of observing each measurement.²⁶ 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

²⁵Along with other robustness checks, in Section 4, we discuss an alternative specification where we estimate the measurement system jointly with outcomes.

²⁶For estimation, we set $T = 500$. Results are robust if we use larger or smaller numbers and are available upon request.

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) .

In the first stage, 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:

$$\mathcal{L}_1^{(g_1)} = \sum_{i=1}^I \log \left(\frac{1}{T} \sum_{t=1}^T \mathbf{f}_{it}^{\mathbf{M},(g_1)}(M_i) \right) \quad (20)$$

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, taking $\hat{\Xi}$ 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 , $(\mathbf{f}_{it}^{\mathbf{S},(g_2)}(s))$, the probability of observing wage y_i , $(\mathbf{f}_{it}^{\mathbf{Y},(g_2)}(y_i))$, and hours worked h_i , $(\mathbf{f}_{it}^{\mathbf{H},(g_2)}(h_i))$, for individual i , draw t and parameter suggestion (g_2) . The simulated log likelihood in the second stage is given by:

$$\begin{aligned} \mathcal{L}_2^{(g_2)} = \sum_{i=1}^I \log \left(\frac{1}{T} \sum_{t=1}^T \mathbf{f}_{it}^{\mathbf{M},(\hat{\Xi})}(M_i) \times \prod_{s=0}^5 \mathbf{f}_{it}^{\mathbf{S},(g_2)}(s)^{\mathbf{1}[s=s_i]} \right. \\ \left. \times \mathbf{f}_{it}^{\mathbf{H},(g_2)}(h_i)^{\mathbf{1}(e_i=1)} \times \mathbf{f}_{it}^{\mathbf{Y},(g_2)}(y_i)^{\mathbf{1}(e_i=1)} \right) \end{aligned} \quad (21)$$

where s_i represents the observed schooling choice and e_i the observed employment status (with employed taking the value 1) in the data.²⁷

3.2 Empirical Results

Here we present the key empirical findings from our benchmark econometric model described in the previous section. We first discuss estimates of the measurement system mapping unobserved factors to observed BSAG maladjustment syndromes (Section 3.2.1). Next, we discuss the externalizing schooling penalty (Section 3.2.2) followed by the externalizing earnings premium (Section 3.2.3). Thereafter, we assess the role of intermediate choices and outcomes, such as occupation, in explaining our findings (Section 3.2.4).

²⁷Standard 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.

3.2.1 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 patterns hold for both males and females (Table C19 in Appendix C). The negative relationship between the two socio-emotional skills and cognition could reflect the distribution of skill endowments at birth. It could also reflect early childhood investments if the same environments that promote externalizing and internalizing behaviors also slow cognitive development (Heckman and Cunha, 2007). An example would be childhood poverty. The positive relationship 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). 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 Tables C20 and C21 in Appendix C, 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. 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).²⁸

²⁸It is possible that class size captures omitted school-level variables which positively affect student out-

3.2.2 The Externalizing Penalty in School

The marginal effect estimates of the multinomial logit model for educational attainment are reported in Table 5.²⁹ There is a significant negative relationship between externalizing behavior and educational attainment for boys. 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 may reflect how teachers are more likely to punish or refer for special help a male versus a female child 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. Also, the effect of family characteristics is consistent with our initial expectations. 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.

In general, estimates for the schooling model are broadly consistent with literature that studies the impact of emotional problems in school. One of the key pathways relating behavioral problems to low educational attainment is through early educational failures such as repeating a grade or falling behind in class. If externalizing or internalizing behavior make learning more difficult, this would in part be captured by the strong negative relationship between the two socio-emotional skills and cognition (which is identified from test scores) reported in Table C19 in Appendix C. However, the negative impact of the socio-emotional skills on education is not fully explained by these correlations, suggesting additional mechanisms. For example, McLeod and 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. This could make

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.

²⁹Standard errors for the marginal effects are calculated using the delta method.

schooling more costly.

3.2.3 The Externalizing Premium on the Labor Market

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 6.³⁰ 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.³¹ The evidence points to different ways that externalizing behavior increases earnings for males and females. It tends to raise wages for externalizing males, while it tends to increase labor supply on the intensive margin for externalizing females. The positive effects on hourly wages or weekly hours worked demonstrate that externalizing behavior is productive on the labor market even though it is counter-productive in school, especially for boys. This is a novel finding in the literature on the economic consequences of childhood misbehavior.

One possible explanation for the externalizing premium in the labor market is that externalizing behavior captures unobserved but correlated personality traits. Several studies have examined the relationship between externalizing and internalizing behaviors and better-known measures, such as the “Big 5” personality traits. 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).³² It is possible that high-externalizing individuals earn more for some of the same reasons that agreeable people earn less, such as a distaste for competition and negotiating. This point relates to earlier work that explores how economic

³⁰Selection into employment is discussed in the following section.

³¹Using our crude model, we considered an alternative specification where we control for hours worked in the wage equation. For males, the relationship between externalizing and wages increases slightly after we control for hours worked. For females, the relationship becomes negative and is insignificant. These results are available upon request.

³²To explain why, Barry and Friedman (1998) show that individuals with higher levels of agreeableness 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.

preferences relate to standard measures of socio-emotional skill (Becker et al., 2012). In robustness exercises, we test for this possibility. In Appendix A.6, we show that 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.³³

Internalizing behavior is negatively related to both productivity in the labor market 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.

Note that in the benchmark model (Column [2]), the labor outcome equations condition on the schooling choices. To evaluate whether including endogenous schooling choices affects the estimated effects of the unobserved skills in an intuitive way, in the same table, we also report estimates when we exclude the schooling outcomes from the outcome equations (Column [1]). Excluding schooling variables allows us to estimate the net impact of skills on labor market outcomes. Doing so increases the point estimates of the effect of cognition on hourly wages for both males and females. It also reduces the point estimate of the effect of externalizing on hourly wages for males, though only slightly. Since externalizing reduces schooling for males and schooling improves wages, it is not surprising that excluding schooling would generate a smaller net effect of externalizing on wages. What is notable is that the coefficient is still positive after including schooling, suggesting that more externalizing males earn higher wages despite the negative impact of externalizing on schooling.

³³One 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 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.

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.

3.2.4 Externalizing and Other Outcomes

To further examine why externalizing behavior increases earnings, we study its relationship to intermediate outcomes, such as fertility and marriage. The aim is to assess mechanisms underlying the externalizing premium. 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.

To explore these mechanisms, we assess how estimated coefficients change when we add endogenous intermediate outcome variables to the wage and hours equations. Results for hourly wage and weekly hours worked are reported in Tables D27 and D28 in Appendix D. We start from the benchmark model (Column [1]) and add a dummy variable for being married by age 33 (Column [2]) and reported number of children by age 33 (Column [3]).³⁴

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. Estimates are found in Table D29 in Appendix D. Externalizing males and females are both more likely to have a larger number of children by age 33, but based on the outcome equations (Tables D27 and D28 in Appendix D), having more children is somewhat irrelevant to earnings for males, but is associated with a large drop in both wages and hours for females. In Figure A2 in Appendix A, we show that female earnings are much lower for women with children in comparison to women without children. For males, there is no discernible relationship. Findings relating externalizing behavior, number of children and earnings suggest that the relatively low net impact of externalizing on women is attributable to two countervailing effects, which are (i) higher fertility, which lowers earnings, and (ii) better labor market outcomes. Finally, we add months of experience and occupational choices as controls (Columns [4] and [5]). Doing so does not appreciably

³⁴We keep the measurement system mapping latent skills to observed measurements of misbehavior as in the benchmark model.

alter the estimated relationship between externalizing behavior and labor market outcomes.

To visualize 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. Finally, we produce this figure separately for males and females to illustrate gender differences. 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.

A related concern is that the estimated externalizing behavior earnings premium could reflect selection into employment. For example, if high externalizing individuals dislike employment and thus select into employment only when they are highly productive due to unobserved factors, our estimates could be upwardly biased. In Appendix D.2, we perform several analyses that provide some evidence that selection into employment does not drive our estimates. For example, we show that our main results are largely unchanged if we impute earnings of individuals who are unemployed at age 33, but are employed in other waves, which drastically reduces the number of individuals with missing earnings. These results provide additional evidence that the positive labor market returns to externalizing behavior are not driven by differential sorting into employment.³⁵

In summary, though externalizing behavior is related to a host of economic outcomes that also predict earnings, positive labor market returns to externalizing behavior are not driven by differential sorting into these outcomes. This is further evidence that externalizing behavior generates higher earnings despite lowering educational attainment. High-externalizing males are either more productive or are otherwise able to secure higher payoffs for their labor, and high-externalizing females work more hours.

³⁵As an additional robustness check, we also experimented with a formal Heckman selection model for hourly wages using partnership and number of children as exclusion restrictions. We do not present these results since they suggest a similar story to the one presented in Appendix D.2 and because the exclusion restrictions are difficult to defend.

4 Sensitivity Analysis and Replication

This section discusses a host of sensitivity analyses, beginning with changes to assumptions on the measurement system. We also discuss replication of our main empirical findings in a variety of longitudinal data sets.

4.1 Sensitivity Analysis

4.1.1 Alternative Dedicated Measurements

Throughout the paper, we have assumed there are three unobservable skills, externalizing behavior, internalizing behavior and cognition, which are identified from measures of childhood classroom misbehavior and test scores. In the estimation of the measurement system that links the unobserved skills to the maladjustment syndromes and test scores, we designated one particular measure as a sole measurement of each skill. In this section, we discuss the implications of these assumptions. As explained in Section 3.1.2, we have chosen “hostility towards children” as the dedicated measurement for externalizing behavior, “depression” for internalizing behavior, and “verbal ability” for cognition. To assess sensitivity, we re-estimate the model iterating over all possible candidates for the dedicated measurements of the two socio-emotional skills.³⁶ 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 2.³⁷ The dashed bars indicate the results from the benchmark econometric model for males and females.

There are several points to note about the figure. While different dedicated measurement choices imply different magnitudes of the effects on education and earnings, externalizing behavior almost always has a significantly positive earnings premium for females and in a majority of cases has a significantly positive premium for males. Moreover, the benchmark specification is not the one that produces the largest externalizing earnings premium for males or females. Remarkably, in no specification do we find significant evidence against our main result from the benchmark model. The specifications under which the earnings premium becomes insignificant tend to be those in which withdrawal or unforthcomingness is chosen as the dedicated measure for internalizing behavior. As shown in Appendix E.1, in such cases, depression loads heavily on the “externalizing” factor. In this case, we identify a factor that is a mixture of what we typically regard as outwardly expressed externalizing behavior and

³⁶We re-estimate the model under all possible combinations of dedicated measurements as described in Table 1.

³⁷Additional findings using alternative dedicated factor assumptions are reported in Appendix E.1.

inwardly expressed internalizing behavior, and the impact of the “externalizing” factor on earnings is muted, which is expected given the negative correlation between depression and productivity. That said, the positive returns to externalizing behavior in the labor market do not require that depression be the dedicated measurement for internalizing behavior. For example, results are similar when the BSAG measure “writing off of adults and adult standards” is chosen as the dedicated measurement for internalizing behavior (see Tables E35 and E36 in Appendix E).

Findings from this exercise illustrate the fundamental identification problem in measuring underlying traits, discussed in Almlund et al. (2011). Creating a summary variable of measurements (as we did in our preliminary analysis) is simple, but implicitly imposes a number of unattractive assumptions. The measurement system in our benchmark model permits the relaxation of some assumptions, though a minimal set of assumptions, including which variable to use as a dedicated measurement, is still required for identification — and the analyst must choose which to use. A benefit of the measurement system is that such assumptions are explicit, and highlight the trade-off between letting the data guide the analysis versus imposing just enough structure to identify economically meaningful objects. In our case, it is possible to construct an externalizing factor that maps to depression and which has a substantially smaller effect on earnings due to the negative correlation between depression and earnings. However, doing so appears to contradict the standard interpretation of externalizing behavior. 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.

4.1.2 Testing the Three-Factor Assumption

The grouping of the factors into cognition, externalizing behavior and internalizing behavior has been previously validated in the literature, as described in Section 2. However, it is still possible that our results are influenced by additional factors that determine both choices and outcomes and that have been omitted in our analysis. We test for this possibility in two ways.

First, if an important fourth factor has been omitted, then the model with only three factors should make poor predictions on sample covariances between outcomes and choices. In Tables E37 and E38 in Appendix E, we present the simulated covariances between schooling levels and outcomes (i.e., wages and hours worked) against their sample counterparts, for males and females respectively. As is clear from the tables, our model with three factors has a good sample fit, suggesting that the benchmark model adequately accounts for the

observed relationship between choices and outcomes.

Second, we implement an extension of the benchmark model that allows for a fourth latent factor. Results are reported in Appendix E.2. In both male and female samples, we find that this fourth factor is insignificant in the schooling equation but does have some predictive power in the wages and hours equations. However, including the fourth factor does not affect the estimated relationship between externalizing behavior, schooling and earnings. In particular, even after controlling for the fourth factor, externalizing behavior still significantly reduces educational attainment and increases wages for males, while for females externalizing behavior continues to increase hours worked despite having no impact on schooling.

4.1.3 Imposing Independence of Factors

If one is willing to assume independence across factors, we can relax other assumptions on the measurement system. In particular, we can dispense with two of the three dedicated measures and let most measurements load on all three factors. In this scenario, identification of the three factors still requires one measurement dedicated to a single factor and a second dedicated to two of the three factors. We refer to Williams (2018) for a detailed discussion on the identification of linear latent factor models. We present results from an alternative model with independent factors in Appendix E.3. For this analysis, we chose verbal ability as a dedicated measure for cognition and depression as a semi-dedicated measure for internalizing behavior and cognition. Results under this specification resemble those from our benchmark model that permits correlation across factors.

4.1.4 Joint Estimation of the Measurement System, Choices and Outcomes

In our benchmark model, we estimated the model in two steps. We estimated the measurement system in a first step, and the educational choice and labor market outcome equations in a second step. In Appendix E.4 we present results where we estimate the measurement system jointly with choices and outcomes. Results remain largely unchanged. Heckman, Humphries, and Veramendi (2018) discuss the relative merits of a two-stage estimation and a joint estimation (see Appendix A.12 of their paper). The main motivation for us to pursue the two-stage estimation strategy in the benchmark model is that it makes interpreting the factors easier. In the two-stage estimation, the factors are solely identified from the measurement system, and can be clearly interpreted as underlying skills that account for the classroom misbehavior and tests measured at age 11.

4.1.5 Alternative Models for the Educational Choice

In the benchmark model, we adopt a very flexible model of schooling, a multinomial logit model with six schooling levels. One concern is that the model might be too flexible to pick up the negative effect of externalizing behavior on female schooling choices. In Appendix E.5, when we use more restrictive models such as a linear regression of years of schooling or a multinomial logit with four coarser schooling levels, we show that the results from these alternative specifications are largely consistent with the results from the benchmark model. That is, the externalizing behavior has a significant negative impact on schooling for boys, but the effect is insignificant for girls.

4.1.6 Alternative Set of Controls

One possible concern is that our results are sensitive to the controls used in the schooling and labor market outcome equations. To test that, in Appendix E.6, we report an alternative set of estimates where we modify the set of controls for labor market outcomes while keeping the controls for education constant. We show that the positive relationship between externalizing behavior and earnings emerges as soon as we control for internalizing behavior and cognition. Also, we report an alternative set of estimates where we assume the sets of controls Z and X are empty. We show that the negative relationship between externalizing behavior and educational attainment and the positive relationship between externalizing behavior and earnings remain in a model without any observable controls. While we continue to control for these variables in our benchmark model to eliminate potential biases, our main results are qualitatively robust to excluding them.

4.2 Robustness Across Data Sets

One possible concern is that our findings are specific to the Great Britain in the 1950s. To test that, we replicate our main analysis in a variety of 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 variable construction are found in Appendix F.

A concern when comparing estimates across data sets is that each dataset uses a different

scale to measure child behaviors and cognition. Thus, it is difficult to distinguish between differences in estimates arising from context-dependence versus differences due to how skills are measured. We cannot fully address this concern, but attempt to ensure that we rely on measures that have been validated in earlier research. Another concern is that there are important differences in parental and teachers' reports of children's behavior (Achenbach, McConaughy, and Howell, 1987; De Los Reyes and Kazdin, 2005; Ronda, 2017). To make sure the new measures are reasonably comparable to those in the NCDS, we rely on measures constructed from teachers' reports when possible. This is possible in 3 of the 4 replication data sets, since the CNLSY did not interview children's teachers. Finally, to facilitate comparison across data sets, we measure the education outcome by years of schooling.

In each dataset, we link the measure of externalizing behavior to schooling and earnings in a manner similar to the crude analysis described in Section 2.3. Results are summarized in Tables F61 and F62 in Appendix F. 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.³⁸ The point estimate of the impact of externalizing on earnings from the NCDS lies between estimates obtained from other data sets, suggesting that the externalizing earnings premium does not vary systematically across countries or over time.

In sum, we conclude that our main results using a 1958 British cohort — that childhood externalizing behavior negatively affects schooling while positively affecting labor market outcomes — are fairly consistent across data sets. Mixed returns of this particular socio-emotional skill appear to be an empirical regularity that is evident in at least two major economies and in data sets from several different time periods.

³⁸The 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 Appendix G).

5 Heterogeneity in Returns to Externalizing by Tasks

We showed that a socio-emotional skill, externalizing behavior, has mixed effects on two key economic outcomes: schooling and earnings. In this section, we further probe these empirical relationships to better understand the implications for human capital and its accumulation. We build on the idea that 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 Urzua, 2006). Testing for heterogeneity in the returns to skills at work is straightforward. We extend our labor market model to allow for the returns to skills to vary with occupational tasks using the O*NET task-intensity scales as in Acemoglu and Autor (2011). Ideally, we would perform a similar analysis for schooling, but we lack the data on school-specific tasks needed to do it. Thus, we examine different types of schools, comparing the externalizing schooling penalty at comprehensive versus private and grammar schools. As we explain below, these schools have different curricula, culture and practices. We speculate that such differences affect the tasks needed to successfully complete them.

5.1 Heterogeneous Skill Returns Across Occupational Tasks

We ask whether externalizing is valuable for work in general or if the premium depends on an individual's occupation. For example, externalizing individuals may be more energetic and thus be more productive in almost any occupation they choose. Alternatively, the skill may be productive on average, but less so (or even counterproductive) in some occupations. This leads us to specify and estimate a task model, where skills affect earnings through their impact on completing certain tasks, which vary in their intensity across occupations. The task-based framework is inspired by the model of comparative advantage of Almlund et al. (2011). It allows us to test whether returns to skills are heterogeneous, and possibly negative, across tasks. We outline a streamlined task-based framework. We show with a simple example that in the task-based framework, a single skill can have not just different but opposite returns in one occupational task than in another.

Suppose an individual i possesses three skills, externalizing (f_1), internalizing (f_2) and cognitive skill (f_3). In her job, she needs to complete two tasks, say an abstract/social task ($k = 1$) and a routine/manual task ($k = 2$). Let $T_{i,k}$ be her productivity in performing task k and it is determined by the three skills and the schooling level (s_i):

$$T_{i,k} = \tau_k(f_{i,1}, f_{i,2}, f_{i,3}, s_i), \quad k = 1, 2 \tag{22}$$

Her labor market earnings are determined by her productivity in each task, the intensity with which each task is required in her occupation of choice (Υ_i), and an individual productivity component (Ψ) that captures additional effects of skills and education that do not interact with the tasks intensity (e.g., preference for working long hours):

$$y_i = \Upsilon_i(T_{i,1}, T_{i,2}, \Psi(f_{i,1}, f_{i,2}, f_{i,3}, s_i)). \quad (23)$$

A special case of this general formulation is when we assume linear relationships in functions τ_k , Υ_i and Ψ in (22) and (23). In particular, let $\iota_{i,1}$ be the intensity of abstract/social tasks and $\iota_{i,2}$ the intensity of routine/manual tasks required in individual i 's occupation in Υ_i , then we can rewrite equation (23) as

$$\begin{aligned} y_i &= \iota_{i,1}T_{i,1} + \iota_{i,2}T_{i,2} + \psi_1f_{i,1} + \psi_2f_{i,2} + \psi_3f_{i,3} + \psi_s s_i + \psi_0 \\ &= \iota_{i,1} \left(\sum_{j=1}^3 \tau_{1,j}f_{i,j} + \tau_{1,s}s_i + \tau_{1,0} \right) + \iota_{i,2} \left(\sum_{j=1}^3 \tau_{2,j}f_{i,j} + \tau_{2,s}s_i + \tau_{2,0} \right) + \sum_{j=1}^3 \psi_j f_{i,j} + \psi_s s_i + \psi_0 \\ &= \sum_{j=1}^3 \psi_j f_{i,j} + \sum_{j=1}^3 \tau_{1,j} \cdot (\iota_{i,1} \times f_{i,j}) + \tau_{1,0} \cdot \iota_{i,1} \\ &\quad + \sum_{j=1}^3 \tau_{2,j} \cdot (\iota_{i,2} \times f_{i,j}) + \tau_{2,0} \cdot \iota_{i,2} + \alpha_s s_i + \psi_0 \end{aligned} \quad (24)$$

where $\alpha_s = \iota_{i,1}\tau_{1,s} + \iota_{i,2}\tau_{2,s} + \psi_s$.

Equation (24) highlights how the labor market returns of different skills will depend on the combination of tasks required in their chosen occupation. In particular, the return to skill j for individual i will depend on the skill's general productivity (ψ_j), its task productivity effects ($\tau_{1,j}$ and $\tau_{2,j}$), and on the task intensities at the individual's chosen occupation ($\iota_{i,1}$ and $\iota_{i,2}$).

The task intensities ($\iota_{i,1}$ and $\iota_{i,2}$) are observed in our data. We follow a similar methodology as in Acemoglu and Autor (2011) and Autor and Handel (2013) to construct the task intensity for each individual i in occupation k (ι_{ik}). The task intensities are composite measures of O*NET Work Activities and Work Context Importance scales.³⁹ 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

³⁹The 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.

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 two composite scales were constructed using factor analysis.

We extend our main econometric model by replacing the earnings outcome equation with equation (24) to estimate heterogeneous skill returns by tasks. The results from the extended model are found in Table 7. Since the task intensities ($\iota_{i,1}$ and $\iota_{i,2}$) are observed, the estimated interactions between skills and task intensities are measures of skills task productivity effects ($\tau_{k,j}$ s). A positive (negative) interaction between task k and skill j implies that skill j increases (decreases) individual productivity in performing task k .

Using this framework, we find evidence of heterogeneity in the returns to externalizing behavior and cognitive skills in the labor market. As in our main specification, on average, males and females face positive returns to externalizing behavior and cognition in the labor market ($\psi_1 > 0$ and $\psi_3 > 0$). The average individual also faces a negative return to the internalizing behavior ($\psi_2 < 0$). However, for males, the returns to externalizing behavior are smaller in occupations that are intensive in abstract and social tasks ($\tau_{1,1} < 0$) and larger in occupations that are intensive in manual and routine tasks ($\tau_{2,1} > 0$). These results imply that externalizing behavior is counterproductive for tasks that require large amounts of social interactions and productive for tasks that involve manual and routine activities. Also, we find that, for males, the returns to internalizing behavior are smaller in occupations that are intensive in manual and routine tasks ($\tau_{2,2} < 0$).

Despite this heterogeneity, it is important to note that the externalizing behavior labor market premium is predominantly positive. 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 of 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. Finally, it is notable that, for females, interactions with tasks are insignificant. This is not surprising since the positive effect of externalizing behavior on females’ earnings works mostly through longer hours. Thus, the results here suggest that a high-externalizing female tends to increase hours

regardless of her occupation.

5.2 Extending the Task-Based Framework to Schooling

We extend the idea that externalizing behavior is more or less productive for different tasks to our earlier results on schooling. As with earnings, it is possible that externalizing behavior is unproductive in school in general, perhaps because it captures a high disutility of learning new material. Alternatively, there may be heterogeneity depending on which tasks are emphasized in a school.

Lacking task data, however, it is not clear whether there is heterogeneity across schools in which tasks are emphasized. Nor is it clear how such differences would affect the returns of education on the labor market. We are able to make some progress on this question by leveraging differences in types of schools. The subjects in our sample who grew up in England and Wales experienced a major change in the state-funded secondary school system in their teens, when the tripartite education system was replaced by comprehensive schools throughout late 1960s and 1970s.⁴⁰ The reform led to four main types of schools: grammar schools, private schools, comprehensive schools, and secondary modern schools. We group grammar and private schools together to form the “non-comprehensive” school group (26% of our sample) and comprehensive and secondary modern schools together to form the “comprehensive” school group (74% of our sample). The non-comprehensive school group is selective on academic achievement with the aim of preparing students for higher learning. The comprehensive school group is non-selective and prepares students for a broader set of career paths, including higher education as well as vocational training.

When we split the sample by type of school individuals attended (comprehensive versus non-comprehensive), we find evidence of heterogeneous returns of externalizing behavior across these two types of schools (Table 8). More specifically, we extend our benchmark model to allow the educational attainment decision to depend on skills and their interactions with the dummy for comprehensive schools. Across school types, externalizing behavior increases the likelihood of a lower educational outcome. However, note the significant negative effect on the interaction term for the lowest educational outcome for both males and females who attended comprehensive schools. In other words, comprehensive schools appear

⁴⁰Under the tripartite educational system, pupils were allocated to one of three types of secondary schools according to their performance on an examination: grammar schools, secondary technical school and secondary modern school. This was the prevailing system under the Conservative governments between 1951 and 1964, but was actively discouraged by the Labour government after 1965. By 1976, the tripartite system was formally abolished in England and Wales, and replaced by the comprehensive system, in which students were not selected on the basis of academic achievement or aptitude. An excellent overview of the introduction of the comprehensive system is found in Harmon and Walker (2000).

to penalize externalizing behavior less than grammar and private schools for relatively early educational outcomes, i.e., those happening as secondary school ends.⁴¹⁴²

How do we interpret this finding? We show in Section 5.1 that externalizing behavior tends to increase an individual’s productivity in manual and routine tasks and decrease productivity in abstract and social tasks, especially for men. One possibility is that different types of schools emphasize different tasks. Non-comprehensive schools may emphasize abstract and social tasks while comprehensives may emphasize manual and routine tasks. Another possibility is that different types of schools emphasize another set of tasks that do not interact with externalizing behavior on the labor market, but do in school. Without carefully measuring tasks in schools, we cannot test these hypotheses. A third possibility, which does not exclude the other two, is that comprehensive schools penalize externalizing behavior less, but at the cost of offering a lower quality of education. In Appendix A.7, we compare the earnings of individuals that attended comprehensive versus grammar or private schools. On average, individuals that attended a comprehensive school do indeed earn significantly less than those that did not. However, this difference is mainly driven by selection into schooling. The earnings gap becomes insignificant once we control for family characteristics and cognition (the latter measured by test scores taken at age 11, so before matriculation into secondary school). This finding is consistent with prior literature, including Harmon and Walker (2000), who find that these school types do not have a direct effect on earnings. Our results provide suggestive evidence that it is possible to accommodate high-externalizing individuals in schools without harming the quality of education or the labor market outcomes of everyone else. This is policy relevant since the set of tasks that different types of schools emphasize is presumably related to school curricula, programming and climate, which are modifiable factors. This possibility would require further examination and data collection, and thus opens up avenues for future research. We return to this point in the Conclusion.

⁴¹The tolerance for externalizing behavior in comprehensive schools is documented by John-Paul Flintoff in his book, *Comp: A Survivor’s Tale*. In the book, Flintoff recounts his experience in Holland Park School, London’s flagship comprehensive school, as a boy with “posh voice and precocious liberal conscience” tried to blend in classmates who “would prefer to throw furniture out of the window or set your books on fire.”

⁴²In results available upon request, we exploit the change in the school type over a child’s adolescence, as is done in Harmon and Walker (2000), to estimate the effect of exposure to a comprehensive education for externalizing individuals. We estimate a reduced-form version of our structural model linking years of education to skills and different measures of exposure to comprehensive schools. In all models, we estimate a positive interaction between exposure to comprehensive education and externalizing behavior for schooling.

6 Implications

Economic research has broadened the notion of what constitutes human capital to include not only work experience, cognition or educational attainment, but also health and socio-emotional skills. Earlier work has also recognized that returns to skills have different prices across sectors. Despite potential differences in returns, however, the skills that constitute human capital have typically been seen as enhancing productivity across all domains. Our results challenge this view. A more complete understanding of human capital should accommodate the idea that some skills are not only less productive but actually *counter-productive* for some tasks and thus for some occupations or in certain economic contexts. This may not hold for every component of human capital. It is difficult to imagine a task where cognition or mechanical skill are actually counter-productive. The reason is that high cognition or mechanical skill can simply remain idle in contexts where they are not useful. However, our findings suggest the same is not generally true for other skills, including the socio-emotional skills we study, which can thus have both positive and negative returns.

Empirically, our findings highlight how context-dependence complicates the evaluation of skill. Estimated returns may not generalize beyond the specific domain in which they are measured. A better understanding of the tasks specific to each context could offer clues. For example, the hype surrounding *grit* as a crucial skill for success in life may have been an artifact of examining the skill in very specific contexts where it may be highly productive, such as test-taking, where persistence and goal-setting are useful (Duckworth et al., 2007; Borghans et al., 2008). However, newer evidence shows that grit is unproductive in settings where the cost of being too persistent can be high (Lucas et al., 2015), for example, for tasks that require recognizing and passing over difficult items. A more careful understanding of the tasks needed for success in the contexts in which grit was initially measured may have been helpful in interpreting estimated returns and understanding their limitations.

The task framework thus provides a useful model to better understand heterogeneity in the returns to skills. However, it need not be limited to studying occupations, which current data allow. With proper data collection, the task framework could be extended to understand different kinds of schools to better assess if different sets of tasks could successfully educate students and prepare them for the labor market while minimizing the externalizing penalty. The task framework could also be extended to other economic domains, such as parenting or interpersonal relationships. Individuals who are patient and compassionate may be less successful in running a business in a cutthroat industry, for example, but more successful in creating value as good parents. Finally, the task framework could be interacted with economic environments that can also affect how tasks align with skills, including political

institutions, legal systems, cultural norms and socio-demographic characteristics. Indeed, in Appendix G, we show suggestive evidence that the externalizing premium may not extend to children who grew up in poverty.⁴³ This provides further evidence that understanding context is crucial when evaluating different components of human capital.

7 Conclusion

Using several methods and data sets, we demonstrated that the same socio-emotional skill, externalizing behavior, can be productive in the labor market and counterproductive in school. Additional results show heterogeneity by tasks and across school types. We discussed several general implications of our findings for how we should evaluate different skills.

We are cautious in suggesting specific implications for policy. In particular, the externalizing premium we identify does not justify policies encouraging externalizing behavior. A primary reason is that there is little evidence on how to promote externalizing behavior in a way that would leave other skills unchanged. For example, it is possible that externalizing and internalizing behaviors are inextricably linked—indeed, our findings on the correlations between the two skills suggest this is a possibility. If so, attempts to modify one skill could modify the other in a way that harms children. Even if it were possible to modify externalizing behavior without modifying other skills, it is unclear if doing so would increase welfare given the possibility of negative spillover effects in the classroom if externalizing children are disruptive and limit other students’ learning (Henneberger, Coffman, and Gest, 2016). Recall, our results show that externalizing behavior loads heavily onto the maladjustment syndrome “hostility towards children.” Given documented negative impacts of bullying on education, policies increasing hostility among schoolchildren are likely to be unproductive (Brown and Taylor, 2008; Carrell and Hoekstra, 2010; Carrell, Hoekstra, and Kuka, 2018).

An alternative is to develop policies that accommodate externalizing behavior rather than penalizing it. The idea is to shift the price of a skill in certain contexts rather than to modify the skill itself. We provide suggestive evidence that this is possible and could have happened to some extent in the 1960s and 1970s in England and Wales. However, our analysis of school types is suggestive and preliminary. Today, there is wide variation in the types of schools children can attend, including charter, Montessori, magnet, military, and religious schools, not to mention increasingly-relevant online options or home-schooling. Different schooling options often have drastically different teaching philosophies that likely emphasize

⁴³This finding is consistent with Heckman, Pinto, and Savelyev (2013). However, we do not present these results because we cannot reject that returns to externalizing are not statistically different across groups, which may be due to small sample sizes.

different sets of tasks that favor different types of students. An important question is which types of schools effectively prepare students for the labor market while also accommodating skills, such as externalizing behavior. To address this question, future research could extend task data collection to schools to study how returns to externalizing behavior vary across school types. Similar task data collection could also be applied to other important economic contexts (e.g., parenting) and to other components of human capital to understand the precise mechanisms through which they affect individuals over the lifecycle.

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., 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.
- Agan, Amanda Y. 2011. "Non-Cognitive Skills and Crime." University of Chicago.
- Aizer, Anna. 2009. "Neighborhood Violence and Urban Youth." In *The Problems of Disadvantaged Youth: An Economic Perspective*, edited by Jonanthan Gruber. University of Chicago Press, 275–308.
- 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.
- Barry, Bruce and Raymond A Friedman. 1998. "Bargainer Characteristics in Distributive and Integrative Negotiation." *Journal of Personality and Social Psychology* 74 (2):345.
- Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4 (1):453–478.
- Becker, Gary Stanley. 1967. *Human Capital and the Personal Distribution of Income: An Analytical Approach*. Institute of Public Administration.
- 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.
- Blanden, Jo, Paul Gregg, and Lindsey Macmillan. 2007. "Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education." *Economic Journal* 117 (519):C43–C60.
- 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.
- Brown, Sarah and Karl Taylor. 2008. "Bullying, Education and Earnings: Evidence from the National Child Development Study." *Economics of Education Review* 27 (4):387–401.
- Campbell, Susan B, Daniel S Shaw, and Miles Gilliom. 2000. "Early Externalizing Behavior Problems: Tod-

- dlers and Preschoolers at Risk for Later Maladjustment.” *Development and Psychopathology* 12 (3):467–488.
- Card, David. 2012. “Earnings, Schooling, and Ability Revisited.” In *35th Anniversary Retrospective (Research in Labor Economics, Volume 35)*, edited by Solomon W. Polachek and Konstantinos Tatsiramos. Emerald Group Publishing Limited, 111–136.
- Carrell, Scott E, Mark Hoekstra, and Elira Kuka. 2018. “The Long-Run Effects of Disruptive Peers.” *American Economic Review* 108 (11):3377–3415.
- Carrell, Scott E and Mark L Hoekstra. 2010. “Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone’s Kids.” *American Economic Journal: Applied Economics* 2 (1):211–228.
- Chase-Lansdale, P Lindsay, Andrew J Cherlin, and Kathleen E Kiernan. 1995. “The Long-term Effects of Parental Divorce on the Mental Health of Young Adults: A Developmental Perspective.” *Child Development* 66 (6):1614–1634.
- 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.
- Duckworth, Angela L, Christopher Peterson, Michael D Matthews, and Dennis R Kelly. 2007. “Grit: Perseverance and Passion for Long-Term Goals.” *Journal of Personality and Social Psychology* 92 (6):1087.
- 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.
- Farmer, Elizabeth MZ. 1993. “Externalizing Behavior in the Life Course The Transition From School to Work.” *Journal of Emotional and Behavioral Disorders* 1 (3):179–188.
- . 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.
- Fronstin, Paul, David H Greenberg, and Philip K Robins. 2005. “The Labor Market Consequences of Childhood Maladjustment.” *Social Science Quarterly* 86 (s1):1170–1195.
- Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M Chang, and Sally Grantham-McGregor. 2013. “Labor Market Returns to Early Childhood Stimulation: a 20-year Followup to an Experimental Intervention in Jamaica.” Tech. rep., National Bureau of Economic Research.

- 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.
- Gibbons, Robert and Michael Waldman. 2004. "Task-specific Human Capital." *American Economic Review* 94 (2):203–207.
- Gregory, Mary K. 1977. "Sex Bias in School Referrals." *Journal of School Psychology* 15 (1):5–8.
- 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.
- Harmon, Colm and Ian Walker. 2000. "The Returns to the Quantity and Quality of Education: Evidence for Men in England and Wales." *Economica* 67 (265):19–35.
- Heckman, James J and Flavio Cunha. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2):31–47.
- Heckman, James J., John Eric Humphries, and Gregory Veramendi. 2018. "Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking." *Journal of Political Economy* 126 (S1):197–246.
- 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 Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *The Journal of Labor Economics* 24 (3):411–482.
- Henneberger, Angela K, Donna L Coffman, and Scott D Gest. 2016. "The Effect of Having Aggressive Friends on Aggressive Behavior in Childhood: Using Propensity Scores to Strengthen Causal Inference." *Social Development* 26 (2):295–309.
- Jackson, Michelle. 2006. "Personality Traits and Occupational Attainment." *European Sociological Review* 22 (2):187–199.
- 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.
- Kautz, Tim, James J Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. 2014. "Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success." Tech. rep., National Bureau of Economic Research.
- Kendler, Kenneth S. and John Myers. 2014. "The Boundaries of the Internalizing and Externalizing Genetic Spectra in Men and Women." *Psychological 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.
- . 2018. "Selection into Entrepreneurship and Self-Employment." Tech. rep., National Bureau of Economic Research.
- Lucas, Gale M, Jonathan Gratch, Lin Cheng, and Stacy Marsella. 2015. "When the Going Gets Tough: Grit Predicts Costly Perseverance." *Journal of Research in Personality* 59:15–22.
- Lundberg, Shelly. 2013. "The College Type: Personality and Educational Inequality." *Journal of Labor Economics* 31 (3):421–441.
- 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.
- Neidell, Matthew and Jane Waldfogel. 2010. "Cognitive and Noncognitive Peer Effects in Early Education." *The Review of Economics and Statistics* 92 (3):562–576.
- Prada, María F and Sergio Urzúa. 2017. "One Size Does Not Fit All: Multiple Dimensions of Ability, College Attendance, and Earnings." *Journal of Labor Economics* 35 (4):953–991.
- Regier, Darrel A., Emily A. Kuhl, and David J. Kupfer. 2013. "The DSM-5: Classification and Criteria Changes." *World Psychiatry* 12 (2):92–98.
- Ronda, Victor. 2017. "When Mothers and Teachers Disagree: Observer Reports and Children's Noncognitive Skills." Mimeo, Johns Hopkins University.
- 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.
- 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.
- Stott, Denis Herbert, Emily G Sykes, and NC Marston. 1974. *The Social Adjustment of Children: Manual to the Bristol Social Adjustment Guides*. Hodder & Stoughton.
- Williams, Benjamin. 2018. "Identification of the Linear Factor Model." *forthcoming at Econometrics Reviews*.
- 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.
- Yeung, W Jean, Miriam R Linver, and Jeanne Brooks-Gunn. 2002. "How Money Matters for Young Children's Development: Parental Investment and Family Processes." *Child Development* 73 (6):1861–1879.

8 Tables and Figures

Table 1: MEASUREMENTS USED TO IDENTIFY UNOBSERVED SKILLS

Unobserved Skill	Measures
Externalizing Behavior	◊ Hostility Towards Adults
	◊ Hostility Towards Children
	◊ Anxiety for Acceptance by Adults
	◊ Anxiety for Acceptance by Children
	◊ Restlessness
	◊ Inconsequential Behavior
	◊ Writing Off of Adults and Adult Standards
Internalizing Behavior	◊ Depression
	◊ Withdrawal
	◊ Unforthcomingness
	◊ Writing Off of Adults and Adult Standards
Cognition	◊ Reading Comprehension Test Score
	◊ Mathematics Test Score
	◊ Non Verbal Score on General Ability Test
	◊ Verbal Score on General Ability Test

Notes: This table lists the three unobserved skills used in the empirical analysis (externalizing behavior, internalizing behavior and cognition) and the observed variables used to identify them. Measures for externalizing and internalizing behaviors are drawn from the BSAG maladjustment variables derived from teachers' reports of misbehavior. For cognition, a series of aptitude test scores are used as measures. See Section 2.2 for further details.

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: PRELIMINARY ANALYSIS: EDUCATIONAL ATTAINMENT

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Misbehavior	-0.316 (0.013)	-0.119 (0.014)							
Externalizing			-0.277 (0.014)	-0.105 (0.014)	-0.197 (0.015)	-0.077 (0.016)	-0.083 (0.016)	-0.105 (0.019)	-0.050 (0.026)
Internalizing					-0.164 (0.015)	-0.060 (0.016)	-0.056 (0.016)	-0.036 (0.020)	-0.082 (0.025)
Cognition		0.748 (0.016)		0.757 (0.015)		0.748 (0.016)	0.673 (0.016)	0.652 (0.022)	0.704 (0.024)
Family Controls	()	()	()	()	()	()	(X)	(X)	(X)
Obs.	7241	7241	7241	7241	7241	7241	7241	3573	3668

Notes: This table contains parameter estimates from an ordered probit model used to link unobserved skills to educational attainment. We estimate the ordered probability of choosing one of six schooling levels on a set of observable variables along with crude measures of unobserved skills. To construct the crude measures of the three unobserved skills, we sum up all variables used to measure that skill according to Table 1 and then normalize each unobserved skill to have mean zero and standard deviation one. Models [1]-[7] include all individuals and a gender dummy, Model [8] includes only males and Model [9] only females. Standard errors in parentheses.

Table 4: PRELIMINARY ANALYSIS: LOG WEEKLY EARNINGS

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Misbehavior	-0.089 (0.008)	-0.024 (0.009)							
Externalizing			-0.059 (0.008)	-0.013 (0.009)	-0.000 (0.008)	0.025 (0.009)	0.032 (0.009)	0.020 (0.008)	0.041 (0.020)
Internalizing				-0.091 (0.010)		-0.055 (0.009)	-0.047 (0.009)	-0.055 (0.008)	-0.031 (0.020)
Cognition		0.189 (0.009)			0.197 (0.009)	0.189 (0.009)	0.079 (0.010)	0.067 (0.009)	0.103 (0.021)
Education	()	()	()	()	()	()	(X)	(X)	(X)
Obs.	4888	4888	4888	4888	4888	4888	4888	2643	2245

Notes: This table contains parameter estimates from OLS regressions used to link socio-emotional and cognitive skills to earnings. We regress log earnings of workers on a set of observable variables along with crude measures of unobserved skills. To construct the crude measures of the three unobserved skills, we sum up all variables used to measure that skill according to Table 1 and then normalize each unobserved skill to have mean zero and standard deviation one. Models [1]-[7] include all individuals and a gender dummy, Model [8] includes only males and Model [9] only females. Standard errors in parentheses.

Table 5: EDUCATION ATTAINMENT, MARGINAL EFFECTS

	Males					
	No Qual.	CSE	O-lvl	A-lvl	H.Edu	H.Deg
Externalizing Behavior	0.014 (0.003)	0.021 (0.010)	0.004 (0.019)	-0.006 (0.017)	-0.022 (0.015)	-0.011 (0.007)
Internalizing Behavior	0.003 (0.004)	0.004 (0.011)	0.009 (0.021)	-0.005 (0.019)	-0.008 (0.017)	-0.003 (0.008)
Cognition	-0.063 (0.005)	-0.075 (0.008)	-0.064 (0.014)	0.051 (0.012)	0.069 (0.010)	0.081 (0.007)
	Females					
	No Qual.	CSE	O-lvl	A-lvl	H.Edu	H.Deg
Externalizing Behavior	0.003 (0.003)	-0.025 (0.012)	0.013 (0.021)	0.006 (0.013)	0.009 (0.016)	-0.005 (0.006)
Internalizing Behavior	0.014 (0.004)	0.034 (0.012)	-0.015 (0.021)	-0.015 (0.014)	-0.019 (0.017)	0.002 (0.007)
Cognition	-0.071 (0.006)	-0.102 (0.009)	0.003 (0.015)	0.055 (0.009)	0.057 (0.011)	0.058 (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. For the full set of parameter estimates, see Table C22 in Appendix C.

Table 6: 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 (0.018)	0.064 (0.017)	0.023 (0.024)	0.013 (0.020)	0.012 (0.008)	0.015 (0.008)	0.047 (0.024)	0.047 (0.025)
Internalizing Behavior	-0.099 (0.020)	-0.096 (0.018)	-0.043 (0.027)	-0.021 (0.022)	-0.014 (0.010)	-0.018 (0.009)	-0.023 (0.027)	-0.020 (0.026)
Cognition	0.106 (0.011)	0.025 (0.011)	0.163 (0.015)	0.044 (0.013)	-0.015 (0.005)	-0.007 (0.006)	0.078 (0.016)	0.021 (0.017)
Educational Attainment	()	(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 C23 and C24 in Appendix C.

Table 7: LOG WEEKLY EARNINGS AND OCCUPATIONAL TASKS

	[Males]		[Females]	
	(1)	(2)	(3)	(4)
Externalizing Behavior	0.069 (0.018)	0.059 (0.016)	0.087 (0.028)	0.077 (0.029)
Internalizing Behavior	-0.104 (0.020)	-0.090 (0.018)	-0.063 (0.032)	-0.049 (0.035)
Cognition	0.019 (0.012)	0.015 (0.011)	0.028 (0.019)	0.050 (0.020)
Ext. x Abstract		-0.026 (0.015)		0.003 (0.038)
Int. x Abstract		0.011 (0.015)		-0.023 (0.042)
Cog. x Abstract		-0.004 (0.010)		-0.017 (0.023)
Abstract Intensity		0.045 (0.027)		0.114 (0.060)
Ext. x Routine	0.036 (0.015)		-0.012 (0.031)	
Int. x Routine	-0.033 (0.015)		0.018 (0.034)	
Cog. x Routine	-0.006 (0.010)		0.012 (0.021)	
Routine Intensity	0.041 (0.020)		0.196 (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 C25 in Appendix C

Table 8: EDUCATION ATTAINMENT BY SCHOOL TYPE, MARGINAL EFFECTS

	Males					
	No Qual.	CSE	O-lvl	A-lvl	H.Edu	H.Deg
Externalizing Behavior	0.012 (0.003)	0.008 (0.010)	0.047 (0.021)	0.010 (0.018)	-0.060 (0.030)	-0.017 (0.015)
Internalizing Behavior	0.006 (0.002)	0.023 (0.015)	-0.019 (0.031)	0.006 (0.023)	-0.002 (0.032)	-0.014 (0.015)
Cognition	-0.047 (0.012)	-0.065 (0.033)	-0.085 (0.034)	0.028 (0.021)	0.059 (0.030)	0.109 (0.041)
Externalizing \times Comprehensive	-0.006 (0.002)	0.004 (0.014)	-0.039 (0.024)	-0.007 (0.021)	0.048 (0.035)	-0.000 (0.013)
Internalizing \times Comprehensive	-0.002 (0.003)	-0.018 (0.016)	0.035 (0.035)	-0.013 (0.027)	-0.024 (0.039)	0.021 (0.020)
Cognition \times Comprehensive	0.008 (0.005)	-0.007 (0.014)	-0.015 (0.028)	-0.011 (0.021)	0.040 (0.035)	-0.015 (0.010)
	Females					
	No Qual.	CSE	O-lvl	A-lvl	H.Edu	H.Deg
Externalizing Behavior	0.002 (0.001)	-0.010 (0.016)	0.067 (0.024)	-0.024 (0.022)	-0.008 (0.028)	-0.028 (0.020)
Internalizing Behavior	0.011 (0.003)	0.023 (0.017)	-0.049 (0.036)	0.019 (0.020)	-0.021 (0.034)	0.017 (0.013)
Cognition	-0.032 (0.010)	-0.075 (0.037)	-0.114 (0.040)	0.071 (0.027)	0.049 (0.032)	0.101 (0.049)
Externalizing \times Comprehensive	-0.005 (0.001)	-0.015 (0.011)	-0.062 (0.028)	0.028 (0.027)	0.025 (0.033)	0.028 (0.024)
Internalizing \times Comprehensive	-0.003 (0.002)	0.010 (0.018)	0.040 (0.039)	-0.032 (0.025)	0.006 (0.038)	-0.022 (0.017)
Cognition \times Comprehensive	0.004 (0.003)	-0.006 (0.017)	0.064 (0.033)	-0.033 (0.017)	0.002 (0.032)	-0.031 (0.014)

Notes: This table lists marginal effects estimates from a multinomial logit model used to link socio-emotional and cognitive skills to educational attainment across school types. We estimate educational attainment on a set of observable variables along with the unobserved skills and their interaction with the comprehensive school dummy. The comprehensive group includes children enrolled in comprehensive or secondary modern secondary education. 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 C26 in Appendix C

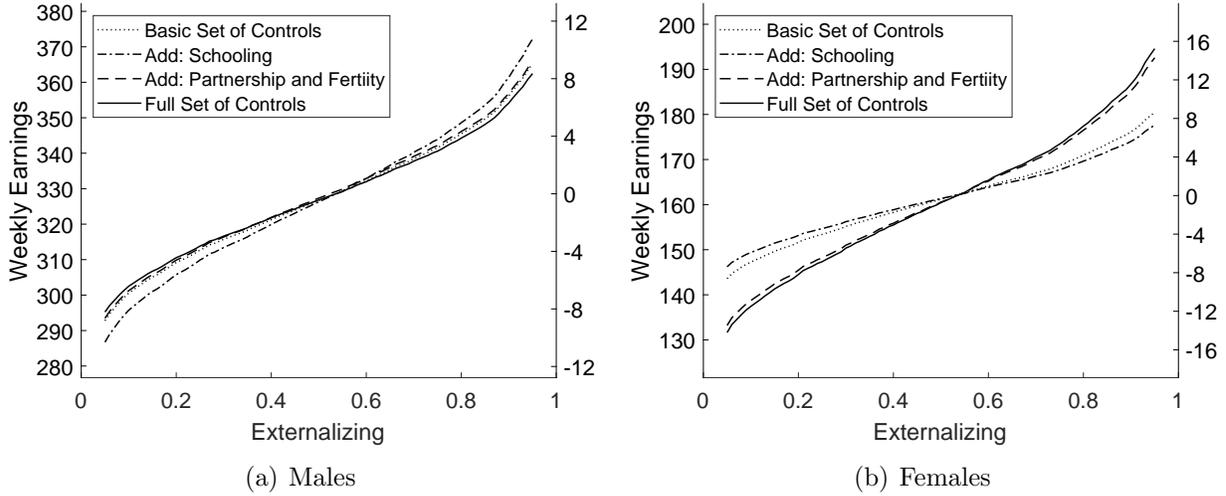


Figure 1: DECOMPOSITION OF EFFECTS OF EXTERNALIZING ON WEEKLY EARNINGS: Figure 1 visualizes the results from regressing weekly earnings on a varying set of controls presented in Tables D27 and D28. It illustrates how the predicted weekly earnings in regression models with different sets of controls vary, 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.

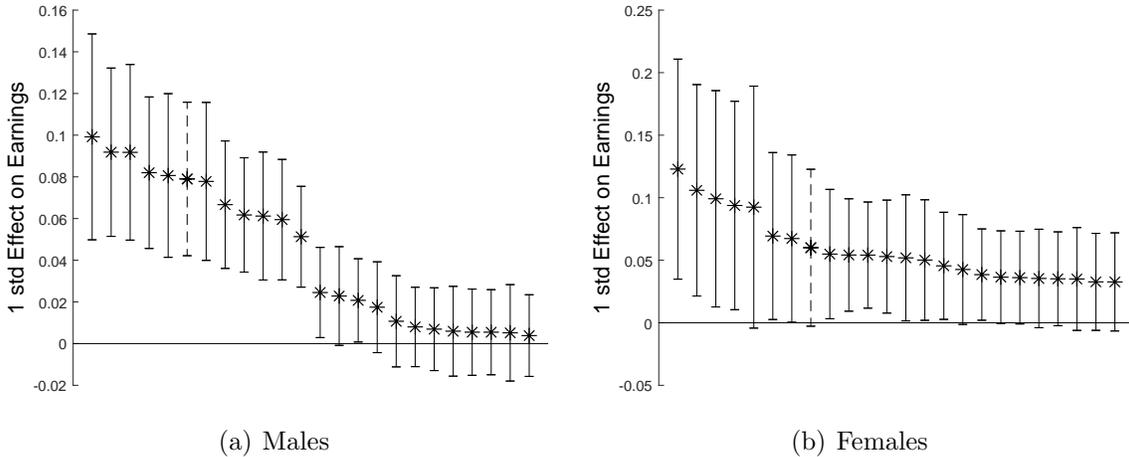


Figure 2: DISTRIBUTION OF EFFECTS OF EXTERNALIZING ON EARNINGS: Figure 2 visualizes the effects on weekly earnings from 1 standard deviation increase in externalizing behavior from specifications that span all possible combinations of the dedicated measurements for externalizing and internalizing behaviors. It summarizes the results reported in Tables E35 and E36. The dashed bars indicate results from our benchmark model.