

Skill formation and the trouble with child noncognitive skill measures

EMILIA DEL BONO
ISER, University of Essex

JOSH KINSLER
Department of Economics, University of Georgia

RONNI PAVAN
Department of Economics, University of Rochester

This paper develops a framework to address issues of contamination in parent-reported measures of child noncognitive skills. We estimate a dynamic model in which child and parental skills evolve jointly and leverage information provided by teachers and interviewers to deal with contamination of parent-reported measures. The model also allows us to examine the relative importance of mothers and fathers in the evolution of child skills. Our findings reveal that ignoring contamination significantly underestimates the role of maternal non-cognitive skills in the evolution of child noncognitive skills. Additionally, we find evidence of stronger feedback effects from child skills to mothers than fathers. Simulation exercises demonstrate how contamination can distort evaluations of early childhood policies, underscoring the importance of robust measurement approaches.

KEYWORDS. Children, human capital, dynamic factor analysis, measurement, policy.

JEL CLASSIFICATION. C13, C18, I38, J13, J24.

1. INTRODUCTION

Many studies have shown that human capital skills developed in early childhood are strong predictors of later life outcomes such as educational attainment, earnings, and health (Todd and Wolpin (2007), Cunha, Heckman, Lochner, and Masterov (2006), Mof-

Emilia Del Bono: edelbono@essex.ac.uk

Josh Kinsler: jkinsler@uga.edu

Ronni Pavan: ronni.pavan@rochester.edu

We are grateful for the comments received by Francesco Agostinelli, Esteban Aucejo, and Sonia Bhalotra. We would like to thank seminar participants at Arizona State University, the University of Essex, the University of Oxford, the University of Pennsylvania, the University of Oklahoma, Clemson University, and conference participants at EALE, the Royal Economic Society, AASLE, and SOLE. We acknowledge funding from the National Science Foundation, award 1725270. This research was also funded in part by the UK Economic and Social Research Council through the Centre on Micro Social Change (MiSoC) [ES/S012486]. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission.

fitt, Arseneault, Belsky, and Caspi (2011), Conti, Heckman, and Pinto (2016)). As a consequence, the success of publicly funded policies such as universal child care, Head Start (for the US) or Sure Start (for the UK) are often measured according to their effects on children's skills. Recently, there has been a growing interest in the impact such policies have on a particular aspect of child human capital, noncognitive skills (Deming (2009), Baker and Milligan (2015)). Noncognitive skills aid in the development of cognitive skills throughout early childhood and directly impact labor market outcomes (Cunha, Heckman, and Schennach (2010), ACFM+ (2020), Heckman, Stixrud, and Urzua (2006)). Yet, how do we measure child noncognitive skills, such as emotional stability, motivation, and self-regulation, when children are too young to understand complex questions and provide accurate reports of their attitudes and behaviors?

To date, much of the analysis regarding child noncognitive skill formation relies on surveys where parents assess the emotional and social behavior of their child.¹ These assessments are naturally subjective and can reflect not only the noncognitive skills of the child but also the skills and traits of the parents, such as their well-being or mental health. The trouble with relying on parent-reported measures of child skills is that parental skills are a direct input in the production function of child skills, and estimating their impact is challenging when child noncognitive skill measures are contaminated. The problem becomes even more complex in a dynamic setting, where parental skills are time varying and can evolve in response to child skill development.

In this paper, we estimate a dynamic model of child skill formation that accounts for measurement contamination and the joint evolution of child and parent skills. Building on the framework of Cunha, Heckman, and Schennach (2010), we introduce three main innovations. First, we allow parent-reported measures of child noncognitive skill to be contaminated by parental skills. Second, we explicitly model the joint formation and evolution of child and parent skills over time, capturing feedback effects between them. Third, we distinguish between maternal and paternal contributions, enabling us to assess the relative roles of mothers and fathers in shaping child development.

We use the model to address two key research questions. First, we ask how estimates of child skill dynamics differ when measurement contamination is ignored. We show that standard models that do not address this issue misrepresent the persistence and drivers of noncognitive skill development. Second, we examine the consequences of relying on contaminated measures in reduced-form policy evaluations. Even when treatment is exogenous and appropriate causal methods are applied, estimates of the effects of policies like universal child care on child noncognitive skills can be biased if contaminated measures are used as outcomes. Our findings highlight the importance of accounting for contamination when employing structural and reduced-form models to study early childhood skill development.

¹This is true for the US, where the most commonly used data for this type of analysis is the National Longitudinal Mother–Child Supplement (Deming (2009), Cunha, Heckman, and Schennach (2010)), as well as the UK, where researchers (Del Bono, Francesconi, Kelly, and Sacker (2016), Hernández-Alava and Popli (2017)) have mainly relied on the Millennium Cohort Study (University College London, UCL Social Research Institute, Centre for Longitudinal Studies (2024)).

We use data from the Millennium Cohort Study (MCS), a large prospective study of infants born between 2000 and 2002 in the United Kingdom. Successive interviews took place every few years and include cognitive and noncognitive assessments of child development. The data offers important features that we leverage to identify the effects of interest. Specifically, the noncognitive skills of the child are evaluated by parents at every wave, but also by interviewers or teachers such that we have measures from different evaluators at every age of the child. An additional attractive feature of the MCS is the availability of parental skills, and specifically parental mental health at each wave.² Finally, data on child noncognitive skills is usually provided by mothers, but for a small proportion of the sample fathers report on child skills, which offers the opportunity to examine the role of fathers in child skill formation.

We begin with descriptive analysis highlighting the subjective nature of child noncognitive skill measures. We compare the correlation between mother-reported child noncognitive skills and maternal noncognitive skills with the correlation between teacher (or interviewer)-reported child noncognitive skills and maternal noncognitive skills. We show that maternal noncognitive skill measures are highly correlated with child noncognitive skill measures only when using child noncognitive skill measures *reported by the mother*. Additionally, low non-cognitive skill mothers evaluate their children more negatively relative to teachers than high noncognitive skill mothers. When the father is the main survey respondent, it is the father's noncognitive skill measures that best predict child noncognitive skills. These patterns suggest that the identity and attributes of the child's evaluator can impact measures of child noncognitive skills.

To fully address issues of contamination and account more generally for the importance of parental skills in child skill formation, we formulate a joint model of child and parent skills production in a dynamic setting. A key innovation is to relax the assumption that there is no contemporaneous correlation in child skill measures other than through child skills. In particular, we allow parent, teacher, and interviewer-reported measures of child noncognitive skills to be contaminated. This contamination can be viewed as a component of skill measurement error that is potentially correlated across noncognitive skill measures and, in the case of parents, with the measures of other skills. Our model is identified under the assumption that contamination in the child noncognitive skill measures reported by different evaluators are independent of each other. We provide supportive evidence for this assumption and investigate the consequences of plausible violations.

We estimate our model using a two-step approach similar to [Agostinelli and Wiswall \(2025\)](#), [ACFM+ \(2020\)](#), and [Attanasio, Meghir, and Nix \(2020\)](#). The results show that a 1 standard deviation increase in maternal noncognitive skills when the child is age 3 improves noncognitive skills at age 11 by a statistically significant 0.1 standard deviations. Paternal noncognitive skills do not have a significant impact on child skills development, but paternal cognitive skills have a statistically significant positive effect on child cognitive and noncognitive skills development. Our model also captures the joint evolution of parent and child skills, allowing for bidirectional influences. In contrast to

²Parental mental health conditions may affect the reporting of children's mental health status and other child noncognitive skills ([NWNS+ \(2000\)](#), [Sawyer, Streiner, and Baghurst \(1998\)](#)). Here, we think of parental mental health as a specific type of parental noncognitive skill relevant to the development of the child.

previous studies, we allow parental noncognitive skills to evolve in response to child development. We find that a one standard deviation increase in a child's noncognitive skill at age 3 leads to a 0.12 standard deviation increase in maternal noncognitive skill when the child is age 11. The effect on paternal skills is also positive but approximately half as large.³

To address our first research question—how ignoring measurement contamination biases estimates of child skill dynamics—we compare our results to those from a conventional model that uses only parent-reported measures of child noncognitive skill and assumes the measures are uncontaminated. In the parent-only model, the estimated effect of maternal noncognitive skill on child non-cognitive skill is substantially smaller and not statistically significant. Also, the self-productivity of child noncognitive skills is significantly larger. These patterns arise because the contaminated parent-reported measures of child non-cognitive skill absorb variation attributable to maternal traits, biasing the estimated effects of maternal noncognitive skills and the self-productivity of child-noncognitive skills.

To address our second research question—how contamination can bias reduced-form estimates of early childhood policies—we use our model to simulate latent skills and observed measures consistent with a universal child care intervention (Haeck, Lebihan, Lefebvre, and Merrigan (2022), Yamaguchi, Asai, and Kambayashi (2018), Baker, Gruber, and Milligan (2008)). In a first simulation, we assume that universal child care only leads to a one standard deviation increase in maternal noncognitive skill at baseline. We find that *measured* child noncognitive skill rises by up to 0.15 standard deviations in the initial period, an effect that arises entirely from contamination, as child skills are not directly affected by the policy. Over time, the increase in maternal skills influences child development, but even after 8 years, roughly half of the observed increase in child noncognitive skill measures is due to contamination rather than true skill gains. In a second exercise, we simulate a policy that directly increases both maternal and child noncognitive skills by 0.2 standard deviations, consistent with estimates from Baker, Gruber, and Milligan (2008), who find that expanded access to free child care in Quebec affected both outcomes (negatively). Even in this case, we find that up to one-quarter of the observed increase in *measured* child noncognitive skill can be attributed to contamination.

Our paper is the first to propose a methodology to tackle systematic error in measures of child noncognitive skills in a dynamic model of skills formation. The existing literature in economics has largely ignored this issue (Cunha, Heckman, and Schennach (2010), ACFM+ (2020)). One explanation for this are data constraints, since dealing with the contamination of parent-reported measures of child noncognitive skills requires the availability of measures from multiple evaluators. While this type of data is frequently collected and analyzed in the psychological and psychometric literature on child development and is increasingly preferred to “single-informant” data (KMAE+ (2003), DL-RAWT+ (2015), Martel, Markon, and Smith (2017)), it is less readily available in large and

³Using data from Pakistan, SBBM+ (2024) estimate a reduced-form model in which maternal mental health measured 12 months (36 months) since the birth of the child is a function of the child socioemotional and physical skills at 6 months (12 months), but do not find significant impacts. It is possible that the interaction between child and maternal skills manifests more clearly when children are older.

representative surveys. One notable exception is the work conducted by [Johnston, Proper, Pudney, and Shields \(2014\)](#), who use the 2004 Survey of Mental Health of Children and Young People in Britain to examine the effects of child mental health on education. The survey provides measures of child mental health reported by parents, teachers, and the children themselves, who are all assumed to be biased informants. The setting is static and identification relies on diagnostic assessments from a panel of psychiatric experts, who are assumed to be unaffected by systematic bias.⁴

The idea of relying on multiple evaluators or multiple measurement methods to purge measures of contamination instead has a long history in the broader psychometric and applied statistics fields ([Campbell and Fiske \(1959\)](#), [Joreskog \(1971\)](#)). Using multiple evaluators can reduce what is known as common source bias, which might arise from rater-specific effects or self-report bias ([Podsakoff, MacKenzie, Lee, and Podsakoff \(2003\)](#)). Our main contribution to this literature is to allow for the bias to be a function of the evaluator's latent traits.

Our study also adds to a well-established literature that explores the impacts of maternal mental health on children's outcomes. A substantial part of this literature investigates the effects of maternal stress in pregnancy or during the post-natal period on child development using either natural experiments—which create exogenous variation in maternal psychological well-being ([Black, Devereux, and Salvanes \(2016\)](#), [Persson and Rossin-Slater \(2018\)](#))—or randomized control trials ([Baranov, Bhalotra, Biroli, and Maselko \(2020\)](#)). By modeling the joint dynamics of child and parental skills, our setup is closer to [Ronda \(2017\)](#), who focuses only on the effect of maternal mental health on child cognitive skills, and [ACFM+ \(2020\)](#) who also consider effects on children's non-cognitive skills. None of these papers however addresses the problem of contamination of parental reports of child noncognitive skills measures on which we focus here, nor discusses explicitly the impact of child skills on parental skills, a link that has been emphasized mostly in the health literature ([Kvist, Nielsen, and Simonsen \(2013\)](#), [Kuhn and Carter \(2006\)](#), [Davis and Carter \(2008\)](#), [Choe, Olson, and Sameroff \(2014\)](#), [Hastings \(2002\)](#)).

2. DATA

The data for our analysis are from the Millennium Cohort Study (MCS), a prospective study of infants born between 2000 and 2002 and representative of the overall UK population.⁵ The first wave of data collection took place when infants were around 9 months

⁴In a more recent study, [Feng, Han, Heckman, and Kautz \(2022\)](#) compare child, guardian, and teacher reports of child noncognitive skills using data from a longitudinal survey of children in primary schools in the Mianzhu region of China. They document that teacher reports are more predictive of later children cognitive and behavioral outcomes in school. Although relevant and interesting, these results do not directly speak to concerns regarding contamination. A measure can be both more predictive and have systematic bias. Our framework allows for both channels, where measurement error can directly affect predictive power and contamination can lead to systematic biases. Quantifying the role of contamination, unique to our analysis, is essential for properly assessing the nature of policy counterfactuals like those presented in Section 6.2.

⁵Infants born on eligible dates in eligible areas were selected from the Child Benefit Register, a universal benefit in the UK at the time. The sampling design allowed for overrepresentation of areas with high levels of childhood deprivation and high proportions of ethnic minorities ([Plewis \(2007\)](#)).

old and includes 18,552 children. Interviews were conducted at ages 3, 5, 7, 11, 14, and 17. Our analysis uses data from the first 4 waves (up to age 11), that is, the last year of primary school. We focus on this age range since this is when children are less able to self-assess their socioemotional well-being and reliance on different evaluators is most acute.

Sample selection The initial sample consists of singleton children and those with initial interviews not conducted by proxy (19,048 children). We select families with two parents (14,648 children) and exclude cases with missing data on gender, ethnicity, maternal age, and basic family demographics (14,598 children). Only families who appear in at least one wave from age 3 onward are included, as this is when cognitive outcomes are first recorded, yielding a final sample of 12,530 children (see Table 1). Over time, observations are dropped after parents separate or due to attrition, creating an unbalanced panel. Dropout is more likely among male children, those with less educated parents, and those from minority ethnic groups, though the regression results are robust to accounting for these factors.⁶

TABLE 1. Descriptive statistics.

	All Families $t = 0$ or $t = 1$		Two-Parent Families $t = 0$ or $t = 1$		Two-Parent Families $t = 0$ or $t = 1$		Two-Parent Families $t = 1$ or Later	
	Mean	std. dev.	Mean	std. dev.	Mean	std. dev.	Mean	std. dev.
	(1)		(2)		(3)		(4)	
% Male	0.51		0.51		0.51		0.51	
% White	0.87		0.88		0.88		0.88	
% Single mothers	0.15		0.00		0.00		0.00	
% First born child	0.42		0.41		0.41		0.40	
Child's age	0.93	0.52	0.91	0.48	0.91	0.48	3.33	0.90
Number of siblings	0.90	1.02	0.91	1.00	0.91	1.00	1.22	0.99
Mother's age at birth	28.89	5.84	29.83	5.35	29.84	5.35	30.05	5.23
Father's age at birth	32.09	5.90	32.34	5.75	32.35	5.74	32.48	5.63
Mother's years of schooling	12.06	1.79	12.28	1.82	12.28	1.82	12.35	1.83
Father's years of schooling	12.13	1.90	12.18	1.91	12.18	1.91	12.22	1.91
England	0.82		0.83		0.83		0.83	
Wales	0.05		0.05		0.05		0.05	
Scotland	0.09		0.09		0.09		0.09	
Northern Ireland	0.03		0.03		0.03		0.03	
<i>N</i>	19,048		14,648		14,598		12,530	

Note: UK Millennium Cohort Study. Sampling weights used throughout. Data on father's schooling only available for fathers in two-parent families. Column (1) is based on the whole sample of children who enter the study for the first time either in $t = 0$ or $t = 1$ (these correspond to the first and second wave of the study, when children were 9 months and 3 years old, respectively); column (2) restricts to the subsample of two-parent families; column (3) restricts further by eliminating observations with missing values on the demographics; column (4) shows our final estimation sample, which includes only two-parent families observed from $t = 1$ onwards.

⁶In particular, the baseline regression results of Table 4 are very similar when we consider only children that remain in the sample for the whole period.

TABLE 2. Child cognitive measures.

	<i>t</i> = 1 (Age 3)	<i>t</i> = 2 (Age 5)	<i>t</i> = 3 (Age 7)	<i>t</i> = 4 (Age 11)
<i>Administered by interviewer</i>				
Bracken School Readiness	✓			
BAS Naming Vocabulary	✓	✓		
BAS Picture Similarities		✓		
BAS Patterns Comprehension		✓	✓	
BAS Word Reading			✓	
NFER Progress in Maths			✓	
BAS Verbal Similarities				✓
Cambridge Gambling Task: quality				✓
Spatial Working Memory Task: strategy				✓
Spatial Working Memory Task: total errors				✓
<i>Assessed by teacher</i>				
FSP (Reading, writing, calculating, etc.)		✓		
Subject assessment (English, Math, and Science)			✓	✓

Note: UK Millennium Cohort Study.

Table 1 shows that there is very little difference in the gender and ethnic composition of these samples, as well as their geographical distribution. The age of the child (in years) at the first interview is also virtually identical. There is however a substantial difference with respect to sociodemographic characteristics. Mothers are usually older and more educated among the sample of two-parent families and in our final estimation sample.⁷

Child cognitive outcomes The cognitive skills of the child are measured using tests administered by the interviewer as well as teacher assessments. In each wave, at least two measures of child cognitive skills are available (see Table 2).

Some of the tests administered by the interviewer come from the British Ability Scales or BAS (Elliott, Smith, and McCulloch (1996)). At age 3, we rely additionally on the Bracken School Readiness Assessment (Bracken (2002)), which is considered a good indicator of success in education. At age 7, we use a variant of the National Foundation for Educational Research (NFER) Progress in Mathematics test. At age 11, we have measures from the Cambridge Gambling Task (CGT) and the Spatial Working Memory (SWM) task (RJOS+ (1998)).⁸

⁷Education is measured as years of schooling. This measure is obtained assuming: (i) all individuals with no qualifications left school before the end of compulsory education (with either 8, 9, or 10 years of schooling depending on their date of birth), (ii) those with a certificate of secondary education left school at the end of the compulsory schooling period (with either 9, 10, or 11 years of schooling depending on their date of birth), (iii) those with O-levels or equivalent qualifications have 11 years of schooling, (iv) those with A-levels or equivalent qualifications have 13 years of schooling, (v) those with a diploma in higher education have 14 years of schooling, and those with a degree or higher level of education have 15 years of schooling.

⁸All measures collected by the interviewer were obtained using Computer Assisted Personal Interviewing (CAPI) by interviewers who were specifically trained, but did not have a psychology background. Where appropriate, our analysis uses age-adjusted ability scores, which reflect the raw score and the difficulty of the items administered.

TABLE 3. Child noncognitive measures.

	<i>t</i> = 1 (Age 3)	<i>t</i> = 2 (Age 5)	<i>t</i> = 3 (Age 7)	<i>t</i> = 4 (Age 11)
<i>Reported by the mother</i>				
SDQ (hyperactivity, conduct, emotional, peer)	✓	✓	✓	✓
CSB (independence, emotional)	✓	✓	✓	
CSB (cooperation)			✓	
<i>Reported by interviewer</i>				
Behavior score (extreme behavior, attention, cooperation)	✓			
<i>Reported by teacher</i>				
FSP (dispositions, social, emotional)		✓		
SDQ (hyperactivity, conduct, emotional, peer)			✓	✓

Note: UK Millennium Cohort Study.

Our measures of child cognitive skills also include teacher assessments. The Early Years Foundation Stage Profile (FSP), reported by the teacher at the end of reception (age 5), describes the child's stage of development in the following areas: (i) personal, social, and emotional development; (ii) communication, language, and literacy; (iii) problem solving, reasoning, and numeracy; (iv) knowledge and understanding of the world; (v) physical development; (vi) creative development. We sum up all scales in all areas, excluding (i) and (v), to construct a measure of cognitive skills. At age 7 and 11, we use the teacher's evaluations across English, Math, and Science.

Child noncognitive outcomes The noncognitive skills of the child are obtained from parents, interviewers, and teachers; in each wave, at least three teacher and/or interviewer measures are available (see Table 3).

Across all waves the main respondent (usually the mother) is asked to complete the Strengths and Difficulties Questionnaire (SDQ). The SDQ is a behavioral questionnaire designed to measure psychological adjustment in children aged 3 to 16 (Goodman (1997)), and identifies five domains: (i) hyperactivity, (ii) conduct problems, (iii) emotional symptoms, (iv) peer problems, and (v) pro-social behavior. We add the scores across the first four components to obtain a measure of child noncognitive skills.⁹ Another measure is derived from the parents' answers to the Child Social Behavior Questionnaire (CSB), which captures the child's ability to perform tasks independently, to concentrate, and to control his/her emotional responses (MHQS+ (2004)). A set of questions to measure child cooperative behavior was added at age 7.

At age 3, the interviewer was asked whether the child had been fidgety, focused, disruptive, etc., during the assessment and the interview. Two questions captured extreme

⁹We exclude the pro-social measure since it was conceived as a separate measure representing strengths or resources (see Goodman (1997) and Goodman, Lamping, and Ploubidis (2010)). Our descriptive patterns and model results are not sensitive to including the pro-social subscore as an additional measure of child noncognitive skill.

behavior (i.e., child is dangerous, child is disruptive), others were combined using principal component analysis to obtain a measure of “focus and attention,” and of “cooperative behavior.” Teacher assessments consist of the component of the FPS (see above) which measures personal, social, and emotional development (age 5), and the SDQ (ages 7 and 11).

Maternal and paternal noncognitive measures The MCS provides measures of the mental health of the respondent and his/her partner at each interview. Specifically, the main respondent and the partner were asked to provide answers to questions from a shortened version of the Kessler questionnaire, used to diagnose mental illness (KBCE+ (2003)). A higher score on these items indicates the presence of psychological distress or depression. We also use the answer to a question on life satisfaction, scaled from 1–10 (1 being “completely dissatisfied” and 10 being “completely satisfied”).

Maternal and paternal cognitive measures Respondents report their own and their partner’s academic qualification, which we transform into years of schooling (see footnote 9). In the first wave of the survey, respondents are also asked about any difficulties they or their partner have reading books, filling forms, or performing everyday math. However, these binary measures are relatively uninformative since approximately 95% of respondents and partners indicate no problems with these activities. As a result, we rely primarily on years of schooling to measure parental cognitive skill.

3. PRELIMINARY EVIDENCE OF DISTORTIONS AND IMPACTS

In this section, we provide descriptive evidence of the presence of contamination in child noncognitive skill measures. We begin by examining raw differences in maternal and teacher evaluations of child noncognitive skills where the measurement tools are identical. At ages 7 and 11, both mothers and teachers evaluate child noncognitive skills using the SDQ. As a reminder, we use four components of the SDQ covering emotional symptoms, behavioral conduct, hyperactivity, and peer problems. There are five questions in each component for a total possible score of 10, where we scale responses such that higher numbers are associated with higher noncognitive skills. In Section 1 of the Supplemental Appendix (Del Bono, Kinsler, and Pavan (2026)), we present the questions and scoring for each SDQ component. The Supplemental Appendix will be referred to throughout the paper as SA.

Figure 1 illustrates the distributions of the differences between maternal and teacher measures in four domains of the SDQ. The maximum and minimum differences are 10 and -10 , but most differences range between 5 and -5 . More importantly, the distributions appear to be skewed slightly to the left, indicating that mothers are evaluating children more negatively on average. Additionally, the differences in maternal and teacher measures are strongly positively correlated across the components of the SDQ (average correlation is 0.25), suggesting systematic differences for the same child. While we cannot determine which respondent, mother, or teacher is closer to the truth, we can certainly conclude that the assessments differ.¹⁰

¹⁰For each SDQ subcomponent, a formal Kolmogorov–Smirnov test rejects that the maternal and teacher distributions are equal. SA Table 1 shows the joint density of maternal and teacher SDQ scores.

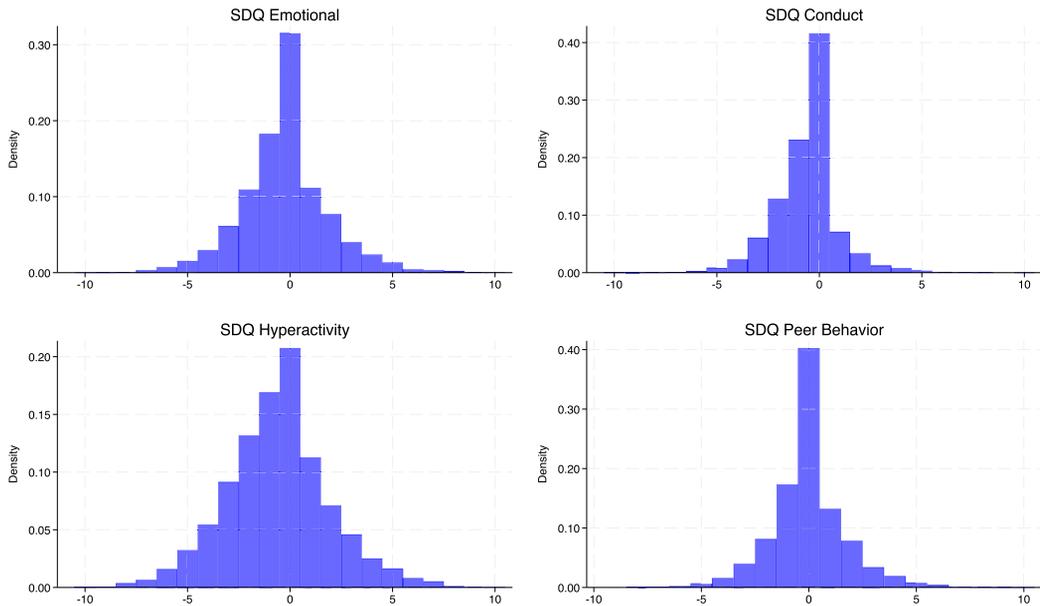


FIGURE 1. Differences between maternal and teacher SDQ measures. *Notes:* UK Millennium Cohort Study. Each panel plots the density of maternal SDQ responses minus teacher SDQ responses when children are ages 7 and 11. SDQ scores range from 0 to 10, with higher numbers indicating more positive behaviors. The sample includes only those children where mothers and teachers provide the four SDQ measures.

A level difference between maternal and teacher measures of child non-cognitive skill can easily be accommodated in standard dynamic latent factor models of skill development. However, if the measures of child noncognitive skills are affected by the skills or traits of the evaluator, key assumptions of the standard model may be violated. Figure 2 provides suggestive evidence that maternal noncognitive skills influence how mothers report about their child's non-cognitive skills. The figure plots the differences between maternal and teacher SDQ measures separately for low and high noncognitive skill mothers, where we proxy for maternal skill using raw Kessler scores.¹¹ The distributions reveal that low noncognitive skill mothers, or mothers with raw Kessler scores less than or equal to 5 (see PSMS+ (2012a)), evaluate their children more negatively relative to teachers than high noncognitive skill mothers across all subcomponents of the SDQ.

While Figure 2 suggests that responder traits can influence responses about child noncognitive skills, it is difficult to ascertain whether the observed differences are jointly significant. Additionally, to construct this figure we utilize only a subset of the available child noncognitive skill measures at specific ages. In the following paragraphs, we take a more systematic approach for documenting the links between child noncognitive skill measures and responder traits using additional measures and survey waves.

We first extract the principal factor for child and parental skills in the ages 3, 5, 7, and 11 waves using the measures discussed in the data section. These skill proxies are

¹¹SA Section 2 presents the questions and scoring for the Kessler survey administered to parents.

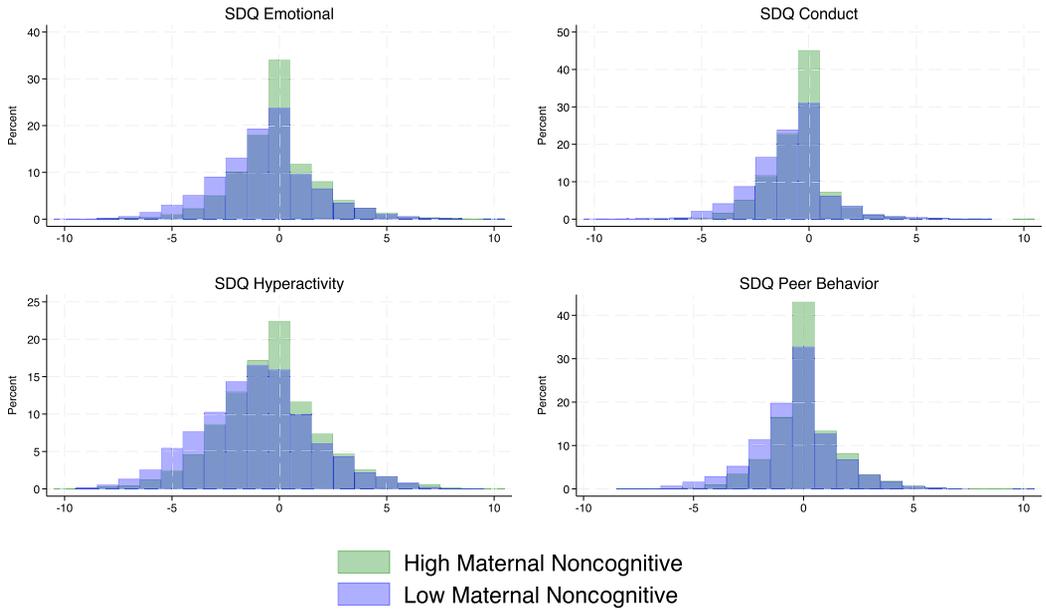


FIGURE 2. Differences between maternal and teacher SDQ measures by maternal skill. *Notes:* UK Millennium Cohort Study. Each panel plots the density of maternal SDQ responses minus teacher SDQ responses when children are ages 7 and 11 separately by maternal Kessler scores. Low noncognitive skill mothers have moderate to severe symptoms on the Kessler scale. SDQ scores range from 0 to 10, with higher numbers indicating more positive behaviors. The sample includes only those children where mothers and teachers provide the four SDQ measures.

noisy versions of the true underlying skills and are functions of different skill measures in different years. Additionally, if even one of the underlying skill measures is missing, the skill proxy will also be missing. The dynamic model we introduce in the next section will handle these concerns. Here, we simply provide additional suggestive evidence of contamination in child noncognitive skill measures that drive our modeling choices later on.

There are four types of individuals who report on a child’s noncognitive skills: mothers, fathers, interviewers, and teachers. An interviewer assesses the child at the end of the age 3 survey, while a teacher assesses the child in all subsequent periods. In every wave, either the mother or father evaluates the child’s noncognitive skills. To illustrate that the identity of the survey respondent can influence the assessment of the child, we create separate child noncognitive skill proxies for each type of respondent. We then explore how these evaluator-specific noncognitive skill proxies relate to child cognitive skills and parental cognitive and non-cognitive skills. The basic idea is that if each type of respondent provides dedicated measures of the child’s true noncognitive skills, the resulting skill proxies should project similarly on the other skill proxies.

Table 4 provides evidence that the identity of the evaluator can impact child noncognitive skill measures. The table reports estimates from the following regression model:

$$NonCog_{it}^r = \alpha_0 + \alpha_1 Cog_{it} + \alpha_2 MotherCog_i + \alpha_3 MotherNonCog_{it} + \alpha_4 X_{it} + u_{it}$$

TABLE 4. Evidence of distortions in child noncognitive skill measures.

	Child Noncognitive Skill _{<i>t</i>}			
	Mother Reported		Teacher/Interv. Reported	
	(1)	(2)	(3)	(4)
Child Cognitive Skill _{<i>t</i>}	0.206 (0.007)	0.191 (0.007)	0.291 (0.007)	0.270 (0.007)
Mother Cognitive Skill _{<i>t</i>}	0.117 (0.007)	0.095 (0.008)	0.021 (0.006)	0.019 (0.007)
Mother Noncognitive Skill _{<i>t</i>}	0.299 (0.007)	0.291 (0.007)	0.071 (0.007)	0.068 (0.007)
Missing Indicators	Y	Y	Y	Y
Demographics	N	Y	N	Y
<i>N</i>	33,905	33,905	26,818	26,818
<i>R</i> ²	0.182	0.208	0.110	0.138

Note: UK Millennium Cohort Study, data from $t = 1$ to $t = 4$ (child ages 3, 5, 7, and 11). Estimates are obtained by OLS. Variables representing child skills and parental skills are transformed by principal component analysis into factors with mean 0 and standard deviation 1. Demographic variables include: gender of the child, ethnicity, age of the child (in months) and its square, maternal age (in years) and its square, number of siblings, weekly family income, and region of birth. Standard errors are clustered at the individual level.

where $t = 1, \dots, 4$ (corresponding to ages 3, 5, 7, and 11) and r corresponds to the person reporting the child's noncognitive skill proxies, either the mother or the interviewer/teacher. The key controls are the principal factors for child cognitive skill (*Cog*), maternal cognitive skill (*MotherCog*), and maternal noncognitive skill (*MotherNonCog*).¹² In the first two columns, we report the coefficients that result when we use the child noncognitive skill proxy generated using the mother's responses as the dependent variable. In the third and fourth columns, we use the child noncognitive skill proxy generated from the teacher/interviewer responses as the dependent variable.¹³

Child noncognitive skills should be strongly correlated with the child cognitive skills and maternal skills since they are all interrelated. However, the key result in Table 4 is the difference in the conditional correlations across the skill measures according to the type of evaluator. When we utilize the mother's measures of child noncognitive skill, there is a strong relationship between child noncognitive skills and mother noncognitive skills. When we utilize the teacher/interviewer's measures of child noncognitive skills, the coefficient on the mother noncognitive skills declines by approximately 80%.¹⁴ Addition-

¹²When constructing the child cognitive skill proxy we do not include teachers' evaluations of the child cognitive abilities since we are concerned they may also be contaminated. In the full statistical model, we employ these measures and account for potential contamination directly.

¹³We assess whether the results are sensitive to demographic controls (X) by including the gender of the child, ethnicity, age of the child (in months) and its square, maternal age (in years) and its square, number of siblings, weekly family income and region of birth in the second and fourth columns. Although the principal factors are age specific (meaning mean zero by wave), we control directly for age since within the same cohort children may differ in their age by several months.

¹⁴All skill proxies are standardized to have a mean of zero and a standard deviation equal to one.

ally, the coefficient on the child cognitive skill increases substantially when we use the teacher/interviewer-generated child noncognitive skill.^{15,16}

While we view the differential impact of maternal noncognitive skills presented in Table 4 as evidence of contamination, an alternative explanation is that parents, teachers, and interviewers are reporting about different underlying latent skills. This concern is especially acute since the maternal, teacher, and interviewer factors for child noncognitive skill are based on different underlying skill measures. To investigate further, we project the raw SDQ measures that are common between parents and teachers onto child cognitive, maternal cognitive, and maternal noncognitive principal factors. If the evidence for contamination in Table 4 arises primarily because informants are reporting about different underlying skills, we would expect them to moderate significantly when we use overlapping measures. The results, presented in Table 5, indicate that the distortions do not moderate and exist across various aspects of child noncognitive skill. The maternal noncognitive factor projects much more strongly on all maternal-reported

TABLE 5. Evidence of distortions using only SDQ measures.

	SDQ Measures							
	Emotional		Conduct		Hyperactivity		Peers	
	Mother (1)	Teacher (2)	Mother (3)	Teacher (4)	Mother (5)	Teacher (6)	Mother (7)	Teacher (8)
Child Cognitive Skill _{<i>t</i>}	0.191 (0.016)	0.228 (0.021)	0.175 (0.013)	0.147 (0.014)	0.512 (0.021)	0.695 (0.027)	0.147 (0.014)	0.184 (0.019)
Mother Cognitive Skill _{<i>t</i>}	0.049 (0.017)	0.035 (0.020)	0.080 (0.013)	0.012 (0.013)	0.143 (0.024)	0.072 (0.028)	0.065 (0.015)	-0.011 (0.019)
Mother Noncognitive Skill _{<i>t</i>}	0.454 (0.018)	0.141 (0.021)	0.318 (0.014)	0.106 (0.015)	0.493 (0.023)	0.169 (0.028)	0.314 (0.015)	0.115 (0.019)
Missing Indicators	Y	Y	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y	Y	Y
<i>N</i>	15,077	9703	15,086	9702	15,052	9702	15,074	9702
<i>R</i> ²	0.106	0.037	0.116	0.064	0.165	0.195	0.110	0.034

Note: UK Millennium Cohort Study, data from *t* = 3 and *t* = 4 (child ages 7 and 11). Estimates are obtained using linear regressions. The dependent variables are raw SDQ scores in a particular domain as reported by mothers and teachers. Demographic controls as in footnote to Table 4. Standard errors are clustered at the individual level.

¹⁵We conducted a formal test of equality of the coefficients in columns (1) and (3) (as well as (2) and (4)) reestimating the model on the sample of observations for which both mother and teacher assessments are available. The results on this sample are virtually identical to those shown in Table 4 and the null hypothesis that the coefficients are the same is rejected in all cases.

¹⁶We explore whether contamination is mainly driven by mothers with low levels of noncognitive skills. In order to do so, we split the sample into above and below average maternal noncognitive skill groups, or exclude mothers with moderate or severe cases of depression according to their Kessler score (see PSMS+ (2012b)). In all cases, the coefficient associated with maternal noncognitive skill is large relative to the corresponding coefficient when teacher-reported measures of child noncognitive skill are used as the dependent variable. We also investigate whether the contamination varies with the number of siblings and the number of years of schooling of the mother. In all cases, we find that the contamination patterns are extremely similar to the results in Table 4.

TABLE 6. Teachers versus interviewers.

N_t reported by:	$t = 1$		$t > 1$	
	Mother (1)	Interviewer (2)	Mother (3)	Teacher (4)
Child Cognitive Skill $_t$	0.188 (0.011)	0.309 (0.011)	0.194 (0.008)	0.264 (0.008)
Mother Cognitive Skill $_t$	0.132 (0.010)	-0.018 (0.010)	0.078 (0.008)	0.036 (0.008)
Mother Noncognitive Skill $_t$	0.268 (0.011)	0.021 (0.010)	0.299 (0.008)	0.091 (0.008)
Missing Indicators	Y	Y	Y	Y
Demographics	Y	Y	Y	Y
N	9391	9343	24,514	17,475
R^2	0.218	0.130	0.207	0.158

Note: UK Millennium Cohort Study, data from $t = 1$ (child age 3) in column 1, and for $t = 2$ to $t = 4$ (child ages 5, 7, and 11). Estimates are obtained using linear regressions. Variables representing child skills and parental skills are transformed by principal component analysis into factors with mean 0 and standard deviation 1. Demographic controls as in footnote to Table 4. Standard errors are clustered at the individual level.

child SDQ measures as compared with teacher-reported child SDQ measures.¹⁷ In Table 6, we further show that interviewer and teacher assessments of child noncognitive skill correlate similarly with other child and parental skills. Moreover, both interviewer and teacher assessments differ significantly from maternal assessments. Overall, the results in Tables 5 and 6 indicate that contamination is relevant for many possible dimensions of children noncognitive skills.

In Table 7, we push the contamination idea one step further by looking at cases where the father is the main survey respondent and reports on the child non-cognitive skills. Although these are only a few cases and the sample is not representative of the total population, the exercise is useful for illustrative purposes. The first column of Table 7 uses the mother-reported measure of child skills as the dependent variable, but here we add paternal cognitive and noncognitive skills to the regression. Similar to the results from Table 4, we find that maternal non-cognitive skills are highly correlated with child noncognitive skills. Paternal skills are also correlated, but the strength of this relationship is weaker. The second column reports results from a similar regression but uses the teacher-reported measures of child noncognitive skills as the dependent variable instead. As before, we see that the correlation between maternal skills and child noncognitive skills is reduced, indeed now maternal and paternal skills are similarly related to child noncognitive skills. The third column uses the father-reported measures of child noncognitive skills. The striking result here is that father noncognitive skills now

¹⁷We conducted a formal test of equality of the coefficients across the mother- and teacher-reported SDQ measures, reestimating the models on the sample of observations for which both mother and teacher assessments are available. The results on this sample are virtually identical to those shown in Table 5 and the null hypothesis that the coefficients are the same is rejected in all cases.

TABLE 7. Additional evidence of distortions in child noncognitive skill measures.

	Child Noncognitive Skill _{<i>t</i>}		
	Mother Reported (1)	Teacher Reported (2)	Father Reported (3)
Child Cognitive Skill	0.186 (0.007)	0.267 (0.007)	0.152 (0.046)
Mother Cognitive Skill	0.080 (0.009)	0.013 (0.007)	0.033 (0.057)
Mother Noncognitive Skill	0.281 (0.007)	0.058 (0.007)	0.155 (0.053)
Father Cognitive Skill	0.034 (0.009)	0.011 (0.007)	0.068 (0.051)
Father Noncognitive Skill	0.044 (0.007)	0.040 (0.007)	0.225 (0.046)
Missing Indicators	Y	Y	Y
Demographics	Y	Y	Y
<i>N</i>	33,905	26,818	443
<i>R</i> ²	0.2117	0.140	0.304

Note: UK Millennium Cohort Study, data from $t = 1$ to $t = 4$ (child ages 3, 5, 7, and 11). Estimates are obtained by OLS. See footnote to Table 4 for the definition of variables. Standard errors are clustered at the individual level.

appear to be strongly correlated to child noncognitive skills.¹⁸ The differences between the mother and father noncognitive skills coefficients across columns one and three are statistically significant, despite the fact that we have a small number of observations where the father is the primary respondent.

The finding that the noncognitive skills of the reporting parent are always strongly related to the noncognitive skills of the child suggests that parental measures of child noncognitive skills may be influenced by parental skills or parental characteristics and, therefore, contaminated. This contamination means that parent-reported measures are correlated within a period and across time. Of course, teacher-reported measures of child noncognitive skills might also be influenced by teacher skills or characteristics, introducing a correlation across all teacher-reported measures within the same period. Ignoring systematic measurement error in parent and teacher reports is likely to bias estimates of skills production. In the next section, we develop a framework for estimating skills production that addresses these contamination issues under specific assumptions.

¹⁸One potential concern would be that when the father is the main respondent, this is a signal that the father plays a more prominent role in the household. This could result in a stronger link between the noncognitive skill of the child and that of the father. However, we examined the link between child noncognitive skill and parental noncognitive skill for the *same* children in column (3) of Table 7 when the *mother* is the main respondent (in a different wave of the study). We found that the pattern in column one reemerges, that is, it is the mother’s noncognitive skills that are most strongly related to the child noncognitive skills.

4. THE MODEL

4.1 Setup

To analyze in detail the dynamics of child skill development and parental emotional well-being, we follow a procedure similar to Cunha et al. (2010). In our model, a child and her parents are followed for T periods. Each child is characterized by cognitive and noncognitive skills, which are unobserved (to the econometrician). Parents are characterized by observed cognitive skills and unobserved noncognitive skills.¹⁹ The skills of the child in household i are denoted by (C_{it}, N_{it}) , where C indicates cognitive skill and N indicates noncognitive skill. The cognitive and noncognitive skills of the parents in household i are given by (C_i^P, N_i^P) , where $P \in \{M, F\}$ for maternal and paternal skills, respectively. All (child and parental) noncognitive skills are assumed to evolve over time. Child cognitive skills also change over time, while parental cognitive skills are assumed to be constant.

At $t = 1$, the first period in our model (corresponding to the interview conducted when the child was 3 years old), six skills are drawn from a joint density. We represent the initial skill draw for household i according to

$$S_{i1} \sim F_1, \quad (1)$$

where $S_{i1} = (C_{i1}, N_{i1}, C_i^M, N_i^M, C_i^F, N_i^F)$ is a six-dimensional vector. Child and parental skills evolve over time according to

$$S_{it+1} = f_t(S_{it}) + v_{it+1} \quad \text{and} \quad v_{it+1} \sim F_{t+1}, \quad (2)$$

with the restriction that $C_{it+1}^P = C_{it}^P = C_{i1}^P$. Since parental cognitive skills are time-invariant, f_t is essentially a four-dimensional function that represents the law of motion or production function of future child skills and parent noncognitive skills.²⁰

The function f_t can be flexibly specified to allow for the self-productivity of skills, and dynamic complementarities or substitutability across skills types. In other words, maternal and paternal skills have an impact on child skills that can be an increasing or decreasing function of past child skills. As is standard in this literature, child cognitive skills can foster noncognitive skills and vice versa. Maternal and paternal noncognitive skills are allowed to evolve over time as a function of all other skills, including their child's cognitive and noncognitive skills. The four elements of the shock, v_{it+1} , can be correlated with each other. While we define f_t and F_t to be general functions, in the estimation section we discuss the parametric assumptions we employ to estimate the model.

The above framework does not explicitly include observed household characteristics in the production of skill. This is because we purge our skill measures of demographic

¹⁹The model can be generalized to the case of unobserved parental cognitive skills. However, since the MCS contains only one informative measure of parental cognitive skills, we cannot treat the underlying skills as unobserved and we need to impose the assumption that parental cognitive skills are observed.

²⁰In our setting, the estimated effect of parental skills on future child skills captures both the direct effect and any indirect effect working through parental investments.

and household variables (see footnote 21 for additional detail) such that the unobserved skill components discussed above are orthogonal to these characteristics. By doing this, we are implicitly allowing demographics to affect the evolution of skill in a linearly separable fashion. An exception to this is the heterogeneity analysis we undertake in Section 6.1.1 where we analyze how the evolution of the skills differs across socioeconomic status and gender.

4.2 Skill measures

Child cognitive and noncognitive skills and parental noncognitive skills are not directly observed, but multiple noisy measures of these skills are available. We assume that the measurements are generated as follows:

$$M_{ijt}^C = \mu_{jt}^C + \alpha_{jt}^C C_{it} + \epsilon_{ijt}^C \tag{3}$$

$$M_{ijt}^{N^P} = \mu_{jt}^{N^P} + \alpha_{jt}^{N^P} N_{it}^P + \epsilon_{ijt}^{N^P} \quad \text{for } P \in \{M, F\} \tag{4}$$

where the above equations refer to measures of child cognitive skills and maternal and paternal noncognitive skills.²¹ The errors, ϵ , are assumed to be independent across individuals, measures, and time periods, and the total number of measures j for each type of skill can vary. Note that there is no measurement equation for parental cognitive skills since we assume these skills are observed.

In the previous section, we presented suggestive evidence that parent-reported measures of child noncognitive skills are likely to be correlated with other key variables, such as parental skills. It is also possible that parent-reported measures of child noncognitive skills are correlated over time. Thus, the measurement equations for parent-reported measures of child noncognitive skills can be written as

$$M_{Pijt}^N = \alpha_{P,1jt}^N N_{it} + \alpha_{P,2jt}^N C_i^P + \alpha_{P,3jt}^N N_{it}^P + \alpha_{P,4jt}^N \theta_i + \epsilon_{P,ijt}^N, \tag{5}$$

where C_i^P and N_{it}^P are parental cognitive and noncognitive skills, and θ_i is a family-specific factor that is time invariant.²² The latter produces a correlation across all

²¹We first residualize all MCS skill measures (including the child noncognitive skill measures discussed below and parental cognitive skill measures) by regressing them on child age in months, gender, ethnicity, an indicator for first born child, parent’s age at time of birth, and indicators for region and area of deprivation. We do this because child and parental characteristics may impact the measures directly. However, these observable characteristics can also capture observable components of skill. In this case, our production technology and measurement system is consistent with a model where observables affect skill accumulation but in a linearly separable fashion from the unobserved skill components. The standard errors presented later will reflect the fact that the residualized measures are estimates themselves. Note that if we estimate a version of the model where we do not residualize the MCS skill measures, the impulse response and policy counterfactuals are mostly unaffected.

²²We assume that any contamination in the child noncognitive skill measures does not have a direct impact on skill accumulation. However, if parents act on their perceptions of child noncognitive skills when choosing investment, contamination will affect the evolution of skills. This is not a problem for contamination driven by C_i^P and N_{it}^P , since parental skills are allowed to affect the evolution of skills as in equation (2). θ_i does not enter the production function and as a result we ignore this potential link between contamination and production. This is consistent with θ_i capturing a family factor that influences how parents respond to survey questions about their children but does not affect their behavior.

parent-reported child skills measures, which is fixed over time. This might arise if, conditional on parental skills, some parents are more or less likely to classify certain child behaviors as problematic.²³ Note that only one parent reports on a child's noncognitive skill in a given survey wave.

In addition to a parent, other survey participants report on the child's non-cognitive skills. These are typically the child's school teachers, although at $t = 1$ (age 3) this is the interviewer since children are not yet in school. The presence of multiple evaluators is key to identify contamination in parent-reported measures of child noncognitive skills, as we will show. Notice, however, that identification does not rely on the assumption that these additional measures are free of contamination. Indeed, we can allow for correlated measurement error in the teacher-reported measures:

$$M_{T,ijt}^N = \alpha_{T,1jt}^N N_{it} + \alpha_{T,2jt}^N T_{it} + \epsilon_{T,ijt}^N. \quad (6)$$

Here, the random component T_{it} is assumed to be independent of all other variables, and is independent across individuals and time periods but constant across contemporaneous measures for the same individual. The presence of T_{it} accounts for the idea that some teachers are more or less likely to classify certain child behaviors as problematic conditional on the child's true noncognitive skills.

An important assumption in our model is that the random teacher effect is independent of all other skills. One concern might be that a teacher's evaluation of child noncognitive skills is contaminated by child cognitive skills, since cognitive skills development is a primary focus of schooling. This would mean that T_{it} is correlated with C_{it} . However, Table 6 shows that the relationship between child cognitive and noncognitive skills are very similar when we use teacher-reported or interviewer-reported measures of child noncognitive skills. Interviewers are likely less focused on cognitive skills development, suggesting that our assumption that T_{it} and C_{it} are uncorrelated is reasonable.

A second source of dependence could arise through the sorting of teachers and students. Note that the random teacher effect in our setting captures a level shift in how teachers evaluate noncognitive child skills. This is not obviously related to typical measures of teacher quality making it less likely that parents and teachers sort according to this feature. Regardless, we look for evidence of sorting by examining whether the conditional correlations between child cognitive skills, maternal skills, and teacher-reported child noncognitive skills change as we include observed and unobserved teacher and school characteristics. The basic idea is that we would expect these correlations to be smaller within schools or teachers if households are strongly sorted into schools and classroom. In Table 8, we display coefficients for a regression of teacher-reported child noncognitive skills on child and maternal skills with and without school or teacher effects.²⁴ The estimates, while statistically different, are economically of the same magnitude suggesting that sorting is likely not a primary concern.

²³The error term $\epsilon_{P,ijt}$ is assumed to be independent across individuals, measures, and time periods.

²⁴Teacher identifiers are only available in wave 3 (child age 7), which is why the sample is smaller in the final column.

TABLE 8. Teacher and school controls.

	Teacher-Reported Measures		
	Baseline (1)	School FE (2)	Teacher FE (Age 7 Only) (3)
Child Cognitive Skill _{<i>t</i>}	0.263 (0.009)	0.243 (0.012)	0.281 (0.043)
Mother Cognitive Skill _{<i>t</i>}	0.036 (0.008)	0.048 (0.013)	0.021 (0.040)
Mother Noncognitive Skill _{<i>t</i>}	0.090 (0.009)	0.075 (0.011)	0.101 (0.035)
Missing Indicators	Y	Y	Y
Demographics	Y	Y	Y
<i>N</i>	17,347	17,347	5536
<i>R</i> ²	0.158	0.480	0.800

Note: UK Millennium Cohort Study, data from child ages 5, 7, and 11. Estimates are obtained using linear regressions. Variables representing child skills and parental skills are transformed by principal component analysis into factors with mean 0 and standard deviation 1. Demographic controls as in footnote to Table 4. Standard errors are clustered at the individual level in column 1, at the school-level in column 2, and at the teacher level in column 3.

A final source of dependence between T_{it} and other model components, including $T_{it'}$ for $t' \neq t$, is through direct communication. Parents and teachers typically meet multiple times over the course of a school year. A low noncognitive skill mother may share negative views about their child’s noncognitive skill with the teacher, thus influencing the teacher’s evaluation directly. Similarly, teachers can share opinions of a student with teachers in other grades. Such communication across survey respondents is likely to induce a positive correlation between T_{it} and N_{it}^M , for example, a violation of our identifying assumption. However, it is important to point out that this violation likely leads us to understate the importance of measurement distortions. The correlation between parent skills and teacher effects will be inappropriately attributed to variability in child non-cognitive skill that will then project even more strongly on maternal skills in a dynamic model.

While we lack data to directly investigate the importance of this communication channel, we analyze the robustness of our key findings to violations of our identifying assumptions by imposing various levels of correlation between T_{it} and N_{it}^M or T_{it} and $T_{it'}$ (among others) and reestimating the model and counterfactuals. As we discuss in more detail in Section 6.2, the policy effect distortions tend to be larger when teacher effects are positively correlated with maternal noncognitive skills, for example.

4.3 Identification

In this section, we briefly describe our approach to identification. A more formal treatment is provided in SA Section 1. The components of the model that need to be identified include: the production function (f_t), the distribution of skill shocks (F_t), and all the parameters of the measurement equations. The key challenge is to pin down the joint distribution of unobserved skills, S_t . Because parental cognitive skills are observed, they

pose no threat to the identification of our model and for the sake of brevity are excluded from the discussion. Once the joint density of S_i is identified, we can identify the law of motion of skills, or production function, as the expectation of one skill conditional on past skills. Suppose for a moment that we know the joint density of all skills S_i , then for skill Y_{it+1} where $Y \in \{C, N, N^M, N^F\}$ define

$$f_{it+1}^Y(Y_{it}) \equiv E(Y_{it+1}|S_{it}) \quad (7)$$

where the mean of v_{it+1} is normalized to zero. We can then recover $v_{it+1}^Y = Y_{it+1} - E(Y_{it+1}|S_{it})$ and identify F_t using the distribution of v_{it+1}^Y .

Although it is clear that f_t and F_t can be identified when the joint distribution of S_i is known, S_i is unobservable. To identify the joint distribution of S_i , we use the measurement model described in the previous section. Following the approach of [Cunha, Heckman, and Schennach \(2010\)](#), it is straightforward to show that the first and second moments of all unobserved skills *other than* the child noncognitive skills can be identified by taking the appropriate covariances between measures. For example, after normalizing the loading factor to one on the first measure of each skill, the following four covariances,

$$\text{Cov}(M_{it}^C, M_{i\tau}^C) = \sigma_{t,\tau}^C \quad \text{for } t \neq \tau \quad (8)$$

$$\text{Cov}(M_{ijt}^C, M_{i\tau}^C) = \alpha_{jt}^C \sigma_{t,\tau}^C \quad \text{for } t \neq \tau \quad (9)$$

$$\text{Cov}(M_{ijt}^C, M_{it}^C) = \alpha_{jt}^C \sigma_t^C \quad (10)$$

$$\text{Cov}(M_{it}^C, M_{i\tau}^{N^P}) = \sigma_{t,\tau}^{CN^P} \quad \text{for } P \in \{M, F\}, \quad (11)$$

can be used to identify the covariance of child cognitive skills over time, the loading factor on child cognitive skills across measures, the variance of child cognitive skills each period, and how child cognitive skills varies with parent noncognitive skills. Similar covariances can be used to identify the second moments related to parental noncognitive skills.

Identifying the second moments related to child noncognitive skills is more challenging. All the measures related to child noncognitive skills have additional unobservables that are common across multiple measures. If we took a strategy similar to the one above, we would not be able to isolate terms related only to N_{it} . For example, the covariance between two teacher-reported measures in the same period will also contain the variance of T_{it} . Additionally, the covariance between a teacher-reported measure and parent-reported measure will contain not only the variance of child noncognitive skill, but also how the child noncognitive skill covaries with parental skills.

Two assumptions are needed to identify the second moments related to the child's noncognitive skills: (1) contamination in the parent-reported measures is independent of the contamination in the teacher-reported measures, and (2) contamination in teacher-reported measures is independent over time. To see how the latter assumption aids identification, consider the covariance between two teacher measures from different periods:

$$\text{Cov}(M_{T,it}^N, M_{T,i\tau}^N) = \sigma_{t,\tau}^N \quad \text{for } t \neq \tau$$

$$\text{Cov}(M_{T,ij}^N, M_{T,i1\tau}^N) = \alpha_{T,1j}^N \sigma_{t,\tau}^N \quad \text{for } t \neq \tau.$$

These two observable quantities identify the covariance of child noncognitive skills across different time periods and the loading factor relative to the teacher measure. Assumption (1) allows us to pin down the variance of child non-cognitive skills each period by taking the covariance between teacher and parent-reported measures of child noncognitive skill. Additional details are provided in SA Section 1.

Once the first and second moments of S_i have been identified, we show that the joint distribution of S_i is nonparametrically identified. This is necessary to allow skills evolution to be nonlinear. The proof broadly follows Theorem 1 of [Cunha, Heckman, and Schennach \(2010\)](#), though we modify it slightly to account for correlated errors in the child non-cognitive measures.

5. ESTIMATION AND RESULTS

Following [Attanasio, Meghir, and Nix \(2020\)](#), [ACFM+ \(2020\)](#), and [Agostinelli and Wiswall \(2025\)](#), estimation of the model proceeds in two steps. First, we flexibly estimate the joint distribution of child and parental skills, along with all the parameters of the measurement system. Second, we draw from the estimated skill distribution and estimate the skill technology. SA Section 1 provides the details for the estimation procedure. While we do not show all the estimates of the first stage, we can use those estimates to provide insight into the size of the parental and teacher distortions present in the measures of child noncognitive skill.

In [Table 9](#), we present the estimated fraction of the variance for each measure that is the result of the true underlying skill and distortion. These fractions do not add to one since part of the variation in each measure is also the result of measurement error. The first panel of the table presents the signal strength of the child cognitive skill measures based upon standardized exams. By construction, there is no contamination in these measures. In the first three periods, most of the measures are close to 50% signal, while in the final period the measures are noisier. Overall, the signal component of our child cognitive measures are lower than those reported in other studies ([Cunha, Heckman, and Schennach \(2010\)](#), [ACFM+ \(2020\)](#)). Between $t = 2$ and $t = 4$, teachers also evaluate the cognitive skill of the child. This is a subjective measure, which is also affected by the teacher random effect. In the second panel of [Table 9](#), we show that there is a nontrivial amount of noise in these measures that will be correlated with the teacher-reported measures of child noncognitive skill.

The third and fourth panels of the table show the degree of contamination in the parent- and teacher-reported measures of child noncognitive skill. The share of the variance in these measures resulting from parent skill, parent random effects, or teacher random effects are quite large. For example, when parents respond regarding the hyperactivity of their child, the share of the variation stemming from distorting components ranges from 12.4% to 18.6% across period. The share of contamination in the teacher responses are even larger. In general, parent-reported measures of child noncognitive skill provide a better signal than teacher-reported measures. This is consistent with parents observing their children over longer periods of time and in different social contexts.

TABLE 9. Contamination in measurements.

	<i>t</i> = 1		<i>t</i> = 2		<i>t</i> = 3		<i>t</i> = 4	
	Signal	Cont.	Signal	Cont.	Signal	Cont.	Signal	Cont.
<i>Child Cognitive—Test Based</i>								
Braken	58.1%	-						
BAS Naming Vocabulary	41.9%	-	36.7%	-				
BAS picture similarity			21.5%	-				
BAS pattern compr.			32.5%	-	33.4%	-		
BAS Word Recognition					52.1%	-		
NFER in Math					53.1%	-		
BAS verbal							25.7%	-
GTC quality							6.9%	-
CANTAB swm strat							12.4%	-
CANTAB swm err							24.2%	-
<i>Child Cognitive—Teacher Reported</i>								
FSP			31.8%	24.9%				
Teacher assessment					61.1%	5.7%	70.7%	3.8%
<i>Child Noncognitive—Parent Reported</i>								
SDQ emotional	20.9%	5.9%	28.6%	8.2%	30.3%	8.4%	34.0%	6.4%
SDQ conduct	35.3%	9.9%	34.6%	12.8%	35.5%	13.8%	29.7%	10.3%
SDQ hyperactivity	33.3%	12.4%	38.1%	18.0%	40.4%	18.6%	41.1%	14.6%
SDQ peer	20.0%	5.6%	33.8%	8.3%	35.5%	7.9%	49.4%	4.8%
Q. Independence	6.5%	0.1%	18.0%	1.0%	26.9%	1.7%		
Q. Emotional	36.4%	10.7%	40.1%	14.2%	40.7%	15.2%		
Q. Cooperation					32.0%	6.8%		
<i>Child Noncognitive—Interviewer/Teacher Reported</i>								
Focus	5.7%	94.1%						
Cooperation	2.7%	37.9%						
Extreme behavior	4.4%	11.5%						
FSP, personal			8.9%	44.9%				
FSP, social			8.5%	65.3%				
FSP, emotional			9.3%	70.9%				
SDQ emotional					6.9%	3.5%	15.6%	3.4%
SDQ conduct					7.8%	37.2%	12.7%	32.6%
SDQ hyperactivity					14.9%	53.6%	18.2%	42.1%
SDQ peer					11.8%	13.6%	17.7%	7.6%

Note: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in Sections 4 and 5. Each entry in this table represents the fraction of the variance of a given measurement that is explained either by the true skill (signal) or by the contamination. Consider, for example, a parent reported measure of the child noncognitive skills. This measure can be written as

$$M_{Pijt}^N = \alpha_{P,1jt}^N N_{it} + \alpha_{P,2jt}^N C_i^P + \alpha_{P,3jt}^N N_{it}^P + \alpha_{P,4jt}^N \theta_i + \epsilon_{P,ijt}^N.$$

The signal corresponds to the fraction of the total variance that is explained by the variance of $\alpha_{P,1jt}^N N_{it}$. The contamination corresponds to the fraction of the variance that is explained by the variance of $\alpha_{P,2jt}^N C_i^P + \alpha_{P,3jt}^N N_{it}^P + \alpha_{P,4jt}^N \theta_i$.

In a second step, we focus on estimating the skills production functions. The details of the procedure are also presented in SA Section 1, and estimates of the skill transition functions are presented in SA Table 2. As we discuss in Del Bono, Kinsler, and Pavan (2022), the estimated parameters are affected by the normalizations we imposed

for identification and therefore discussing those estimates is not very meaningful. However, in the Appendix we present the coefficient estimates themselves to highlight some basic overarching patterns, like self-productivity and cross-productivity in skills. The estimates also show that the noncognitive skill of the child seems to have a statistically significant impact on the evolution of the noncognitive skill of the parents, especially the mother. Finally, there is evidence of nonlinearity in the skill transition functions, though it appears rather small in magnitude. In fact, imposing linearity only mildly reduces the fit of the estimated production functions.

In [Del Bono, Kinsler, and Pavan \(2022\)](#), we also show that some properly defined treatment effects are identified without loss of generality. In the next subsection, following those results, we provide a more intuitive interpretation of the estimates by exploring how a shock to skills in one period impacts skills in a future period in terms of standardized measures.

6. ASSESSING THE IMPACT OF CONTAMINATION

While estimates of the skill production and measurement equations are informative, they do not fully reveal how contamination in child noncognitive skill measures can distort our understanding of skill formation or the evaluation of interventions aimed at improving child outcomes. In this section, we undertake two empirical exercises to highlight the practical consequences of measurement contamination.

In the first exercise, we assess how accounting for contamination alters our understanding of parent–child skill dynamics. We compare treatment effects derived from our dynamic latent factor model to those obtained from alternative specifications that ignore contamination. These treatment effects capture how an exogenous increase in child or parental skill at a given age affects the future trajectory of child and parental skills.²⁵ Our comparisons show that ignoring contamination leads to substantially biased inferences, particularly regarding the role of child noncognitive skills and the influence of parental traits.

In the second exercise, we turn to the evaluation of early childhood policies—such as universal child care—and examine how contamination biases reduced-form estimates of policy effects. Even under random assignment, if the policy also affects parental skills, and if parent-reported child measures are contaminated by those same skills, the observed effect on the child will conflate real skill gains with rater-driven distortions. We use our model to quantify this bias and show that it accounts for a substantial share of the observed effect.

6.1 *Contamination in skill dynamics from latent factor models*

The first exercise we pursue is to increment child and parental skills in various periods and observe how final skills are affected. These treatment effects provide an intuitive

²⁵Many other papers that employ dynamic latent factor models to study skill development examine similar types of treatment effects to gain better insight into skill dynamics. Examples include [Agostinelli and Wiswall \(2025\)](#), [ACFM+ \(2020\)](#), and [Aucejo and James \(2021\)](#).

way to understand the relationships between skills across various periods of childhood. To generate treatment effects, we first simulate child and parental skills in the initial period. We then increase one type of skill, say child noncognitive skill, and simulate how this and all other skills evolve using the estimated production function. We then compare the treatment effects from our main specification with the treatment effects generated by three alternative models that neglect contamination issues.

The first alternative model is one where we ignore the availability of teacher-reported, child noncognitive skill measures. Thus, we reestimate the entire model relying only on parental measures of child noncognitive skills. In this model, we do not include any measurement distortions in child noncognitive skill measures since it is not possible to identify the related parameters when measures come from one type of evaluator only. Most research on child noncognitive skill development is estimated in such a fashion since most datasets lack teacher interviews.

Even when teacher interviews are available, a naïve approach would be to simply combine their assessments of the child with parental assessments and estimate a standard model. We investigate whether this approach yields different conclusions regarding the production of skill. Finally, we estimate a model that relies only on teacher measures of child noncognitive skill. Similar to the parent-only model, we do not allow for measurement distortions (teacher effects) in the child noncognitive measures since the within-teacher covariances across measures are needed to identify the variance of child noncognitive skill and the related factor loadings.

Tables 10 and 11 show how boosting child and parental skills at $t = 1$ (age 3) impacts average child and parental skills at $t = 4$ (age 11) for our main, parent-only, parent-teacher, and teacher-only models.²⁶ All skill boosts are standardized to reflect a one standard deviation increase, and the resulting impact is standardized according to the relevant skill distribution in the final period.²⁷ As an example, the first number in Table 10 indicates that a one standard deviation increase in child cognitive skill at $t = 1$ leads to a 0.578 standard deviation increase in child cognitive skill at $t = 4$. This effect includes not just a self-productivity effect, but all the cross-skill effects accumulating over time.

There are a number of interesting findings in Table 10. Focusing first on our main specification, we find that child cognitive skill and paternal cognitive skill have the largest impact on average child cognitive skill at $t = 4$. For child non-cognitive skill at $t = 4$, it is child noncognitive skill at $t = 1$ and maternal non-cognitive that have the largest impact.

When we compare our main specification with the models that ignore contamination in child noncognitive skill measures, we would expect the parameters that govern

²⁶We can also investigate the skill impacts at $t = 2$ and $t = 3$ after increasing skills at $t = 1$. Alternatively, we can boost skills at ages $t = 2$ and examine the subsequent effects. We focus on the endpoints for illustration purposes.

²⁷As noted earlier, Del Bono, Kinsler, and Pavan (2022) prove that treatment effects anchored to the standard deviation of child skill and based on a translog technology are identified regardless of the location and scale normalizations. In this case, we have a linear production technology because we define the measures as a function of skill. The model is equivalent to defining the measures to be a function of log skill and imposing a translog production function.

TABLE 10. Impulse response: child skills.

	Main Specification		Only Parental Measures		All Measures (No Distortions)		Only Teacher Measures	
	C_4	N_4	C_4	N_4	C_4	N_4	C_4	N_4
+1sd in C_1	0.578 (0.018)	0.110 (0.016)	0.593 (0.017)	0.115 (0.014)	0.633 (0.019)	0.072 (0.017)	0.645 (0.021)	0.097 (0.019)
+1sd in N_1	0.088 (0.019)	0.465 (0.021)	0.074 (0.019)	0.533 (0.014)	0.009 (0.021)	0.488 (0.015)	0.030 (0.015)	0.009 (0.002)
+1sd in C_1^M	0.074 (0.016)	0.008 (0.028)	0.077 (0.016)	0.001 (0.012)	0.069 (0.016)	0.010 (0.011)	0.073 (0.017)	0.022 (0.015)
+1sd in N_1^M	0.005 (0.018)	0.102 (0.031)	-0.006 (0.019)	0.051 (0.015)	0.003 (0.020)	0.049 (0.016)	0.009 (0.018)	0.047 (0.012)
+1sd in C_1^F	0.116 (0.013)	0.044 (0.012)	0.110 (0.013)	0.050 (0.011)	0.099 (0.014)	0.059 (0.010)	0.104 (0.014)	0.031 (0.016)
+1sd in N_1^F	0.023 (0.016)	0.022 (0.013)	0.026 (0.016)	0.025 (0.013)	0.015 (0.018)	0.034 (0.012)	0.019 (0.016)	0.022 (0.011)

Note: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in Sections 4 and 5. In each row, we report the impact on children's skills at age 11 when different skills are shocked by 1sd when the children are 3 years old. The first two columns refer to our main specification. The second two columns present results for a model that does not correct for the contamination and utilizes only parental reported children noncognitive measures. The third two columns refer to a model that does not correct for contamination, but utilizes all available measures. The final two columns refer to a model that does not correct for contamination and utilizes only teacher reported children noncognitive measures. Standard errors are obtained bootstrapping 30 times the original data set at the individual level.

TABLE 11. Impulse response: parental skills.

	Main Specification		Only Parental Measures		All Measures (No Distortions)		Only Teacher Measures	
	N_4^M	N_4^F	N_4^M	N_4^F	N_4^M	N_4^F	N_4^M	N_4^F
+1sd in C_1	0.002 (0.011)	0.022 (0.012)	0.013 (0.011)	0.023 (0.012)	-0.005 (0.012)	0.014 (0.013)	0.045 (0.014)	0.040 (0.015)
+1sd in N_1	0.123 (0.012)	0.056 (0.010)	0.116 (0.013)	0.061 (0.012)	0.115 (0.013)	0.065 (0.012)	-0.002 (0.005)	0.006 (0.006)
+1sd in N_1^M	0.309 (0.014)	0.077 (0.012)	0.286 (0.014)	0.072 (0.012)	0.291 (0.013)	0.074 (0.013)	0.325 (0.013)	0.076 (0.013)
+1sd in N_1^F	0.066 (0.014)	0.359 (0.019)	0.072 (0.013)	0.354 (0.018)	0.080 (0.014)	0.363 (0.018)	0.061 (0.013)	0.362 (0.020)

Note: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in Sections 4 and 5. In each row, we shock a different skill by 1sd when the child is 3 years old, and report the impact on parental skills when the child is 11. The first two columns refer to our main specification. The second two columns present results for a model that does not correct for the contamination and utilizes only parental reported children noncognitive measures. The third two columns refer to a model that does not correct for contamination, but utilizes all available measures. The final two columns refer to a model that does not correct for contamination and utilizes only teacher reported children noncognitive measures. Standard errors are obtained bootstrapping 30 times the original data set at the individual level.

noncognitive skill dynamics to be particularly affected since this is where contamination enters. This is precisely what we find. When only parental measures are employed, the self-productivity of child noncognitive skill is about 15% larger than in our baseline model, a difference that is statistically significant at a 5% level. Also, the estimated effect of maternal noncognitive skills in fostering child noncognitive skills is 50% smaller than in the model that adjusts for distortions. The associated p -value for the difference in the coefficients is 0.06. These findings are not surprising in light of the fact that when only parental measures are available, the mother's skills—which are highly persistent—are partially absorbed by the child noncognitive skills.

In the model that uses teacher and parent measures of child noncognitive skills (but does not model contamination), the estimated cross-productivities of child skills are 35–90% lower than in our main specification and statistically different at a 5% level. This alternative model does not allow for a separate teacher effect in the teacher-reported measures of child noncognitive skill. Therefore, the teacher effect—which is period-specific (or nonpersistent)—is absorbed into the child noncognitive skill, making it less predictive of future cognitive skill and vice versa. This feature also appears in the teacher-only model since there is also no teacher effect. More striking, however, is the lack of self-productivity in noncognitive skill in the teacher-only model. This is again due to the fairly weak correlation of teacher measures over time, likely driven by the presence of independent rater effects.

Previous literature on child skill dynamics has emphasized own and cross-skill complementarities and the influence of parental skill inputs. These exercises illustrate that there are potentially large biases in these parameters when contamination in child noncognitive skill measures is ignored. For example, [Cunha, Heckman, and Schennach \(2010\)](#) estimate a dynamic model of child skills formation using only parent-reported measures. They find that a child's current noncognitive skill is the most important factor for developing future noncognitive skill, particularly for children aged 7 to 14—a group that closely matches our sample. The noncognitive skill self-productivity parameter is nearly an order of magnitude larger than the parameters for child cognitive skills and parental skills. A direct comparison with our findings is difficult since they use a CES technology and do not report easily interpretable treatment effects. However, our findings suggest that correcting for contamination in parent-reported measures of child noncognitive skill effect would likely result in a smaller self-productivity effect and an increase in the parameters associated with parental skills.

Table 11 focuses instead on the impact skill changes at $t = 1$ have on parental noncognitive skills at $t = 4$. Comparing different models, with and without correction for contamination, reveals few differences. This is because the parental measures are not distorted and the self-productivity effects, which range between 0.286 and 0.359 and are generally larger for fathers than mothers, tend to dominate.

Most interesting however is the fact that the noncognitive skill of the child influences the future noncognitive skill of the mother. Specifically, in our main specification a one standard deviation increase in child noncognitive skill at age 3 leads to a statistically significant 0.123 standard deviation increase in maternal noncognitive skill when the

TABLE 12. Heterogeneity analysis: child skills.

	Female		Male		Low Income		High Income	
	C_4	N_4	C_4	N_4	C_4	N_4	C_4	N_4
+1sd in C_1	0.575 (0.026)	0.144 (0.021)	0.587 (0.019)	0.098 (0.027)	0.606 (0.023)	0.136 (0.020)	0.562 (0.028)	0.083 (0.024)
+1sd in N_1	0.097 (0.023)	0.468 (0.032)	0.079 (0.022)	0.438 (0.032)	0.070 (0.020)	0.405 (0.024)	0.108 (0.023)	0.547 (0.037)
+1sd in C_1^M	0.073 (0.023)	0.009 (0.050)	0.080 (0.016)	0.042 (0.043)	0.053 (0.018)	0.013 (0.054)	0.092 (0.018)	-0.009 (0.055)
+1sd in N_1^M	-0.005 (0.025)	0.073 (0.051)	0.016 (0.016)	0.142 (0.032)	-0.015 (0.022)	0.126 (0.033)	0.018 (0.017)	0.054 (0.054)
+1sd in C_1^F	0.113 (0.023)	0.052 (0.016)	0.108 (0.019)	0.040 (0.020)	0.091 (0.022)	0.044 (0.021)	0.137 (0.018)	0.051 (0.023)
+1sd in N_1^F	0.037 (0.022)	0.018 (0.016)	0.009 (0.016)	0.021 (0.016)	0.011 (0.024)	0.007 (0.023)	0.041 (0.015)	0.040 (0.014)

Note: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in Sections 4 and 5. In each row, we report the impact on children’s skills at age 11 when different skills are shocked by 1sd when the children are 3 years old. The models are estimated separately by gender and family income. Standard errors are obtained bootstrapping 30 times the original data set at the individual level.

child is 11.²⁸ This effect is equally strong at different points in time. For example, a one standard deviation boost in child non-cognitive skills at age 5 (age 7) improves maternal noncognitive skills by 0.16 (0.11) standard deviations at age 11 (results not shown). The effect on fathers is weaker, on the order of 0.056 of a standard deviation, and there are no significant impacts of the child cognitive skills on future parental noncognitive skills.

6.1.1 Heterogeneous effects In our main specification, we assume that demographic characteristics such as gender or family income influence the distribution of skills only at the mean. This is equivalent to assuming that the skill transition equation does not vary with these characteristics. We relax this assumption by re-estimating the model separately for girls and boys, and for families below and above median annual family income. Table 12 reports the impulse response functions by gender and level of income. We focus on these aspects as the relationship between parental skills and child skills might differ due to biological differences in aspects of child development, access to information, or exposure to economic shocks (Persson and Rossin-Slater (2018), Ahun, Gapare, Gariépy, and Côté (2021)).

We see a few significant differences between girls and boys. Specifically, female noncognitive skills at $t = 4$ are more responsive to changes in both cognitive and noncognitive skills at $t = 1$ than it is the case for male noncognitive skills. By contrast, the response of cognitive skills to changes in early cognitive or noncognitive endow-

²⁸This result might be due to the fact that our measures of maternal noncognitive skills capture maternal mental health, so it might not extend to other measures, such as locus of control or self-esteem, for example.

ment is basically the same. Most interestingly, we see that an initial boost in maternal noncognitive skills has a larger impact on boys than girls.

Looking at the difference in the impulse response functions by family income, we see again some significant heterogeneity in the production technology. The own and cross-productivity of child cognitive skills are larger for children from low-income families, while the own and cross-productivity of child noncognitive skills are larger for children from high-income families. Moreover, we see that parental cognitive skills have a larger impact on children from high-income families, possibly capturing a complementarity with higher parental investments that these families can afford, such as better quality schools.

Looking at the measurement model, we see that contamination in parental measures of child noncognitive skills is generally larger for boys than girls (except for the SDQ emotional measure), and somewhat smaller for high-income families. We also looked at the impulse response functions for parental skills, but no statistically significant differences emerge in relation to gender or family income.²⁹

6.2 Contamination in reduced-form policy evaluations

While Section 6.1 demonstrated how ignoring contamination distorts treatment effect estimates derived from a dynamic model of skill formation, this section shows that contamination also poses serious threats to design-based evaluations of child-focused policies. This is because any policy or intervention aimed at improving child skills can either directly or indirectly influence parental skills. An evaluation that relies on parental measures of child noncognitive skills can generate spurious impacts, no matter how robust the research design. For example, a number of papers estimate the impact universal child care programs have on child noncognitive skills and/or maternal noncognitive skills.³⁰ As our results illustrate, part of the estimated impact of universal child care on child noncognitive skills could be due to the contamination driven by changes in maternal skills.

The environment we have in mind is as follows. Imagine that to evaluate a child care policy a researcher randomly assigns families to treatment and control groups. The treated families send their children (age 3) to free child-care while in the control group the mother stays home with the child. Mothers report on the noncognitive skill of their child at the end of the age 3 year, and at the end of the experiment (age 5). In follow-up surveys at ages 7 and 11, mothers are again asked to report on the child's noncognitive skills.

Using our model, we can quantify how contamination in parent-reported measures of child noncognitive skills influences estimates of the short and long-run policy effects.³¹ We consider two scenarios. First, we consider a child care policy that shifts ma-

²⁹Results not reported for space considerations.

³⁰See Yamaguchi, Asai, and Kambayashi (2018), Haeck, Lebihan, Lefebvre, and Merrigan (2022), Baker, Gruber, and Milligan (2008), and Datta Gupta and Simonsen (2010) among others.

³¹Our goal is not to externally validate our model by comparing predicted treatment effects to those in the literature, but instead to demonstrate how using contaminated measures can bias treatment effect estimates. Because of this, it does not make sense to examine treatment effects generated by the parent-only, teacher-only, and naive models since the noncognitive skill measures are uncontaminated by assumption.

ternal noncognitive skills by one standard deviation in the first period, holding fixed the initial level of all other child and parental skills. In the context of the above experiment, this would mean that child care has no direct impact on the child, but instead influences maternal noncognitive skills during the period of daycare through, for example, a labor supply response. However, the policy will have a real impact on child noncognitive skills in all subsequent periods through the skills transition function (the indirect effect). Importantly, the change in maternal noncognitive skills will also impact *measures* of child non-cognitive skills (the direct effect). Second, we consider a policy where both maternal and child noncognitive skills increase in the first period by 20% of a standard deviation.³² This means that child care has a direct effect on both mother and child skills when the child is age 3. We then ask how much of the *measured* increase in child noncognitive skills across all ages is the result of contamination, that is, the direct effect of a change in maternal noncognitive skills on measures of child noncognitive skills.

Figures 3 and 4 (corresponding to policy experiments 1 and 2) illustrate that contamination in child noncognitive skills measures can pose a serious threat to policy evaluations. Each panel in the figures represents a simulated SDQ measure of child

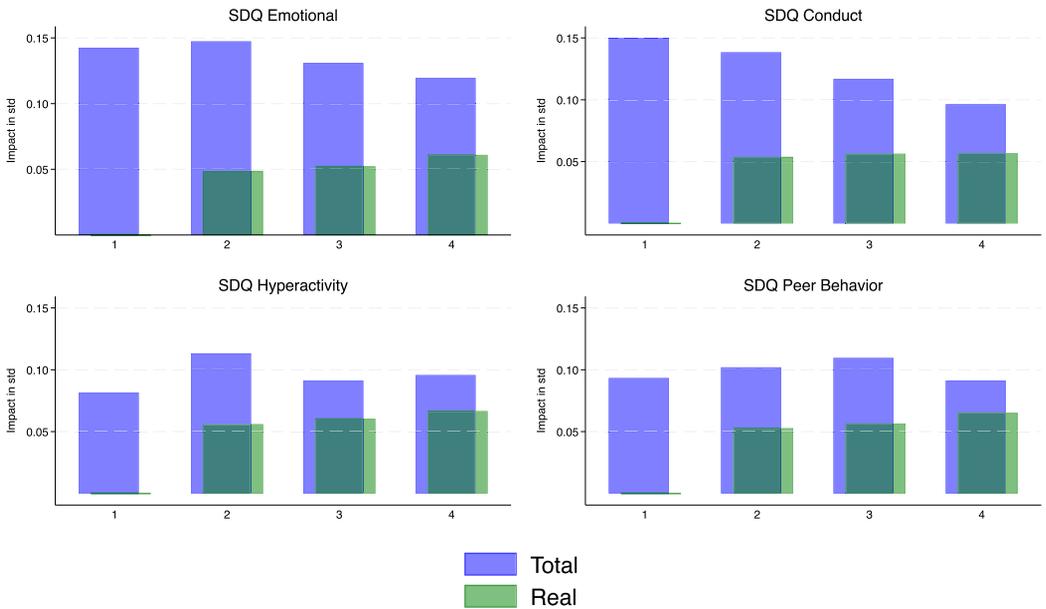


FIGURE 3. Policy evaluation with contamination: Change in maternal skills only. *Notes:* UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in Sections 4 and 5. The blue bars indicate the increase in each SDQ measure in different periods after we increase maternal noncognitive skills by 1sd at $t = 1$. The green bars indicate the same impact after we have removed the contamination from those measures.

³²While a large effect, it is similar in magnitude to estimates in Baker, Gruber, and Milligan (2008). Here, the authors find that the introduction of universal child care subsidies in Quebec increased child care use and maternal labor supply, with negative impacts on maternal well-being and child emotional and social development.

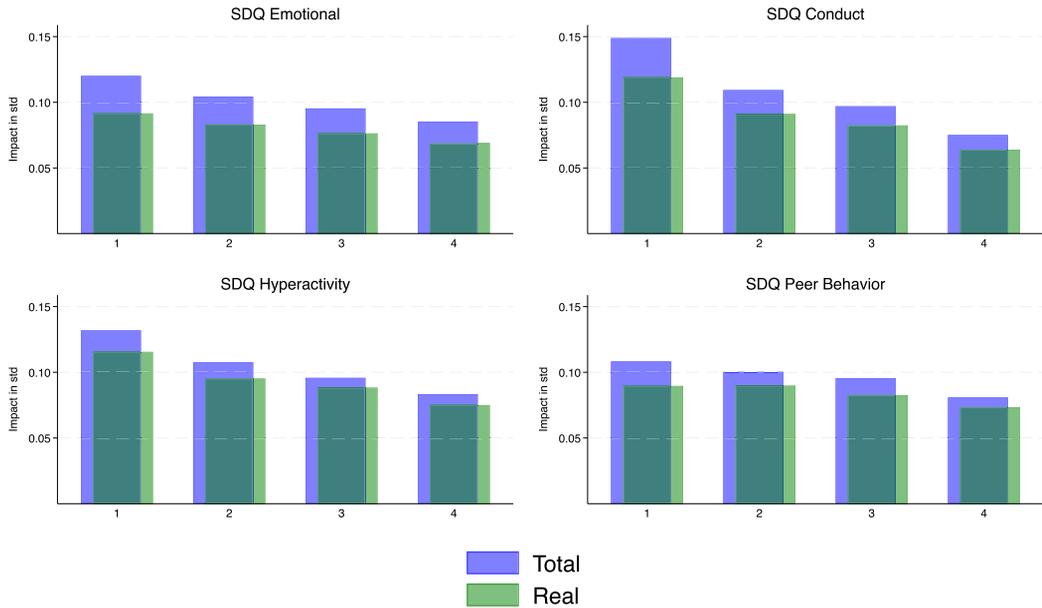


FIGURE 4. Policy evaluation with contamination: Change in maternal and child skills. *Notes:* UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in Sections 4 and 5. The blue bars indicate the increase in each SDQ measure in different periods after we increase maternal noncognitive skills and children noncognitive skills by $0.2sd$ at $t = 1$. The green bars indicate the same impact after we have removed the contamination from those measures.

noncognitive skill, mimicking the exact measures we observe in the data. We focus on the SDQ measures because these are among the most commonly used measures in the literature.³³ Variation along the x-axis of each panel reflects the different waves when child noncognitive skill is measured. The blue bars represent the estimated effect of the policy as measured by average treatment-control differences in maternal reported child noncognitive skills. These differences include contamination arising from maternal noncognitive skills. The green bars instead represent what the estimated effect of the policy would be if researchers could net out the contamination from maternal noncognitive skills.

For the policy experiment where only maternal skill is directly affected, the contamination effects are large. Figure 3 illustrates that the effect on child SDQ measures in the first period are on the order of 0.075 to 0.15 standard deviations. However, these effects are entirely spurious. Any increase in maternal non-cognitive skill can only influence child noncognitive skill in subsequent periods by construction. Thus, starting in period 2 there will be a real effect that works through the production technology. For example, in period 2 the SDQ emotional measure increases by 0.147 standard deviations, but only a third of the effect is real. In some cases, it is not only the magnitude, but the dynamic patterns that are affected by contamination. For SDQ conduct, the evolution of the mea-

³³In SA Tables 3 and 4, we present estimates for all child noncognitive skill measures.

sure indicates that the policy effect fades out over time, but the real change is fairly constant. Ultimately, the purpose of this first exercise is to show the extent to which policy estimates can be biased as the result of contaminated child non-cognitive skill measures. According to our model, a standard deviation increase in maternal noncognitive skills can translate into a 0.15 of a standard deviation spurious increase in child noncognitive skills.

Figure 4 illustrates the impact of a policy in which both maternal and child noncognitive skills increase by 0.2 of a standard deviation in the first period. This policy is motivated by the findings in Baker, Gruber, and Milligan (2008), who found similar sized effects on maternal and child noncognitive skills.³⁴ Looking again at SDQ emotional, we see that in the first period the measure increases by approximately 0.12 standard deviations.³⁵ However, the green bar indicates that almost a quarter of this effect is the result of contamination. Across all periods and measures, we find that the size of the distortion ranges between 3% and 24%. More generally, the direction of the distortion will depend on whether the policy under consideration affects mothers and children in the same direction. If, for example, a universal child care policy negatively affects the noncognitive skills of mothers, but positively affects those of the children, the distortion would lead to an underestimate of the policy on child noncognitive skill.

6.2.1 Correlation between parent and teacher reports To generate the above policy experiments, we estimated our model relying on the assumption that contamination in teacher-reported measures of child noncognitive skills is independent of parent skills and contamination in teacher measures in other periods. We can assess the robustness of our model and counterfactuals to these assumptions by imposing correlations between teacher effects and parent skills or between teacher effects across time. As discussed in Section 4.2, direct communication between these survey respondents is likely to lead to a positive correlation in teacher effects across time or between teacher effects and parental skills. We separately impose pairwise correlations of 0.1 and 0.5 across various contamination sources and reestimate our model and policy effects.³⁶ The results of this exercise are displayed in Table 13.³⁷

The first two columns in the top and bottom panels show the average increase in SDQ measures and the proportion driven by contamination across periods when maternal and child noncognitive skills are both boosted by 0.2 standard deviations in the first period using our baseline model. The remaining columns illustrate the same counterfactual, but allow for correlated contamination between maternal noncognitive skill and

³⁴Baker, Gruber, and Milligan (2008) find that the policy had negative impacts on both mothers and children. For illustrative purposes, we find it convenient to reverse the sign of the impacts.

³⁵The increase is less than 0.2 due to measurement error, as the SDQ emotional is a noisy measure of the child's true underlying skill. Evaluating any policy, including those that do not directly affect maternal noncognitive skill, will suffer from this measurement problem.

³⁶Allowing for correlations across pairs of variables makes it easy to target the same level of relatedness.

³⁷We could also model the effect of correlation in parent and teacher reports on the impulse response estimates shown in Tables 10 and 11. However, we prefer to use treatment effect contamination as our metric of interest since it combines both how the production function parameters change and how the measurement parameters change.

TABLE 13. Policy effect contamination with correlated factors.

t	Baseline		Corr(N_{it}^M, T_{it}) = 0.1		Corr(C_{it}^M, T_{it}) = 0.1		Corr($T_{it}, T_{it'}$) = 0.1	
	Avg. SDQ	%Contam.	Avg. SDQ	%Contam.	Avg. SDQ	%Contam.	Avg. SDQ	%Contam.
1	0.1273	18.4%	0.1277	18.7%	0.1300	17.4%	0.1276	18.3%
2	0.1054	14.8%	0.1061	20.0%	0.1074	14.1%	0.1060	14.4%
3	0.0959	14.0%	0.0967	18.4%	0.0979	13.6%	0.0960	13.9%
4	0.0811	13.3%	0.0819	18.0%	0.0828	12.7%	0.0810	13.0%

t	Baseline		Corr(N_{it}^M, T_{it}) = 0.5		Corr(C_{it}^M, T_{it}) = 0.5		Corr($T_{it}, T_{it'}$) = 0.5	
	Avg. SDQ	%Contam.	Avg. SDQ	%Contam.	Avg. SDQ	%Contam.	Avg. SDQ	%Contam.
1	0.1273	18.4%	0.1284	19.4%	0.1290	17.7%	0.1272	17.9%
2	0.1054	14.8%	0.1068	22.7%	0.1072	13.7%	0.1058	13.7%
3	0.0959	14.0%	0.0949	22.2%	0.0979	13.4%	0.0955	13.7%
4	0.0811	13.3%	0.0796	26.9%	0.0828	12.0%	0.0806	11.5%

Note: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in Sections 4 and 5. The Avg. SDQ columns indicate the average increase across SDQ measures in different periods after we increase maternal and child noncognitive skills by 0.2 SD at $t = 1$. The %Contam. columns indicates the share of the increase that is the result of contamination in the child noncognitive skill measures. The baseline results correspond to Figure 4. The remaining columns result from estimates of the model that allow for correlations between maternal skills and teacher effects and teacher effects across waves. See Sections 4.2 and 6.2 for further discussion.

teachers, maternal cognitive skill and teachers, and teachers over time.³⁸ When teacher assessments are directly influenced by maternal noncognitive skills, the distortionary effects are larger than in our baseline model. This occurs since more of the covariance between maternal and teacher measures is absorbed by contamination as opposed to child noncognitive skill. As a result, maternal noncognitive skill explains more of the variation in child noncognitive skill measures, leading to greater contamination in our policy exercises. In other words, our model that assumes a zero correlation is likely understating the degree of contamination. When teacher effects are related to maternal cognitive skill (columns 5 and 6) or to teacher effects in other periods (columns 7 and 8), the estimated policy effects and degree of contamination are mostly unchanged. While it is true that more of the variability in child noncognitive skill measures is driven by contamination, our policy exercises are not explicitly targeting maternal cognitive skill or teacher effects.

7. CONCLUSION

Researchers are forced to rely upon externally reported measures of child non-cognitive skill when studying skill formation since small children are not capable of assessing their own behaviors and emotional well-being. However, external evaluators bring their

³⁸We also investigated correlations between teacher effects and child noncognitive skill and cognitive skill, potentially arising from sorting. For noncognitive skill, contamination in the policy effects tend to increase, while for cognitive skill the policy effects remain constant. Finally, we allowed for a correlation between maternal noncognitive skill and the interviewer effect. Contamination plays a larger role in the policy effect in the initial period, which tapers away in later periods.

own skills and traits to these evaluations, potentially contaminating measures of child noncognitive skills. In this paper, we show that contamination in measures of child noncognitive skills can substantially affect our basic understanding of child skill dynamics. Additionally, when parental skills contaminate measures of child noncognitive skill, it is difficult to evaluate the effect of policies, which affect both children and parents.

A key finding of the current paper is that having multiple evaluators is critical to mitigating contamination issues. There are a number of available, longitudinal datasets that have this feature, including the 1970 British Cohort Study (BCS70), the Avon Longitudinal Study of Parents and Children (ALSPAC), and the Longitudinal Study of Australian Children, among others. Going forward, data collection efforts aiming to study child development should include evaluations of the child beyond just those provided by parents. Alternatively, surveys could collect more objective measures of child noncognitive skills related to observed behaviors. Of course, many historical data sets lack multiple informants, or only have multiple informants for a subset of observations within or across waves. When only one informant is available, we suggest doing significant robustness exercises around subjective measures, building in correlations that would otherwise be ruled out. If multiple informants are available for a subset of the sample, imposing assumptions on the nature of contamination across observations or time is likely to be a fruitful approach.

Rather than rely on multiple informants to correct for measurement error in subjective measures of noncognitive skills, researchers can also employ incentivized experiments or vignettes. As an example, [FKPS+ \(2021\)](#) use a series of incentivized experiments to measure time preferences, risk preferences, and altruism for children ages 7–9. Incentivized experiments allow researchers to construct non-cognitive skills directly from revealed preferences in a well-defined and controlled context. Similar to experiments, the use of anchoring vignettes can minimize measurement error by identifying heterogeneity in scale use. [Coenen, Golsteyn, Stolp, and Tempelaar \(2021\)](#) use this approach to estimate the impact of personality traits on student performance in higher education. In our setting, this tool would be helpful for correcting for heterogeneity in scale use among teachers. Interest in the development of human capital is unlikely to wane and developing new techniques to address measurement concerns is a fruitful area for additional research.

REFERENCES

- Agostinelli, Francesco and Matthew Wiswall (2025), “Estimating the technology of children’s skill formation.” *Journal of Political Economy*, 133 (3), 846–887. [0137, 0155, 0157]
- Ahun, Marilyn N., Claire Gapare, Geneviève Gariépy, and Sylvana M. Côté (2021), “Sex differences in the association between maternal depression and child and adolescent cognitive development: A systematic review and meta-analysis.” *Psychological Medicine*, 51 (9), 1431–1440. [0161]
- [ACFM+] Atanasio, Orazio, Sara Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina (2020), “Estimating the production function for human capital: Results

from a randomized control trial in Colombia.” *American Economic Review*, 110 (1), 48–85. [0136, 0137, 0138, 0139, 0155, 0157]

Attanasio, Orazio, Costas Meghir, and Emily Nix (2020), “Human capital development and parental investment in India.” *Review of Economic Studies*, 87 (6), 2511–2541. [0137, 0155]

Aucejo, Esteban M. and Jonathan James (2021), “The path to college education: The role of math and verbal skills.” *Journal of Political Economy*, 129 (10), 2905–2946. [0157]

Baker, Michael, Jonathan Gruber, and Kevin Milligan (2008), “Universal child care, maternal labor supply, and family well being.” *Journal of Political Economy*, 116 (4), 709–745. [0138, 0162, 0163, 0165]

Baker, Michael and Kevin Milligan (2015), “Maternity leave and children’s cognitive and behavioral development.” *Journal of Population Economics*, 28 (2), 373–391. [0136]

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko (2020), “Maternal depression, women’s empowerment, and parental investment: Evidence from a randomized controlled trial.” *American Economic Review*, 110 (3), 824–859. [0139]

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2016), “Does grief transfer across generations? Bereavements during pregnancy and child outcomes.” *American Economic Journal: Applied Economics*, 8 (1), 193–223. [0139]

Bracken, Barbara (2002), “Bracken school readiness assessment.” The Psychological Corporation, San Antonio, TX. [0141]

Campbell, Donald T. and Donald W. Fiske (1959), “Convergent and discriminant validation by the multitrait-multimethod matrix.” *Psychological Bulletin*, 56 (2), 81–105. [0139]

Choe, Daniel, Sheryl Olson, and Arnold Sameroff (2014), “Effortful control moderates bidirectional effects between children’s externalizing behavior and their mothers’ depressive symptoms.” *Child Development*, 85, 643–658. [0139]

Coenen, Johan, Bart Golsteyn, Tom Stolp, and Dirk Tempelaar (2021), “Personality traits and academic performance: Correcting self-assessed traits with vignettes.” *PLoS One*, 16 (3). [0167]

Conti, Gabriella, James Heckman, and Rodrigo Pinto (2016), “The effects of two influential early childhood interventions on health and healthy behaviour.” *Economic Journal*, 126, F28–F65. [0136]

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov (2006), “Interpreting the evidence on life cycle skill formation.” In *Handbook of the Economics of Education*, Vol. 1 (Eric A. Hanushek and Finis R. Welch, eds.), 697–812, Elsevier, Amsterdam. Chapter 12. [0135]

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach (2010), “Estimating the technology of cognitive and noncognitive skill formation.” *Econometrica*, 78 (3), 883–931. [0136, 0138, 0154, 0155, 0160]

Datta Gupta, Nabanita and Marianne Simonsen (2010), “Non-cognitive child outcomes and universal high quality child care.” *Journal of Public Economics*, 94 (1), 30–43. [0162]

Davis, Naomi Ornstein, and Alice S. Carter (2008), “Parenting stress in mothers and fathers of toddlers with autism spectrum disorders: Associations with child characteristics.” *Journal of Autism and Developmental Disorders*, 38 (7), 1278. [0139]

[DLRAWT+] De Los Reyes, Andres, Tara M. Augenstein, Mo Wang, Sarah A. Thomas, Deborah A. G. Drabick, Darcy E. Burgers, and Jill Rabinowitz (2015), “The validity of the multi-informant approach to assessing child and adolescent mental health.” *Psychological Bulletin*, 141 (4), 858–900. [0138]

Del Bono, Emilia, Marco Francesconi, Yvonne Kelly, and Amanda Sacker (2016), “Early maternal time investment and early child outcomes.” *The Economic Journal*, 126 (596), F96–F135. [0136]

Del Bono, Emilia, Josh Kinsler, and Ronni Pavan (2022), “Identification of dynamic latent factor models of skill formation with translog production.” 37 (6), *Journal of Applied Econometrics*, 1256–1265. [0156, 0157, 0158]

Del Bono, Emilia, Josh Kinsler, and Ronni Pavan (2026), “Supplement to ‘Skill formation and the trouble with child noncognitive skill measures.’” *Quantitative Economics Supplemental Material*, 17, <https://doi.org/10.3982/QE2297>. [0143]

Deming, David (2009), “Early childhood intervention and life-cycle skill development: Evidence from head start.” *American Economic Journal: Applied Economics*, 3 (1), 111–134. [0136]

Elliott, Colim, Pauline Smith, and Kay McCulloch (1996), *British Ability Scales Second Edition (BAS II). Administration and Scoring Manual*. Nelson, London. [0141]

[FKPS+] Falk, Armin, Fabian Kosse, Pia Pinger, Hannah Schildberg-Horisch, and Thomas Deckers (2021), “Socioeconomic status and inequalities in children’s IQ and economic preferences.” *Journal of Political Economy*, 129 (9), 2504–2545. [0167]

Feng, Shuaizhang, Yujie Han, James J. Heckman, and Tim Kautz (2022), “Comparing the reliability and predictive power of child, teacher, and guardian reports of noncognitive skills.” *Proceedings of the National Academy of Sciences*, 119 (6), e2113992119. [0139]

Goodman, Anna, Donna Lamping, and George Ploubidis (2010), “When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the strengths and difficulties questionnaire (sdq): Data from British parents, teachers and children.” *Journal of Abnormal Child Psychology*, 38, 1179–1191. [0142]

Goodman, Robert (1997), “The strengths and difficulties questionnaire, a research note.” *Journal of Child Psychology and Psychiatry*, 38, 581–586. [0142]

Haeck, Catherine, Laetitia Lebihan, Pierre Lefebvre, and Philip Merrigan (2022), “Child-care, parental behavior and well-being during childhood.” *The B.E. Journal of Economic Analysis & Policy*, 22 (2), 311–360. [0138, 0162]

Hastings, Richard P. (2002), “Parental stress and behaviour problems of children with developmental disability.” *Journal of Intellectual & Developmental Disability*, 27 (3), 149–160. [0139]

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006), “The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior.” *Journal of Labor Economics*, 24 (3), 411–482. [0136]

Hernández-Alava, Mónica and Gurleen Popli (2017), “Children’s development and parental input: Evidence from the uk millennium cohort study.” *Demography*, 54 (2), 485–511. [0136]

Johnston, David, Carol Propper, Stephen Pudney, and Michael Shields (2014), “Child mental health and educational attainment: Multiple observers and the measurement error problem.” *Journal of Applied Econometrics*, 29, 880–900. [0139]

Joreskog, Karl Gustav (1971), “Statistical analysis of sets of congeneric tests.” *Psychometrika*, 36, 109–133. [0139]

[KBCE+] Kessler, Ronal, Peggy Barker, Lisa Colpe, Joan Epstein, Joseph Gfroerer, Eva Hiripi, Mary Howes, Sharon-Lise Normand, Ronald Manderscheid, Ellen Walters, and Alan Zaslavsky (2003), “Screening for serious mental illness in the general population.” *Archives of General Psychiatry*, 60, 184–189. [0143]

[KMAE+] Kraemer, Helena, Jeffrey Measelle, Jennifer Ablow, Marylin Essex, Thomas Boyce, and David Kupfer (2003), “A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives.” *American Journal of Psychiatry*, 160 (9), 1566–1577. [0138]

Kuhn, Jennifer C. and Alice S. Carter (2006), “Maternal self-efficacy and associated parenting cognitions among mothers of children with autism.” *The American journal of Orthopsychiatry*, 76, 564–575. [0139]

Kvist, Anette Primdal, Helena Skyt Nielsen, and Marianne Simonsen (2013), “The importance of children’s adhd for parents’ relationship stability and labor supply.” *Social Science & Medicine*, 88, 30–38. [0139]

Martel, Michelle, Kristian Markon, and Gregory Smith (2017), “Research review: Multi-informant integration in child and adolescent psychopathology diagnosis.” *Journal of Child Psychology and Psychiatry*, 58, 116–128. [0138]

[MHQS+] Melhuish, Edward, Karen Hanna, Louise Quinn, Kathy Sylva, Iram Siraj-Blatchford, Pam Sammons, and Brenda Taggart (2004), *The Effective Pre-School Provision in Northern Ireland Project. Technical Paper 11: Pre-School Experience and Social/Behavioural Development at the End of Year 3 of Primary School*. Stranmillis University Press, Belfast. [0142]

Moffitt, Terrie, Louise Arseneault, Daniel Belsky, and Avshalom Caspi (2011), “A gradient of childhood self-control predicts health, wealth, and public safety.” *Proceedings of the National Academy of Sciences*, 108, 2693–2698. [0135, 0136]

[NWNS+] 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. [0137]

Persson, Petra and Maya Rossin-Slater (2018), “Family ruptures, stress, and the mental health of the next generation.” *American Economic Review*, 108, 1214–1252. [0139, 0161]

Plewis, Ian (2007), “The Millennium Cohort Study: Technical report on sampling.” Working Paper, CLS, London. [0139]

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff (2003), “Common method biases in behavioral research: A critical review of the literature and recommended remedies.” *Journal of Applied Psychology*, 88 (5), 879–903. [0139]

[PSMS+] Prochaska, Judith, Hai-Yen Sun, Wendy Max, Yanling Shi, and Michael Ong (2012a), “Validity study of the k6 scale as a measure of moderate mental distress based on mental health treatment need and utilization.” *International Journal of Methods in Psychiatric Research*, 21, 88–97. [0144]

[PSMS+] Prochaska, Judith J., Hai-Yen Sung, Wendy Max, Yanling Shi, and Michael Ong (2012b), “Validity study of the k6 scale as a measure of moderate mental distress based on mental health treatment need and utilization.” *International Journal of Methods in Psychiatric Research*, 21 (2), 88–97. [0147]

[RJOS+] Robbins, Trevor W., Merle James, Adrian M. Owen, Barbara J. Sahakian, Andrew D. Lawrence, Lynn McInnes, and Patrick Rabbitt (1998), “A study of performance on tests from the cantab battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. Cambridge neuropsychological test automated battery.” *Journal of the International Neuropsychological Society*, 4, 474–490. [0141]

Ronda, Victor (2017), “The effect of maternal psychological distress on children’s cognitive development.” Unpublished manuscript. [0139]

Sawyer, Michael G., David L. Streiner, and Peter Baghurst (1998), “The influence of distress on mothers’ and fathers’ reports of childhood emotional and behavioral problems.” *Journal of Abnormal Child Psychology*, 26 (6), 407–414. [0137]

[SBBM+] Sevim, Dilek, Victoria Baranov, Sonia Bhalotra, Joanna Maselko, and Pietro Biroli (2024), “Trajectories of early childhood skill development and maternal mental health.” *Journal of Human Resources*, 59 (S), S365–S401. [0138]

Todd, Petra E. and Kenneth I. Wolpin (2007), “The production of cognitive achievement in children: Home, school, and racial test score gaps.” *Journal of Human Capital*, 1 (1), 91–136. [0135]

University College London, UCL Social Research Institute, Centre for Longitudinal Studies (2024), “Millenium Cohort Study.” 16th Release. UK Data Service. SN:2000031. [0136]

Yamaguchi, Shintaro, Yukiko Asai, and Ryo Kambayashi (2018), “How does early child-care enrollment affect children, parents, and their interactions?” *Labour Economics*, 55, 56–71. [0138, 0162]

Co-editor Limor Golan handled this manuscript.

Manuscript received 5 December, 2022; final version accepted 3 November, 2025; available online 8 December, 2025.

The replication package for this paper is available at <https://doi.org/10.5281/zenodo.17290197>. The authors were granted an exemption to publish their data because either access to the data is restricted or the authors do not have the right to republish them. Therefore, the replication package only includes the codes but not the data. However, the authors provided the Journal with (or assisted the Journal to obtain) temporary access to the data. The Journal checked the restricted data and the provided codes for their ability to reproduce the results in the paper and approved online appendices. Given the highly demanding nature of the algorithms, the reproducibility checks were run on a simplified version of the code, which is also available in the replication package.

All authors assume responsibility for all aspects of the paper.