I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13713

Skill Formation and the Trouble with Child Non-Cognitive Skill Measures

Emilia Del Bono
Josh Kinsler
Ronni Pavan

SEPTEMBER 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13713

Skill Formation and the Trouble with Child Non-Cognitive Skill Measures

Emilia Del Bono
ISER, University of Essex and IZA

Josh Kinsler
University of Georgia

Ronni Pavan
University of Rochester

SEPTEMBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Skill Formation and the Trouble with Child Non-Cognitive Skill Measures*

Research on child skill formation and related policies typically rely on parent- reported measures of child non-cognitive skills. In this paper, we show that parental assessments of child non-cognitive skills are directly affected by the skills of the parents. We develop a dynamic model of child and parental skill formation that accounts for this contamination and show how standard estimates of the production of skills are affected. We then use our model to illustrate how contamination in parental measures of child non-cognitive skills affects estimates of child development policies that also directly affect parental skills.

JEL Classification: C13, C18, I38, J13, J24

Keywords: children, human capital, dynamic factor analysis, measurement, policy

Corresponding author:
Ronni Pavan
Department of Economics
University of Rochester
Harkness Hall
Rochester, NY 14627
USA
E-mail: 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 no. 1725270. Emilia Del Bono also received support from the Research Centre on Micro-Social Change (MiSoC), award grant ES/S012486/1. The authors are grateful to the Centre for Longitudinal Studies (CLS), UCL Institute of Education, for the use of these data and to the UK Data Service for making them available. However, neither CLS not the UK Data Service bear any responsibility for the analysis or the interpretation of this data.

1 Introduction

Many studies have shown that human capital skills in early childhood are strong predictors of important adult outcomes such as completed education, earnings, and health (Todd & Wolpin, 2007; Cunha *et al.* , 2006; Moffitt *et al.* , 2011; Conti *et al.* , 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, non-cognitive skills.[1] Non-cognitive skills aid in the development of cognitive skills throughout early childhood and directly impact labor market outcomes (Cunha *et al.* , 2010; Agostinelli & Wiswall, 2020; Attanasio *et al.* , 2020; Heckman *et al.* , 2006). Yet, how do we measure child non-cognitive 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 skill formation relies on household surveys where parents assess the emotional and social behavior of their child.[2] These assessments are naturally subjective and can be influenced not only by the non-cognitive skill of the child but also by the skills and traits of the parents. The trouble with relying on parent-reported measures of child skills is most acute in a dynamic setting, where parental skills, the source of contamination, can be seen as a direct input in the production of future child skills. In these models, an important parameter is the effect of parental skills on child skills in the next period. The interpretation of this parameter becomes more challenging if child non-cognitive skills are affected by parental skills simply due to the way they are measured.

The possibility that parental skills influence measures of child non-cognitive skills has

[1] See Deming (2009) or Baker & Milligan (2015) for example.

[2] This 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 *et al.* , 2010), as well as the UK, where researchers have mainly relied on the Millennium Cohort Study (Del Bono *et al.* , 2016; Hernández-Alava & Popli, 2017).

important implications beyond skill production estimates. Consider a policy such as the introduction of universal child care. Assume this policy has no effect on child skills but does impact parental skills, such as maternal mental health (Haeck *et al.* , 2019; Yamaguchi *et al.* , 2018; Baker *et al.* , 2008); assume also that parental skills have no impact on future child skills. If parental skills directly influence the *measures* of child non-cognitive skills available to the researcher, even a thorough and well-executed evaluation will indicate an impact of the policy on child non-cognitive development when there is none. Even when child care policies impact child skills directly, the same problem may occur with the result that our understanding of a policy's impact on children can be seriously influenced by the way in which the policy also affects parents.

In this paper, we develop a model that produces consistent estimates of the true distribution and evolution of child skills even when parental measures of child non-cognitive skills are contaminated. By contamination we mean that parental measures of child non-cognitive skills are influenced by the parents' own skills and traits. Our estimation strategy relies on the availability of multiple evaluators of child non-cognitive skills. For this reason we turn to the Millennium Cohort Study (MCS). The MCS is a large prospective study of infants born between 2000 and 2002 in the United Kingdom. Successive interviews took place when the children were 3, 5, 7, 11, 14 and 17 years old, and include cognitive and non-cognitive assessments of child development.[3] Specifically, the non-cognitive 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 non-cognitive skills measures at each wave. This allows us to jointly model the dynamics of child and parental skills and to consider the implications of contamination in a setting where it is most relevant. The data also offers the opportunity to examine the role of fathers in child skill formation and to analyze the importance of feedback effects running from child non-cognitive skills to parental non-cognitive skills.

We begin with descriptive analysis that illustrates the subjective nature of child non-

[3] We focus on surveys between the ages of 3 and 11.

cognitive skill measures. We first look at the contemporaneous correlation between mother-reported child non-cognitive skills and maternal non-cognitive skills and compare it with the correlation between teacher (or interviewer) reported child non-cognitive skills and maternal non-cognitive skills. We show that maternal non-cognitive skill measures are highly predictive of child non-cognitive skill measures only when using child non-cognitive skill measures *reported by the mother*. By contrast, maternal non-cognitive skills have significantly less predictive power when using teacher (or interviewer) reported measures of child non-cognitive skill. Similarly, maternal non-cognitive skills are strong predictors of the child non-cognitive skill measures only when the mother, not the father, reports the child's non-cognitive skill. When the father is the main survey respondent, it is the father's non-cognitive skill measures that best predict the child non-cognitive skills. These patterns suggest that the identity and attributes of the child's evaluator may significantly impact measures of child non-cognitive skills.

While the descriptive analysis reveals that contamination is a concern, it is not well-suited to solving the problem. To address contamination, we estimate a model of skills formation similar in spirit to Cunha *et al.* (2010) and Agostinelli & Wiswall (2020). The key innovation in our setting is to relax some of the typical assumptions related to the measurement model of child non-cognitive skills. In particular, we allow parent, teacher, and interviewer reported measures of child non-cognitive skills to be contaminated. This contamination can be viewed as a component of skill measurement error that is potentially correlated across non-cognitive skill measures and, in the case of parents, with the measures of other skills. These features differentiate our measurement model from others in the literature, where it is assumed that there is no contemporaneous correlation in the measures other than through child non-cognitive skills (in Cunha *et al.* (2010), measurement error can be correlated across different time periods). We show that our model is identified under the assumption that contamination in the child non-cognitive skill measures reported by different evaluators are independent of each other.

We estimate the model using a two-step approach similar to Agostinelli & Wiswall (2020), Attanasio *et al.* (2020), Attanasio *et al.* (2019). Estimates of the measurement equations show that a significant share of the variation in parent-reported measures of child non-cognitive skills are driven by parental skills. Similarly, the measurement error in teacher reported measures of child non-cognitive skill are highly correlated within period. The estimated skill technology parameters indicate high levels of persistence in skill formation, a significant role of child non-cognitive skills on cognitive skills, and a large impact of paternal cognitive skills on child cognitive skills. There is also evidence that child non-cognitive skills influence the evolution of parental mental health.

To understand how contamination impacts estimates of skill production, we compare the results of our preferred model with two alternatives. The first alternative uses only parental measures of child non-cognitive skills, where it is not possible to allow for contaminated measures since only one evaluator is available. The second alternative takes into account both parent-reported and teacher (or interviewer) reported measures but assumes there is no contamination in any measure. We find important differences between our preferred specification and the two alternatives. For example, we find that the persistence of child non-cognitive skills is 15% higher when only parent measures of child non-cognitive skills are used. This occurs because maternal and paternal skills are themselves highly persistent and these skills contaminate estimates of child skills through the measures in all periods.

In addition to affecting estimates of child skill dynamics, relying on contaminated measures of child non-cognitive skills can have broader implications in terms of policy evaluation. If a policy or program influences parental skills and parental skills contaminate the available *measures* of child non-cognitive skills, estimates of the policy effect on child skills could be partly spurious. In the final section of the paper we illustrate this point through two simulation exercises. We first implement a hypothetical policy where only initial maternal non-cognitive skills are increased by one standard deviation. Child non-cognitive skill measures increase by as much as 15% of a standard deviation in the initial period but this is only a result of contamination. Over time increased maternal skills directly impact child non-cognitive skills, but even so around half of the effect after eight years is attributable to contamination. In a second simulation exercise, we consider a hypothetical policy where both initial maternal and child non-cognitive skills are increased by 20% of a standard deviation. This effect would be in line with the findings in Baker *et al.* (2008), where an increase in the availability of free child care in Quebec had a (negative) impact on both maternal and child non-cognitive skills of approximately the same size. Even in this case, contamination can be responsible for up to a quarter of the increase in measured child non-cognitive skills.

Our paper is the first to propose a methodology to tackle systematic error in measures of child non-cognitive skills in a dynamic model of skills formation. The existing literature in economics has largely ignored this issue (Cunha *et al.* , 2010; Attanasio *et al.* , 2020). One possible explanation for this are data constraints, since dealing with the contamination of parent reported measures of child non-cognitive 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 (Kraemer *et al.* , 2003; De Los Reyes *et al.* , 2015; Martel *et al.* , 2017), it is less readily available in large and representative surveys. One notable exception is the work conducted by Johnston *et al.* (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 outcomes. The survey provides measures of child mental health reported by parents, teachers, and the children themselves, who are all assumed to be biased informants. Here the setting is static and identification relies on the availability of diagnostic assessments from a panel of psychiatric experts, who are assumed not to be affected by systematic bias.

The idea of relying on multiple evaluators or multiple measurement methods to purge measures of contamination has a long history in the broader psychometric and applied statistics fields (Campbell & 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 *et al.* , 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 growing literature that explores the impacts of maternal mental health on children. A substantial part of this literature investigates maternal stress in pregnancy or during the post-natal period on child development using either quasi-natural experiments - which create exogenous variation in maternal psychological well-being (Black *et al.* , 2016; Persson & Rossin-Slater, 2018) - or randomised control trials (Baranov *et al.* , 2017). A contribution of our work is to also consider the potential impact of children on parental mental health, a link that has been mainly emphasized in medical studies (Kuhn & Carter, 2006; Davis & Carter, 2008; Choe *et al.* , 2014; Hastings, 2002).

The rest of the paper proceeds as follows. Sections 2 and 3 present the data and preliminary evidence that suggests how child non-cognitive skill measures may be contaminated. In Section 4 we present a model of skill dynamics meant to handle this contamination and discuss the necessary identification assumptions. Section 5 describes estimation and presents the results. Section 6 evaluates how contamination impacts policy evaluation. Section 7 concludes.

2 Data

The data for our analysis are from the Millennium Cohort Study (MCS). The MCS is a large prospective study of infants born between 2000 and 2002 in the United Kingdom. It is representative of the overall UK population of newborns.[4] The first wave of data collection took place when the infants were around 9 months old and includes data on 18,552 children. The sampling design allowed for over-representation of areas with high levels of childhood deprivation and high proportions of ethnic minorities (Plewis 2007).

At the first interview, the main respondent was asked about pregnancy, birth, infant health, infant development, their own mental health, health behavior and the family social and economic circumstances. Successive interviews took place when the children were 3, 5, 7, 11, 14 and 17 years old. Data for the present analysis is restricted to interviews up to age 11, i.e. the last year of primary school. During each sweep of the study, the children were administered a series of cognitive assessments by a specially trained interviewer. At age 5, 7 and 11 the study collected additional information about the child's academic ability and

[4] Infants born on eligible dates in eligible areas were selected from the Child Benefit Register, a universal benefit in the UK at the time.

socio-emotional development as reported by the main primary school teacher.

Sample selection — We operate a number of sample restrictions (see Table 1). First, we consider only singleton children and families where the initial interview was not obtained by proxy, starting with a sample of 19,048 children who are interviewed for the first time during the first or second wave of the study. We then focus on two-parent families, as our analysis takes into consideration maternal as well as paternal measures of cognitive and non-cognitive skills. This restricts the sample to 14,648 children/families. We then delete a few cases where there is missing information on: gender, ethnicity, maternal age and other basic family demographics (14,598 children). As cognitive outcomes start being recorded at age 3, we consider only families that appear in the study at least once from the second wave onwards (i.e. delete families that only reply to the first interview). Our main sample thus consists of a total of 12,530 children/families observed from age 3 onwards.

One important feature of our sample restrictions is that any observation where the parents separate is dropped from the sample as soon as the family becomes a single-parent family. As a result of this, as well as general panel attrition, our estimation sample includes slightly fewer than eight thousand children in the last wave. Children that drop out from our sample due to attrition or family separation are more likely to be male, less likely to be white, more likely to have siblings, and have less educated parents. Despite this, the regression analysis we present in Section 3 appears to be robust to sampling composition.[5]

Table 1 below shows basic demographic and family background variables for the initial sample of singleton births, the sub-sample of two-parent families, and our estimation sample. For ease of comparison we show descriptive statistics using information collected at the time of the earliest available interview (columns 1-3). This is conducted when the child was 9 months old, but for less than 5% of cases it takes place at age 3 as some new families were recruited into the study at wave 2. As we can see, 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.

[5]In 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.

There is however a significant difference with respect to socio-demographic characteristics. Mothers are usually older and more educated among the sample of two-parent families.[6] As we would expect, this positive selection is a bit more marked for our final estimation sample (column 4), as low socio-economic status families are less likely to respond to the second interview and stay in the panel. Paternal characteristics differ much less across these samples.

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 of the survey, at least two measures of child cognitive skills are available (see Table 2). The tests administered by the interviewer during the early years come mainly from the British Ability Scales or BAS (Elliott *et al.* , 1996). In addition, at age 3 the children were assessed according to some of the components of the Bracken School Readiness Assessment (Bracken, 2002), which is considered a good indicator of success in formal education. We also use information obtained at age 7 from a variant of the National Foundation for Educational Research (NFER) Progress in Mathematics test. At age 11, child cognitive skills are assessed through the Cambridge Gambling Task (CGT) and the Spatial Working Memory (SWM) task (Robbins *et al.* , 1998).[7]

Our measures of child cognitive skills also include teacher assessments. The first of these is the Early Years Foundation Stage Profile (FSP). This assessment is performed by the teacher during the reception year, when UK children are aged 5, and describes each child's development and learning achievements in the following areas: (i) personal, social and emotional development; (ii) communication, language and literacy; (iii) problem solving,

[6]Education is measured as years of schooling. This measure is obtained assuming that: (i) all individuals with no qualifications have 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 have 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.

[7]All 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.

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 information from a teacher questionnaire, which reports the teacher's evaluation of the MCS child across the main subjects (English, Math, and Science).

Child non-cognitive outcomes — At each interview, the main respondent - who is the mother of the child in the vast majority of cases, but could also be the father or another member of the family - is asked to complete the Child Strengths and Difficulties Questionnaire (SDQ). The SDQ is a behavioral screening questionnaire designed to measure psychological adjustment in children aged 3 to 16 (Goodman, 1997). It identifies five different components: (i) hyperactivity/inattention, (ii) conduct problems, (iii) emotional symptoms, (iv) peer problems, and (v) pro-social behavior. We take the first four sub-scores as our main measures of respondent-reported child non-cognitive skills. Another measure reported by the respondent is captured by selected items from the Child Social Behavior Questionnaire (CSB), measuring the child's ability to perform tasks independently, to concentrate, and to control his/her emotional responses (Melhuish *et al.* , 2004). A set of questions to measure child cooperative behavior was introduced at age 7 (see Table 3).

At age 3, the interviewer reported his/her own assessment of child behavior as part of a module aimed at providing information on the general conditions of the assessment. The interviewer was asked whether the child had been fidgety, focused, disruptive, etc., during the assessment and the interview. There are 10 questions in total. Two questions were aimed at capturing extreme behavior (i.e. child is dangerous, child is disruptive), others were combined using principal component analysis to obtain two additional summary scores, one centered on focus and attention, the other on the child's cooperative behavior.

Lastly, we consider teacher assessments of the child's social and emotional development. There is one area of the FPS assessment measured at age 5 (see above) which is relevant in this case: (i) personal, social and emotional development. This is further subdivided into three components: (a) dispositions and attitudes, (b) social development, and (c) emotional development. Each of them is scored between 0-9 and is considered separately.

At age 7 and 11, teachers are administered the teacher-version of the SDQ, from which we take four components: (i) hyperactivity/inattention, (ii) conduct problems, and (iii) emotional symptoms and (iv) peer problems. In each wave of the survey, at least three teacher and/or interviewer measures of child non-cognitive skills are available.

Maternal and paternal non-cognitive 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 (if present) were asked to provide answers to questions from a shortened version of the Kessler questionnaire, a screening device frequently used to diagnose mental illness (Kessler *et al.* , 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 — In the MCS there are relatively few high quality measures of respondent and partner cognitive skill. Respondents report their own and their partner's academic qualification, which we transform into years of schooling (see footnote 5 for more details). 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. Finally, in the age 14 MCS survey, respondents and partners are asked to complete a word task. The results of this word task are highly correlated with years of schooling, but are not used here as the word task is completed by only half of our estimation sample. 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 non-cognitive skill measures. We first extract the principal factor for child and parental skills in each wave using the measures discussed in the previous section.[8] The skill proxies we generate are 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 suggestive evidence of contamination in child non-cognitive skill measures that drive our modeling choices later on.

There are four types of individuals who report on a child's non-cognitive 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.[9] In every wave either the mother or father evaluates the child's non-cognitive skills. To illustrate that the identity of the survey respondent can influence the assessment of the child, we create separate child non-cognitive skill proxies for each type of respondent. We then explore how these evaluator-specific non-cognitive 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 non-cognitive skills, the resulting skill proxies should project similarly on the other skill proxies.

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

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

where $t = 1, ..., 4$ (corresponding to ages 3, 5, 7, and 11) and r corresponds to the person reporting the child's non-cognitive skill proxies, either the mother or the interviewer/teacher.[10] The key controls are the principal factors for child cognitive skill (Cog),

[8] As mentioned previously, parental cognitive skill is simply equal to parental years of schooling. The key patterns we discuss in this section are unchanged if we replace years of schooling with separate dummy variables for each level of academic qualification.

[9] In the UK, all children begin school by the age of 5.

[10] At age 3 we use interviewer responses, and at ages 5, 7, and 11 we use teacher responses.

maternal cognitive skill ($MotherCog$), and maternal non-cognitive skill ($MotherNonCog$).[11] In the first two columns, we report the coefficients that result when we use the child non-cognitive skill proxy generated using the mother's responses as the dependent variable. In the third and fourth columns we use the child non-cognitive skill proxy generated from the teacher/interviewer responses as the dependent variable.[12]

Child non-cognitive 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 non-cognitive skill, there is a strong relationship between child non-cognitive skills and mother non-cognitive skills. When we utilize the teacher/interviewer's measures of child non-cognitive skills, the coefficient on the mother non-cognitive skills declines by approximately 80%.[13] Additionally, the coefficient on the child cognitive skill increases significantly when we use the teacher/interviewer-generated child non-cognitive skill. While we cannot determine which respondent, mother or teacher, is closer to the truth, we can certainly conclude that the assessments differ significantly.[14]

One reason the assessments might differ, apart from contamination, is that parents and teachers are reporting about a different underlying latent skill. In fact, not all parent and teacher measures of child non-cognitive skills overlap. However, when we construct the child's non-cognitive skill factor using only those measures that are common between

[11] When 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.

[12] We 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 2nd and 4th 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.

[13] All skill measures are standardized to have a mean of zero and a standard deviation equal to one.

[14] We explore whether contamination is mainly driven by mothers with low levels of non-cognitive skills. In order to do so, we split the sample into above and below average maternal non-cognitive skill groups, or exclude mothers with moderate or severe cases of depression according to their Kessler score (see Prochaska *et al.* (2012)). In all cases the coefficient associated with maternal non-cognitive skill is large relative to the corresponding coefficient when teacher-reported measures of child non-cognitive skill are used as the dependent variable.

parents and teachers, we find the same patterns. Moreover, similar results are obtained when we replace the principal factor with the underlying SDQ measures.[15] Both results are in online Appendix Table 1, which shows quite clearly that contamination is relevant for many possible dimensions of children non-cognitive skills.

In Table 5 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 5 uses the mother-reported measure of child skills as the dependent variable, but here we add paternal cognitive and non-cognitive skills to the regression. Similar to the results from Table 4 we find that maternal non-cognitive skills are highly correlated with child non-cognitive skills. Paternal skills are also correlated, but the strength of this relationship is significantly weaker. The second column reports results from a similar regression but uses the teacher-reported measures of child non-cognitive skills as the dependent variable instead. As before, we see that the correlation between maternal skills and child non-cognitive skills is reduced, indeed now maternal and paternal skills are similarly related to child non-cognitive skills. The third column uses the father-reported measures of child non-cognitive skills. The striking result here is that father non-cognitive skills now appear to be strongly correlated to child non-cognitive skills.[16] The differences between the mother and father non-cognitive 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 non-cognitive skills of the reporting parent is always strongly

[15]Specifically, we combine the SDQ sub-scores on 'emotional symptoms' and 'peer problems' into a measure of *internalizing* behaviour, and the SDQ sub-scores on 'hyperactivity/inattention' and 'conduct problems' into a measure of *externalizing* behaviour (Goodman *et al.* , 2010; Moroni *et al.* , 2019).

[16]One 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 non-cognitive skill of the child and that of the father. However, we examined the link between child non-cognitive skill and parental non-cognitive skill for the *same* children in column (3) of Table 5 when the *mother* is the main respondent (in a different wave of the study). We found that the pattern in column one re-emerges, i.e. it is the mother non-cognitive skill that is most strongly related to the child non-cognitive skill.

related to the non-cognitive skills of the child suggests that parental measures of child non-cognitive 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 non-cognitive skills might also be influenced by teacher skills or characteristics, introducing a correlation across all teacher-reported measures within the same period. If the goal is to learn about skill evolution and dynamic skill complementarity, how should one proceed when potentially all of the available measures of child non-cognitive skills are flawed? In the next section we develop a framework that will allow us to address the contamination issues under specific assumptions. However, prior to presenting the model, we first illustrate that using any of the contaminated measures to estimate a simple skill technology will likely result in misleading findings.

Table 6 presents estimates of child non-cognitive skills transition functions using combinations of mother and teacher-reported measures of child non-cognitive skills. The first column shows that if a researcher were to use only measures of child non-cognitive skills reported by the mother, child non-cognitive skills would be highly persistent (the persistence coefficient is 0.617). Child cognitive skills beget non-cognitive skills but the magnitude of this effect is rather modest. Additionally, maternal cognitive skills do not appear to be important. In column (3) we estimate the same model, using teacher/interviewer-reported child non-cognitive skills as the dependent variable. Here we find that the persistence in child non-cognitive skill is significantly smaller, exhibiting a coefficient of 0.221. We also find that child and maternal cognitive skills are significantly more important for the evolution of child non-cognitive skills when we rely on the teacher-reported measures. Finally, columns (2) and (4) include both the mother and teacher-reported measures as inputs in the technology. The takeaway from these regressions is that both measures matter, regardless of whose measure of child non-cognitive skill we use as the outcome. This suggests that both evaluators are providing useful information.[17] However, depending on which measure of child non-cognitive skills a researcher mainly relies upon, the technology estimates can

[17]The correlation between parent and teacher proxies of child non-cognitive skills is 0.33. Among all the child and parental skill proxies, this is the highest pairwise correlation for both measures.

be quite different.[18]

The analysis in this section illustrates the inherent subjectivity of child non-cognitive skill measures and the potential impact that this type of systematic measurement error may have on estimates of the skills production function. However, no definitive conclusions can be reached at this point. As mentioned above, the skill proxies we use are noisy versions of the true underlying skills and the level of noise might differ across periods. Moreover, it remains unclear how much contamination there is in the parent-reported measures as opposed to the teacher-reported measures. The model we present in the next section proposes a new framework which is able to address these issues and allows us to quantify not only the level of contamination but also the consequences of ignoring it.

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.[19] Each child is characterized by cognitive and non-cognitive skills, which are unobserved (to the econometrician). Parents are characterized by observed cognitive skills and unobserved non-cognitive skills.[20] The skills of the child in household i are denoted by (C_{it}, N_{it}), where C indicates cognitive skill and N indicates non-cognitive skill. The cognitive and non-cognitive skills of the parents in household i are

[18]Though we do not report them here, we also examined how estimates of the transition function for child cognitive and mother non-cognitive skills vary according to the nature of the evaluator. For child cognitive skills, we find that mother non-cognitive skills become important only when we utilize the teacher-reported measures of child non-cognitive skills. This makes sense since the mother-reported measures of child non-cognitive skills capture some of her own non-cognitive skills. In the mother non-cognitive skill transitions, we find evidence of feedback from child to mother when using mother-reported measures of child non-cognitive skills. This feedback effect essentially disappears when we use instead the teacher-reported measures.

[19]We formulate the model in the context of a two-parent single-child household. For estimation we restrict to two-parent households and include controls for being the first born child.

[20]The 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.

given by (C_i^P, N_{it}^P), where $P \in \{M, F\}$ for maternal and paternal skills, respectively. All (child and parental) non-cognitive skills are assumed to evolve over time. Child cognitive skills also change over time, while parental cognitive skills are assumed to be constant.[21]

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

$$S_{i1} \sim F_1, \tag{1}$$

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

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

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 non-cognitive skills.[22]

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 non-cognitive skills and vice-versa. Maternal and paternal non-cognitive skills are allowed to evolve over time as a function of all other skills, including their child's cognitive and non-cognitive 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

[21]This assumption, mainly driven by a lack of time varying measures of parental cognitive skills, is justified by the fact that parents are in a stage of life where cognitive skills are fully formed and likely stable over time.

[22]In 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.

the production of skill. This is because we purge our skill measures of demographic and household variables (see footnote 20 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. While it might be important to analyze how demographics interact with unobservable skills, for computational convenience this is left for future research.

4.2 Skill Measures

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

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

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

where the above equations refer to measures of child cognitive skills and maternal and paternal non-cognitive skills.[23] 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 present suggestive evidence that parent-reported measures of child non-cognitive skills are likely to be correlated with other key variables, such as parental skills. It is also possible that parent-reported measures of child non-cognitive skills

[23] We first residualize all MCS skill measures (including the child non-cognitive 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.

are correlated over time. Thus, the measurement equations for parent-reported measures of child non-cognitive skills can be written as:

$$M^N_{Pijt} \quad = \quad \alpha^N_{P,1jt} N_{it} + \alpha^N_{P,2jt} C^P_i + \alpha^N_{P,3jt} N^P_{it} + \alpha^N_{P,4jt} \theta_i + \epsilon^N_{P,ijt}, \tag{5}$$

where C^P_i and N^P_{it} are parental cognitive and non-cognitive skills, and θ_i is a family-specific factor that is time invariant.[24] The latter produces a correlation across all 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.[25] Note that only one parent reports on a child's non-cognitive 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 non-cognitive 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:

$$M^N_{T,ijt} \quad = \quad \alpha^N_{T,1jt} N_{it} + \alpha^N_{T,2jt} T_{it} + \epsilon^N_{T,ijt}. \tag{6}$$

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.[26] The presence of T_{it} accounts for the idea that some

[24] We assume that any contamination in the child non-cognitive skill measures does not have a direct impact on skill accumulation. However, if parents act on their perceptions of child non-cognitive skills when choosing investment, contamination will affect the evolution of skills. This is not a problem for contamination driven by C^P_i and N^P_{it}, 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.

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

[26] An important assumption here is that the random teacher effect is independent of all other skills. One concern might be that a teacher's evaluation of child non-cognitive skills is contaminated by child

teachers are more or less likely to classify certain child behaviors as problematic conditional on the child's true non-cognitive skills. The independence of T_{it} across time periods is justified by the fact that a child's teacher typically changes every year.

While the measurement system we employ to identify the dynamic latent factor model is broadly similar to the ones used in the previous literature, what distinguishes our approach is the presence of a within-period correlation in measurement error across measures and the general lack of a dedicated measure of child non-cognitive skills. Although the lack of a dedicated measure is not necessarily a problem for identification (Carneiro *et al.* , 2003), the contemporaneous correlation in measurement error is trickier. Previous researchers have illustrated that it is possible to allow for correlated measurement error across periods (Cunha *et al.* , 2010) but not across contemporaneous measures.

In contrast, there is a large literature in psychometrics and applied statistics that focuses on how to address correlations across measures that reflect the evaluation method rather than the underlying latent skill or trait. The original multi-trait, multi-method framework (MTMM) proposed by Campbell & Fiske (1959) spawned an enormous literature aimed at estimating the true link between latent factors when the measures are potentially contaminated by method effects, including rater-specific biases. However, the most common statistical methods employed, correlated trait-correlated uniqueness (CTCU) and correlated trait-correlated methods (CTCM), assume that the underlying traits (or factors in our model) are uncorrelated with rater-specific effects (see Podsakoff *et al.* (2003) for a lengthy review). This is clearly violated in our model since a key source of contamination in measures of child non-cognitive skill work through parental skills.

More recently, two MTMM type statistical models have been proposed that allow for

cognitive skills, since cognitive skills development is a primary focus of schooling. This would mean that T_{it} is correlated with C_{it}. However, in online Appendix Table 2 we show that the relationship between child cognitive and non-cognitive skills are very similar when we use teacher-reported or interviewer-reported measures of child non-cognitive skills. Interviewers are likely less focused on cognitive skills development, suggesting that our assumption is reasonable. A second source of dependence could arise through the sorting of teachers and students. If students are sorted to classrooms based on teachers' characteristics that are relevant for non-cognitive skill reports, our independence assumption would be violated. In online Appendix Table 3 we show that the link between child cognitive skills and teacher-reported non-cognitive skills is unaffected when we add observed or unobserved teacher characteristics (these are only available for period 3, i.e. when the child is 7 years old). This indicates that sorting is unlikely to be a concern.

the estimation of trait-method correlation: the latent difference (LD) and the latent means (LM) model (Pohl *et al.* , 2008; Pohl & Steyer, 2010). However, these models require strong assumptions on the scale of the available measures and must normalize one of the method effects to achieve identification. Our model does not require these assumptions, instead relying on the fact that the contamination that works through parental skills is also a latent factor for which we have dedicated measures. In a sense, our model brings together two separate approaches in the broader MTMM literature controlling for the contamination effects of a directly measured latent factor while also allowing this latent factor to be correlated with other latent factors. In addition, our model incorporates features of the CTCU model in that we allow for teacher-specific and household-specific biases in measures of child non-cognitive skills. In the next section we show how the availability of multiple evaluators of child non-cognitive skills allows us to identify the joint distribution of skills despite the various sources of contamination in these measures.

4.3 Identification

In this section, we briefly describe our approach to identification. A more formal treatment is provided in online Appendix A. 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_i. 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.[27] 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:

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

[27]As indicated above, a generalization that included unobserved parental cognitive skills would be straightforward in the presence of multiple measures of those skills.

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

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 *et al.* (2010), it is straightforward to show that the first and second moments of all unobserved skills *other than* the child non-cognitive 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,

$$Cov(M^C_{i1t}, M^C_{i1\tau}) = \sigma^C_{t,\tau} \qquad \text{for } t \neq \tau \tag{8}$$

$$Cov(M^C_{ijt}, M^C_{i1\tau}) = \alpha^C_{jt}\sigma^C_{t,\tau} \qquad \text{for } t \neq \tau \tag{9}$$

$$Cov(M^C_{ijt}, M^C_{i1t}) = \alpha^C_{jt}\sigma^C_t \tag{10}$$

$$Cov(M^C_{i1t}, M^{N^P}_{i1\tau}) = \sigma^{CN^P}_{t,\tau} \quad \text{for } P \in \{M, F\} \tag{11}$$

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

Identifying the second moments related to child non-cognitive skills is more challenging. All the measures related to child non-cognitive 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}.

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

$$\begin{aligned} Cov(M^N_{T,i1t}, M^N_{T,i1\tau}) &= \sigma^N_{t,\tau} \qquad \text{for } t \neq \tau \\ Cov(M^N_{T,ijt}, M^N_{T,i1\tau}) &= \alpha^N_{T,1jt}\sigma^N_{t,\tau} \qquad \text{for } t \neq \tau. \end{aligned}$$

These two observable quantities identify the covariance of child non-cognitive 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 non-cognitive skill. Additional details are provided in online Appendix A.

Once all the first and second moments of S_i have been identified, we show that the joint distribution of S_i is non-parametrically identified. This is crucial if we want to allow skills evolution to be non-linear. The proof broadly follows Theorem 1 of Cunha *et al.* (2010), though we modify it slightly to account for correlated errors in the child non-cognitive measures.

5 Estimation and Results

Following Attanasio *et al.* (2019), Attanasio *et al.* (2020), and Agostinelli & Wiswall (2020), 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.

5.1 Joint Distribution of Skill and Measurement Parameters

The joint distribution of child and parental skills across all time periods is governed by $f_t(\cdot)$ and $F_t(\cdot)$ as defined in the previous section. However, in the first step of the estimation process we do not take a stand on these two functions. Instead, we assume that all child and parental skills are jointly distributed according to a mixture of normal distributions. The flexibility of the mixture distribution puts few restrictions on the precise form that $f_t(\cdot)$ and $F_t(\cdot)$ ultimately take. We parameterize these functions in the second step of our estimation approach.

The joint distribution of child and parental skill takes the following form,

$$g(S_i) = \pi g_1(\gamma_1, \Sigma_1) + (1-\pi) g_2(\gamma_2, \Sigma_2), \tag{12}$$

where g_k is a normal density with mean γ_k and variance Σ_k for $k = (1, 2)$.[28] π is the mixture weight and determines the probability that a household is drawn from either $g_1(\cdot)$ or $g_2(\cdot)$. Σ_k is constrained to be a symmetric matrix with a positive main diagonal. There are 16 unobserved skills: child cognitive and non-cognitive in each of four periods, and maternal and paternal non-cognitive skill in each of four periods. There are also two observed skills, maternal and paternal cognitive skill. As a result, γ_k is a vector of 18 unknown means, while Σ_k is a matrix with 18 unknown variances, and 144 unknown covariances.[29] All of these parameters will be estimated in the first step.

In addition to estimating the parameters that govern the joint distribution of skill, the first step also yields estimates of the measurement equation parameters. The functional form for the measurement equations is provided in equations (3)-(6).[30] However, we also need to specify the distributions of the measurement errors (ϵ^C_{ijt}, ϵ^{NP}_{ijt}, $\epsilon^N_{P,ijt}$, and $\epsilon^N_{T,ijt}$ for $P \in \{M, F\}$) and the teacher (T_{it}) and parent (θ_i) random effects. The measurement errors are normally distributed with means equal to zero and variances to be estimated. The teacher and parent random effects are standard normals to set their scale without loss of generality. All of these components are assumed to be independent of each other and all skills.

[28] As discussed in the identification section, we need to normalize the mean of skill to zero. To do this we set $\gamma_2 = \frac{\pi}{1-\pi}\gamma_1$.

[29] Although we assume parental cognitive skill is observed, we still estimate the parameters of the two mixing distributions that give rise to the observed distribution.

[30] At $t = 2$, $t = 3$ and $t = 4$ (i.e. age 5, 7, and 11), the teacher also provides a measure of the child's cognitive skill. We assume that this measure includes not only the child's cognitive skill, but also the random teacher effect. This generates a correlation across these measures that is driven in part by the teacher.

Finally, we make one additional restriction to ease the computational burden. Rather than estimate separate loading factors and measurement variances for maternal and paternal measures of child non-cognitive skill, we constrain all of the paternal parameters to be scalars of the maternal parameters. For example, the father's loading on the child's non-cognitive skill in period t, $\alpha_{F,1jt}$, is equal to $\alpha_{F,1t} \times \alpha_{M,1jt}$ for all j. Similarly, the loading on the father's cognitive skill in the child's non-cognitive skill measures, $\alpha_{F,2jt}$, is given by $\alpha_{F,2t} \times \alpha_{M,2jt}$ for all j. Notice that we allow these scaling factors to vary by period.

We estimate the skill and measurement parameters outlined above using maximum likelihood. The likelihood contribution of household i is based strictly on the observed measures and skills associated with household i. Define M_i as the vector of all measurements and observed skills associated with household i across all time periods.[31] This vector can be written as

$$M_i = \mathrm{A} S_i^+ + \epsilon_i \tag{13}$$

where A is a matrix with 23 columns and as many rows as the total number of skill measurements for household i. The number of columns, 23, is the sum of the 18 skills and 5 random effects associated with measurement contamination. Therefore, with probability π the vector of measurements will be normally distributed with a mean of $\mu_1^M = \mathrm{A}\gamma_1$ and a covariance matrix $\Sigma_1^M = \mathrm{A}\Sigma_1\mathrm{A}' + \Sigma_\epsilon$. With probability $1-\pi$, the vector of measurements will be normally distributed with a mean $\mu_2^M = \mathrm{A}\gamma_2$ and a covariance matrix $\Sigma_2^M = \mathrm{A}\Sigma_2\mathrm{A}' + \Sigma_\epsilon$. Σ_ϵ is the (diagonal) covariance matrix of the measurement error. The log-likelihood contribution of household i is then given by

$$L_i(\gamma_1, \Sigma_1, \mu_2, \Sigma_2, \pi, \mathrm{A}, \Sigma_\epsilon) = \pi h(M_i; \mu_1^M, \Sigma_1^M) + (1-\pi) h(M_i; \mu_2^M, \Sigma_2^M) \tag{14}$$

where $h(\cdot)$ is the multivariate normal density function with the given mean and variance. We maximize the likelihood using a quasi-Newton optimization algorithm.[32]

[31] In practice there are missing measures across households and periods.

[32] We do not observe teacher identifiers for each period in our sample, and therefore assume each child has a different teacher. For the one period we can observe teacher identifiers ($t = 3$) we find that the average teacher is observed with fewer than two children in the sample. Thus, our assumption regarding different teachers is not particularly restrictive. At $t = 1$, interviewers assess child non-cognitive skill. On

The maximum likelihood procedure yields estimates of the joint skill distribution ($\hat{\gamma}_k$, $\hat{\Sigma}_k$, $\hat{\pi}$) and of the measurement related parameters (A and Σ_ϵ). The former are used primarily as an input into the second estimation step so we do not discuss these results directly. However, we can use the estimated measurement parameters to provide insight into the size of the parental and teacher distortions present in the measures of child non-cognitive skill.

In Table 7, 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 quite a bit noisier. Between $t = 2$ and $t = 4$ teachers are also asked to evaluate the cognitive skill of the child. This is a subjective measure which we allow to be affected by the teacher random effect. In the second panel of Table 7 we show that there is a non-trivial amount of noise in these measures that will be correlated with the teacher reported measures of child non-cognitive skill.

The third and fourth panels of the table show the degree of contamination in the parent and teacher reported measures of child non-cognitive 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 non-cognitive 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.

average, each interviewer is observed with more than thirty children. We have experimented with versions of the likelihood that allows for correlations in child skills within interviewer and found little change in the estimated parameters. For that reason we pursue the simpler model.

5.2 Skill Evolution

In a second step, we focus on estimating the skill transition functions, $f_t(S_{it})$. Although S_{it} is unobserved, we can simulate S_i for a large number of potential households using the previously obtained estimates of the mixture distribution ($\hat{\gamma}_k$, $\hat{\Sigma}_k$, and $\hat{\pi}$). Estimating a skill transition function using the simulated data simply requires a functional form for $f_t(S_{it})$.

For our preferred specification we assume that the skill transition function takes the following form,

$$Y_{it+1} = \beta^t_{Y,1} S_{it} + \beta^t_{Y,2}(Y_{it} \times S_{it}) + v^Y_{it+1} \tag{15}$$

for $Y \in \{C, N, N^M, N^F\}$.[33] S_{it} represents the full vector of child and parental skill at time t, while Y_{it+1} and Y_{it} are scalars that reflect the level of specific skill Y in periods $t+1$ and t respectively. The vector $\beta^t_{Y,1}$ captures the linear impact of past skills on next period skill, while the vector $\beta^t_{Y,2}$ captures how the impact of previous skills varies with the level of skill Y in the previous period. The latter term allows for non-linearity in skill evolution over time, capturing potential patterns of substitutability and complementarity across skills.[34] Notice that the parameters of the skill transition function are allowed to vary by period.

Estimates of the skill transition functions are presented in online Appendix Table 4.[35] Skills are not standardized (the observed measures are), so it is difficult to compare effect sizes either across columns or rows. In the next section 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. However, we present the coefficient estimates themselves to highlight some basic overarching patterns. First, self-

[33] Recall that parental cognitive skills are assumed to be fixed and thus we do not model their evolution.

[34] Note that the estimated parameters of the production function are sensitive to the normalizations imposed on the measurement system to identify the location and scale of the latent factors, and this would be true even if age-invariant measures were available. However, treatment effects based on the estimated linear technology are unaffected by location and scale normalizations even in the absence of age invariant measures. For a proof of this result and a more detailed discussion see Del Bono *et al.* (2020).

[35] Standard errors are calculated through a bootstrap procedure. We randomly sample with replacement from our original sample, and repeat the first estimation step. We then simulate data using the estimated skill distribution based on the bootstrap sample, and re-estimate the production function. We do this 30 times and calculate the standard deviation of the skill technology estimates across bootstrap samples.

productivity is generally high. This means, for example, that the strongest predictor for child cognitive skills is child cognitive skills in the previous period. There is also evidence of cross-productivity in skill formation. Especially in the early years, child non-cognitive skills appear to foster her cognitive skills, but this is not so much the case later on. Additionally, the non-cognitive skill of the child seems to have a statistically significant impact on the evolution of the non-cognitive skill of the parents, especially the mother.

Finally, there is evidence of non-linearity in the skill transition functions, though it appears rather small in magnitude. For example, at older ages the marginal impact of child cognitive skill at time t on child non-cognitive skill at time $t + 1$ is decreasing in the level of child non-cognitive skill at time t. There are also statistically significant interactions between parental non-cognitive skill and child non-cognitive skill in producing parental non-cognitive skill in the next period. Again, the non-linear effects are small.

6 Assessing the Impact of Distortions

While estimates of the skill transition and measurement equations are informative, it is difficult to fully understand how distortions in child non-cognitive skill measures impact our understanding of skill dynamics. In this section, we pursue two empirical exercises to shed further light on this matter.

6.1 Model Based Treatment Effects

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 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 non-cognitive skill, and simulate how this and all other skills evolve using the estimated production function.[36] We then compare the

[36]To simulate a skill boost in period $t > 1$, we use the initial skill draw and production function to generate skills through period $t - 1$ and then boost the relevant skill in period t. Skills in periods $t + 1$ are then affected through the production function.

treatment effects from our main specification with the treatment effects generated by two alternative models that neglect contamination issues.

The first alternative model is one where we ignore the availability of teacher-reported, child non-cognitive skill measures. Thus, we re-estimate the entire model relying only on parental measures of child non-cognitive skills. In this model we do not include any measurement distortions in child non-cognitive 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 non-cognitive 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 significantly different conclusions regarding the production of skill.

Tables 8 and 9 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, and parent-teacher models.[37] 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.[38] As an example, the first number in Table 8 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 8. 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

[37] We 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.

[38] As noted earlier, Del Bono *et al.* (2020) 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.

child non-cognitive 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 non-cognitive skill measures, we would expect the parameters that govern non-cognitive 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 non-cognitive skill is about 15% larger than in our baseline model. Also, the estimated effect of maternal non-cognitive skills in fostering child non-cognitive skills is 50% smaller than in the model that adjusts for distortions. 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 non-cognitive skills.

In the model that uses both teacher and parent measures of child non-cognitive skills (but does not model contamination) the estimated cross-productivities of child skills are 35-90% lower than in our main specification. This alternative model does not allow for a separate teacher effect in the teacher-reported measures of child non-cognitive skill. Therefore, the teacher effect – which is period-specific (or non persistent) - is absorbed into the child non-cognitive skill, making it less predictive of future cognitive skill and vice-versa. Previous literature on child skill dynamics has emphasized cross-skill complementarities and the influence of parental skill inputs. This exercise illustrates that there are potentially large biases in these parameters when contamination in child non-cognitive skill measures is ignored.

Table 9 focuses instead on the impact skill changes at $t = 1$ have on parental non-cognitive skills at $t = 4$. Boosting maternal or paternal non-cognitive skill when the child is 3 leads to approximately a third of a standard deviation increase in the corresponding skill when the child is age 11. Most interesting is the fact that the non-cognitive skill of the child can influence the non-cognitive skill of the mother. A one standard deviation increase in child non-cognitive skill at age 3 leads to a 0.123 standard deviation increase in maternal non-cognitive skill when the child is 11. For fathers, the effect is only 0.056. There are few differences between the main specification and the no contamination models since the parental measures themselves are not distorted and the self-productivity effects tend to dominate.

In addition to estimating the model under alternative assumptions about measurement availability and structure, we also investigate whether there is important heterogeneity in skill production and contamination by child or household characteristics. First, we re-estimate the model separately for girls and boys, but find almost no important differences in the estimated skill technology and measurement systems. Part of this likely reflects the fact that we adjust the measures for gender in a first step. Second, we split the sample evenly into low-SES and high-SES groups (defined according to household income measured at age 3) and re-estimate. Here we find mild heterogeneity in the production technology. The own and cross-productivity of child cognitive skills are larger for low-SES households, while the own and cross-productivity of child non-cognitive skills are larger for high-SES households. Also, parental cognitive skills have a larger impact on children in high-SES families. For the measurement system, we find that the contamination in parental measures of child non-cognitive skill is somewhat smaller for high-SES households, though this is not always the case. Because the heterogeneous patterns we discover are slight, we do not report these results.

6.2 Evaluating the Impact of Child Care Policies

The treatment effects studied in the previous section show how accounting for contamination in measures of child non-cognitive skills influences our predictions for child and parental skill evolution. Yet, the impact of reporting bias extends further. The measurement distortions evident in Table 7 can be relevant for reduced-form policy analysis regardless of the extent to which the treatment effects from a dynamic latent factor model are biased. Any child-centered policy that either directly or indirectly influences parents can generate spurious impacts on *measures* of child non-cognitive skill. For example, a number of papers estimate the impact universal child care programs have on child non-cognitive skills and/or maternal non-cognitive skills.[39] As our results illustrate, part of the

[39] See Yamaguchi *et al.* (2018), Haeck *et al.* (2019), Baker *et al.* (2008) and Datta Gupta & Simonsen (2010) among others.

estimated impact of universal child care on child non-cognitive skills could be due to the contamination driven by changes in maternal skill.

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 non-cognitive 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 non-cognitive skills.

Using our model we can quantify how contamination in parent-reported measures of child non-cognitive skills influences estimates of the short and long-run policy effects. We consider two scenarios. First, we consider a child care policy that shifts maternal non-cognitive 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 non-cognitive skills during the period of daycare through, for example, a labor supply response. However, the policy will have a real impact on child non-cognitive skills in all subsequent periods through the skills transition function (the indirect effect). Importantly, the change in maternal non-cognitive skills will also impact *measures* of child non-cognitive skills (the direct effect). Second, we consider a policy where both maternal and child non-cognitive skills increase in the first period by 20% of a standard deviation.[40] 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 non-cognitive skills across all ages is the result of contamination, i.e. the direct effect of a change in maternal non-cognitive skills on measures of child non-cognitive skills.

Figures 1 and 2 (corresponding to policy experiments 1 and 2) illustrate that contamination in child non-cognitive skills measures can pose a serious threat to policy evaluations.

[40]This is a large effect, but it is similar in magnitude to the estimates in Baker *et al.* (2008). Here the authors find that the introduction of universal child care subsidies in Quebec increased child care use and maternal labour supply, with negative impacts on maternal well-being and child emotional and social development.

Each panel in the figures represents a simulated SDQ measure of child non-cognitive 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.[41] Variation along the x-axis of each panel reflects the different waves when child non-cognitive skill is measured. The dark vertical bars represent the 'real' effect of the policy in measured child non-cognitive skill once contamination has been netted out. The light grey bars show the overall effect of the policy over time, including the 'real' effect and the effect resulting from contamination.

For the policy experiment where only maternal skill is directly affected, the contamination effects are large. Figure 1 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 non-cognitive 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 measure 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 non-cognitive skills can translate into a 0.15 of a standard deviation spurious increase in child non-cognitive skills.

Figure 2 illustrates the impact of a policy in which both maternal and child non-cognitive skills increase by 0.2 of a standard deviation in the first period. This policy is motivated by the findings in Baker *et al.* (2008), who found similar sized effects on maternal and child non-cognitive skills.[42] Looking again at SDQ emotional, we see that in the first

[41] In online Appendix Tables 5 and 6, we present tabular versions of the estimates for all child non-cognitive skill measures.

[42] Baker *et al.* (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.

period the measure increases by approximately 0.12 standard deviations.[43] However, the dark grey 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 non-cognitive skills of mothers, but positively affects those of the children, the distortion would lead to an underestimate of the policy on child non-cognitive skill.

These simulations suggest that relying on parental measures of child non-cognitive skill to evaluate programs or policies that also influence parental non-cognitive skills is highly problematic. We are not the first to point this out, as Baker *et al.* (2008) explicitly recognize this as a potential threat to their findings. Are there clear solutions to this issue? First, it appears that having multiple evaluators of the child is useful. This is because while each evaluator is potentially biased, when the distortions are unrelated across evaluators we can obtain a clearer signal of child non-cognitive skills. Second, objective measures of child behavior, such as the number of emotional outbursts in the past week, might be less sensitive to parental distortion. Additional research is needed to determine whether this is a reasonable alternative.

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 own skills and traits to these evaluations, potentially contaminating measures of child non-cognitive skills. In this paper we show that contamination in measures of child non-cognitive skills can significantly affect our basic understanding of child skill dynamics. Additionally, when

[43] The 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 non-cognitive skill, will suffer from this measurement problem.

parental skills contaminate measures of child non-cognitive 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. Going forward, data collection efforts that seek to study child development should attempt to include evaluations of the child beyond just parents. Alternatively, surveys could collect more objective measures of child non-cognitive skills related directly to observed behaviors.

The measurement issues we highlight are not necessarily specific to child non-cognitive skills. We assume that adult reports of their own non-cognitive skill are uncontaminated, but it would be interesting to obtain external evaluations or direct behaviors that could speak to this. Other literatures have faced similar issues. As an example, self-reported health status is likely impacted by individual interpretations of what constitutes excellent health or by a reference-level of health.[44] Researchers attempt to deal with these measurement concerns by using fixed-effect type models or incorporating objective health measures, such as blood pressure or cholesterol. 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.

[44] See for example Baker *et al.* (2004), Bound (1991).

References

Agostinelli, Francesco, & Wiswall, Matthew. 2020. *Estimating the Technology of Children's Skill Formation*. NBER Working Paper 22442.

Attanasio, Orazio, Meghir, Costas, & Nix, Emily. 2019. Human capital development and parental investment in India. *Review of Economic Studies*. forthcoming.

Attanasio, Orazio, Cattan, Sara, Fitzsimons, Emla, Meghir, Costas, & Rubio-Codina, Marta. 2020. Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia. *American Economic Review*, **110**(1), 48–85.

Baker, Michael, & Milligan, Kevin. 2015. Maternity leave and children's cognitive and behavioral development. *Journal of Population Economics*, **28**(2), 373–391.

Baker, Michael, Stabile, Mark, & Deri, Catherine. 2004. What Do Self-Reported, Objective, Measures of Health Measure? *Journal of Human Resources*, **39**(4), 1067–1093.

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

Baranov, V., Bhalotra, S., Biroli, P., & Maselko, J. 2017. *Maternal Depression,Women's Empowerment, and Parental Investment: Evidence from a Large Randomized Control Trial*. IZA Discussion Paper 11187.

Black, Sandra E., Devereux, Paul J., & Salvanes, Kjell G. 2016. Does Grief Transfer Across Gen-erations? Bereavements During Pregnancy and Child Outcomes. *American Economic Journal: Applied Economics*, **8**(1), 193223.

Bound, John. 1991. Self-Reported Versus Objective Measures of Health in Retirement Models. *Journal of Human Resources*, **26**(1), 106–138.

Bracken, B. 2002. *Bracken School Readiness Assessment*. San Antonio, TX: The Psychological Corporation.

Campbell, Donald T., & Fiske, Donald W. 1959. Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, **56**(2), 81–105.

Carneiro, Pedro, Hansen, Karsten T., & Heckman, James J. 2003. Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice. *International Economic Review*, **44**(2), 361–422.

Choe, D., Olson, S., & Sameroff, A. 2014. Effortful Control Moderates Bidirectional Effects Between Children's Externalizing Behavior and Their Mothers' Depressive Symptoms. *Child Development*, **85**, 643–658.

Conti, Gabriella, Heckman, James, & Pinto, Rodrigo. 2016. The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour. *Economic Journal*, **126**, F28–F65.

Cunha, Flavio, Heckman, James J., Lochner, Lance, & Masterov, D. 2006. Interpreting the evidence on life cycle skill formation. *Chap. 12, pages 697–812 of:* Hanushek, E.A., & Welch, F. (eds), *Handbook of the Economics of Education*, vol. 1. Amsterdam: Elsevier.

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

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

Davis, Naomi Ornstein, & Carter, Alice S. 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.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. 2015. The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, **141**(4), 858–900.

Del Bono, Emilia, Francesconi, Marco, Kelly, Yvonne, & Sacker, Amanda. 2016. Early Maternal Time Investment and Early Child Outcomes. *The Economic Journal*, **126**(596), F96–F135.

Del Bono, Emilia, Kinsler, Josh, & Pavan, Ronni. 2020. *A Note on the Importance of Normalizations in Dynamic Latent Factor Models of Skill Formation*. Working Paper.

Deming, David. 2009. Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics*, **3**(1), 111–134.

Elliott, C., Smith, P., & McCulloch, K. 1996. *British Ability Scales Second Edition (BAS II). Administration and Scoring Manual*. London: Nelson.

Goodman, A., Lamping, D., & Ploubidis, G. 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**, 1179–91.

Goodman, R. 1997. The Strengths and Difficulties Questionnaire, A Research Note. *Journal of Child Psychology and Psychiatry*, **38**, 581–586.

Haeck, Catherine, Lebihan, Laetitia, Lefebvre, Pierre, & Merrigan, Philip. 2019. *Childcare, parental behavior and well-being during childhood*. unpublished manuscript.

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

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

Hernández-Alava, Mónica, & Popli, Gurleen. 2017. Children's Development and Parental Input: Evidence From the UK Millennium Cohort Study. *Demography*, **54**(2), 485–511.

Johnston, David, Propper, Carol, Pudney, Stephen, & Shields, Michael. 2014. Child mental health and educational attainment: multiple observers and the measurement error problem. *Journal of Applied Econometrics*, **29**, 880–900.

Joreskog, K.G. 1971. Statistical analysis of sets of congeneric tests. *Psychometrika*, **36**, 109–133.

Kessler, R., Barker, P., Colpe, L., Epstein, J., Gfroerer, J., Hiripi, E., Howes, M., Normand, S., Manderscheid, R., Walters, E., & Zaslavsky, A. 2003. Screening for serious mental illness in the general population. *Archives of General Psychiatry*, **60**, 184–189.

Kraemer, H., Measelle, J., Ablow, J., Essex, M., Boyce, W., & Kupfer, D. 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), 1566–77.

Kuhn, Jennifer C., & Carter, Alice S. 2006. Maternal self-efficacy and associated parenting cognitions among mothers of children with autism. *The American journal of Orthopsychiatry*, **76**, 564–75.

Martel, M., Markon, K., & Smith, G. 2017. Research Review: Multi-informant integration in child and adolescent psychopathology diagnosis. *Journal of Child Psychology and Psychiatry*, **58**, 116128.

Melhuish, E., Hanna, K., Quinn, L., Sylva, K., Siraj-Blatchford, I., Sammons, P., & Taggart, B. 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*. Belfast: Stranmillis University Press.

Moffitt, T., Arseneault, Louise, Belsky, Daniel, & Caspi, Avshalom. 2011. A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, **108**, 2693–2698.

Moroni, Gloria, Nicoletti, Cheti, & Tominey, Emma. 2019. *Child Socio-Emotional Skills: The Role of Parental Inputs*. IZA Discussion Paper 12432.

Persson, Petra, & Rossin-Slater, M. 2018. Family Ruptures, Stress, and the Mental Health of the Next Generation. *American Economic Review*, **108**, 1214–52.

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

Pohl, Steffi, & Steyer, Rolf. 2010. Modeling Common Traits and Method Effects in Multitrait-Multimethod Analysis. *Multivariate Behavioral Research*, **45**, 45–72.

Pohl, Steffi, Steyer, Rolf, & Kraus, Katrin. 2008. Modeling method effects as individual causal effects. *Journal of the Royal Statistical Society Series A*, **171**(1), 41–63.

Prochaska, Judith J., Y., Sung H., W., Max, Y., Shi, & M., Ong. 2012. 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.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., & Rabbitt, P. 1998. Screening for serious mental illness in the general population. *Journal of the International Neuropsychological Society*, **4**, 474–490.

Todd, Petra E., & Wolpin, Kenneth I. 2007. The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, **1**(1), 91–136.

Yamaguchi, Shintaro, Asai, Yukiko, & Kambayashi, Ryo. 2018. How does early childcare enrollment affect children, parents, and their interactions? *Labour Economics*, **55**, 56–71.

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	
N.	19,048		14,648		14,598		12,530	

Notes: 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 sub-sample 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.

Table 2: Child Cognitive Measures

	$t = 1$ (age 3)	$t = 2$ (age 5)	$t = 3$ (age 7)	$t = 4$ (age 11)
Administered by interviewer				
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				✓
Assessed by teacher				
FSP (Reading, writing, calculating etc.)		✓		
Subject assessment (English, Maths and Science)			✓	✓

Notes: UK Millennium Cohort Study.

Table 3: Child Non-Cognitive Measures

	$t = 1$ (age 3)	$t = 2$ (age 5)	$t = 3$ (age 7)	$t = 4$ (age 11)
Reported by the mother				
SDQ (hyperactivity, conduct, emotional, peer)	✓	✓	✓	✓
CSB (independence, emotional, cooperation)	✓	✓	✓	
Reported by interviewer				
Behavior score (extreme behaviour, attention, cooperation)	✓			
Reported by teacher				
FSP (dispositions, social, emotional)		✓		
SDQ (hyperactivity, conduct, emotional, peer)			✓	✓

Notes: UK Millennium Cohort Study.

Table 4: Evidence of Distortions in Child Non-Cognitive Skill Measures

	Child Non-Cognitive Skill$_t$			
	Mother Reported		Teacher/Interv. Reported	
	(1)	(2)	(3)	(4)
Child Cognitive Skill$_t$	0.206**	0.191**	0.291**	0.270**
	(0.007)	(0.007)	(0.007)	(0.007)
Mother Cognitive Skill$_t$	0.117**	0.094**	0.021**	0.019*
	(0.007)	(0.008)	(0.006)	(0.006)
Mother Non-Cognitive Skill$_t$	0.299**	0.291**	0.071**	0.068**
	(0.007)	(0.007)	(0.007)	(0.007)
Missing Indicators	Y	Y	Y	Y
Demographics	N	Y	N	Y
N	33,905	33,905	26,818	26,818
R^2	0.182	0.208	0.1106	0.138

Notes: UK Millennium Cohort Study, data from $t = 1$ to $t = 4$ (child age 3, 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 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Table 5: Additional Evidence of Distortions in Child Non-Cognitive Skill Measures

	Child Non-Cognitive Skill$_t$		
	Mother Reported	Teacher Reported	Father Reported
	(1)	(2)	(3)
Child Cognitive Skill	0.186**	0.267**	0.152**
	(0.007)	(0.007)	(0.046)
Mother Cognitive Skill	0.080**	0.013^{+}	0.033
	(0.009)	(0.007)	(0.057)
Mother Non-Cognitive Skill	0.281**	0.058**	0.155*
	(0.007)	(0.007)	(0.053)
Father Cognitive Skill	0.034**	0.011	0.068
	(0.009)	(0.007)	(0.051)
Father Non-Cognitive Skill	0.044**	0.040**	0.225**
	(0.007)	(0.007)	(0.046)
Missing Indicators	Y	Y	Y
Demographics	Y	Y	Y
N	33,905	26,818	443
R^2	0.207	0.138	0.304

Notes: UK Millennium Cohort Study, data from $t = 1$ to $t = 4$ (child age 3, 5, 7 and 11). Estimates are obtained using linear regressions. See footnote to Table 4 for the definition of variables. Standard errors are clustered at the individual level.$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Table 6: Skill Technology Distortions

	Child Non-Cognitive Skill$_{t+1}$			
	Mother Reported		Teacher Reported	
	(1)	(2)	(3)	(4)
Child Cognitive Skill$_t$	0.057**	0.039**	0.156**	0.134**
	(0.006)	(0.006)	(0.010)	(0.010)
Mother Reported Child Non-Cognitive Skill$_t$	0.617**	0.604**		0.134**
	(0.007)	(0.007)		(0.010)
Teacher Reported Child Non-Cognitive Skill$_t$		0.064**	0.221**	0.200**
		(0.007)	(0.011)	(0.012)
Mother Cognitive Skill	0.016**	0.018 **	0.035**	0.023*
	(0.006)	(0.006)	(0.009)	(0.009)
Mother Non-Cognitive Skill$_t$	0.070**	0.072**	0.059**	0.027**
	(0.006)	(0.007)	(0.009)	(0.010)
Father Cognitive Skill	0.023**	0.025*	0.033**	0.032**
	(0.006)	(0.006)	(0.009)	(0.010)
Father Non-Cognitive Skill$_t$	0.024**	0.024**	0.031**	0.025*
	(0.006)	(0.006)	(0.009)	(0.009)
N	22,371	18,271	13,648	13,197
R^2	0.466	0.465	0.166	0.180

Notes: UK Millennium Cohort Study, data from $t = 1$ to $t = 4$ (child age 3, 5, 7 and 11). Estimates are obtained using linear regressions. Definition of variables and demographic controls as in footnote to Table 4. Standard errors are clustered at the individual level.$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Table 7: Contamination in measurements

	$t=1$		$t=2$		$t=3$		$t=4$	
	signal	cont.	signal	cont.	signal	cont.	signal	cont.
Child Cognitive - Test Based								
Braken	58.1%	-						
BAS Naming Vocabulary	41.9%	-	36.7%	-				
BAS picture similarity			21.5%	-				
BAS pattern compr.			32.5%	-	33.3%	-		
BAS Word Recognition					52.1%	-		
NFER in Math					53.0%	-		
BAS verbal							25.7%	-
GTC quality							6.9%	-
CANTAB swm strat							12.4%	-
CANTAB swm err							24.1%	-
Child Cognitive - Teacher Reported								
FSP			31.8%	24.9%				
Teacher assessment					61.1%	5.7%	70.7%	3.8%
Child Non-Cognitive - Parent Reported								
SDQ emotional	20.9%	5.9%	28.6%	8.1%	30.3%	8.4%	34.0%	6.3%
SDQ conduct	35.3%	9.9%	34.6%	12.8%	35.5%	13.7%	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%	33.7%	8.3%	35.5%	7.8%	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%	40.6%	15.2%		
Q. Cooperation					32.0%	6.8%		
Child Non-Cognitive - Intverviewer/Teacher Reported								
Focus	5.7%	94.1%						
Cooperation	2.7%	37.9%						
Extreme behavior	4.4%	11.4%						
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%	37.1%	12.7%	32.6%
SDQ hyperactivity					14.9%	53.6%	18.2%	42.1%
SDQ peer					11.8%	13.6%	17.6%	7.6%

Notes: 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 non-cognitive skills. This measure can be written as

$$M^N_{Pijt} = \alpha^N_{P,1jt}N_{it} + \alpha^N_{P,2jt}C^P_i + \alpha^N_{P,3jt}N^P_{it} + \alpha^N_{P,4jt}\theta_i + \epsilon^N_{P,ijt}.$$

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

Table 8: Impulse Response: Child Skills

	Main Specification		Only Parental measures		All measures (No distortions)	
	C_4	N_4	C_4	N_4	C_4	N_4
$+1sd$ in C_1	0.578*	0.110*	0.593	0.115	0.633	0.072
	(0.018)	(0.016)				
$+1sd$ in N_1	0.088*	0.465*	0.074	0.533	0.009	0.488
	(0.019)	(0.021)				
$+1sd$ in C_1^M	0.074*	0.008	0.077	0.001	0.069	0.010
	(0.016)	(0.028)				
$+1sd$ in N_1^M	0.005	0.102*	-0.006	0.051	0.003	0.049
	(0.018)	(0.031)				
$+1sd$ in C_1^F	0.116*	0.044*	0.110	0.050	0.099	0.059
	(0.013)	(0.012)				
$+1sd$ in N_1^F	0.023	0.022	0.026	0.025	0.015	0.034
	(0.016)	(0.013)				

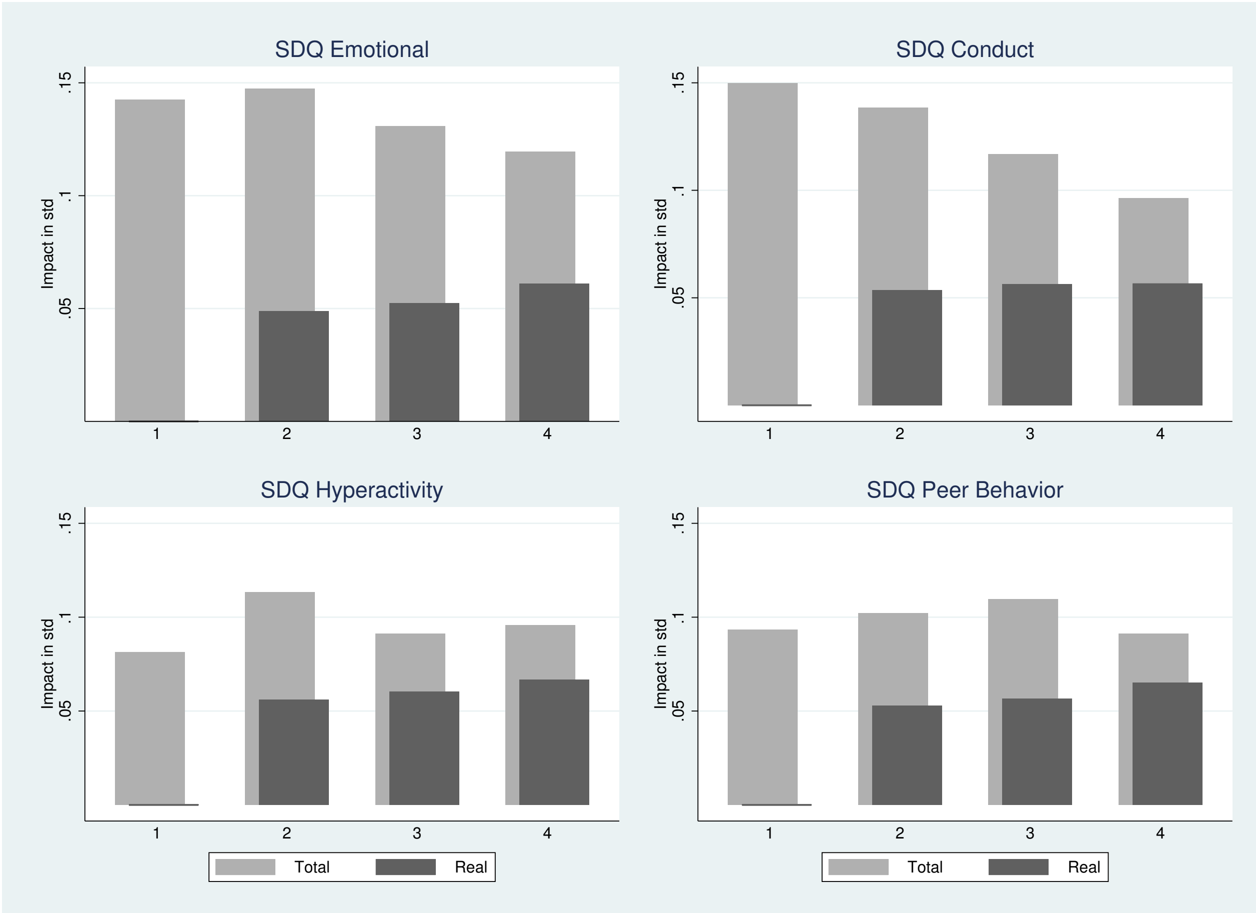
Notes: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in sections 4 and 5. In each row we report the impact on children skills at age 11 when different skills are shocked by 1sd when the children are 3 years old. The first two columns refer to our main specification. The second two columns present results for a model that does not correct for the contamination and utilizes only parental reported children non-cognitive measures. The last two columns refer to a model that does not correct for contamination, but utilizes all available measures. Standard errors are obtained bootstrapping 30 times the original data set at the individual level and are available only for the main specification.* $p < 0.05$

Table 9: Impulse Response: Parental Skills

	Main Specification		Only Parental measures		All measures (No distortions)	
	N_4^M	N_4^F	N_4^M	N_4^F	N_4^M	N_4^F
$+1sd$ in C_1	0.002	0.022	0.013	0.023	-0.005	0.014
	(0.011)	(0.012)				
$+1sd$ in N_1	0.123*	0.056*	0.116	0.061	0.115	0.065
	(0.012)	(0.010)				
$+1sd$ in N_1^M	0.309*	0.077*	0.286	0.072	0.291	0.074
	(0.014)	(0.012)				
$+1sd$ in N_1^F	0.066*	0.359*	0.072	0.354	0.080	0.363
	(0.014)	(0.019)				

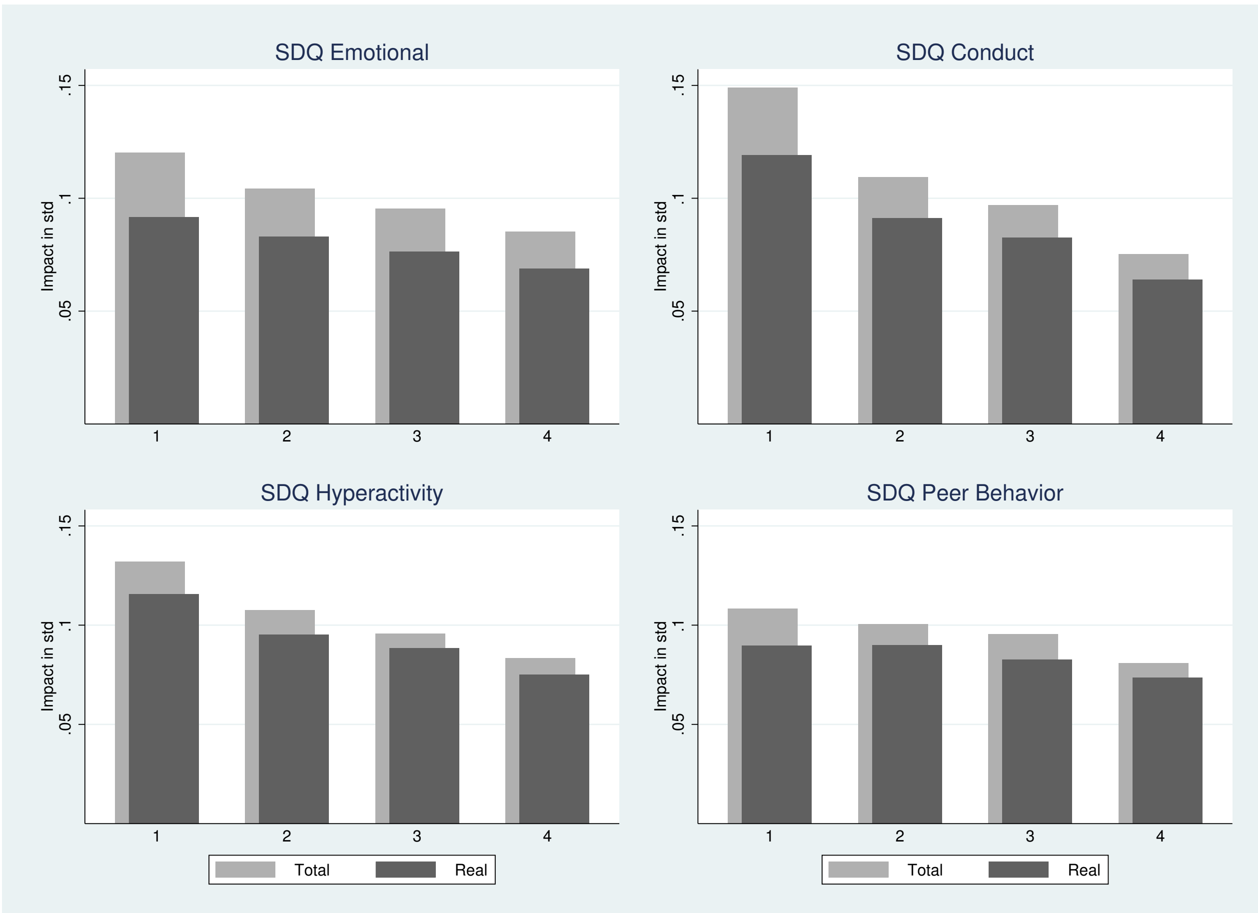
Notes: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in sections 4 and 5. In each row we shock a different skill by 1sd when the child is 3 years old, and report the impact on parental skills when the child is 11. The first two columns refer to our main specification. The second two columns present results for a model that does not correct for the contamination and utilizes only parental reported children non-cognitive measures. The last two columns refer to a model that does not correct for contamination, but utilizes all available measures. Standard errors are obtained bootstrapping 30 times the original data set at the individual level and are available only for the main specification.* $p < 0.05$

Figure 1: 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 light grey bars indicate the increase in each SDQ measure in different periods after we increase maternal non-cognitive skills by 1sd at $t = 1$. The dark bars indicate the same impact after we have removed the contamination from those measures.

Figure 2: 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 light grey bars indicate the increase in each SDQ measure in different periods after we increase maternal non-cognitive skills and children non cognitive skills by 0.2sd at $t = 1$. The dark bars indicate the same impact after we have removed the contamination from those measures.

Appendices

A Identification of the Model

In this section, we consider the identification of f_t, F_t, and all the parameters of the measurement equations. We proceed in two steps. First, we show how to identify the first and second moments of the joint density of S_i (along with the measurement equations). Next, we show that the density of S_i is non-parametrically identified. 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. For skill Y_{it+1} for $Y \in \{C, N, N^M, N^F\}$ define:

$$f^Y_{t+1}(Y_{it}) \equiv E(Y_{it+1}|S_{it})$$

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

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 turn to the measurement model (see equations 3 to 6 for the definition of the loading factors and their subscripts). Consider first the identification of the second moments of the joint density of S_i. Following Cunha et al. (2010), we normalize $\alpha^Y_{1t} = 1$ for all skills (normalize $\alpha^N_{T,11t} = 1$ for child non-cognitive skill). Also, normalize $\alpha^N_{P,411} = \alpha^N_{T,21t} = 1$ to set the scale of the random variables θ_i and T_i respectively. Finally, we normalize the means of S and ϵ to be equal to zero. As mentioned in the paper, all measures can be purged of the effect of observables in a first stage, i.e. we can identify the impact of observables looking at the conditional means of the measures. This is consistent with a model in which observables may affect directly the measure but also affect the production function of skills, although in a linearly separable fashion.

Using the measurement equations of Equation (5), consider the following covariances:

$$
\begin{aligned}
Cov_1 = Cov(M_{i1t}^{N^P}, M_{i1\tau}^{N^P}) &= \sigma_{t,\tau}^{N^P} \qquad \text{for } t \neq \tau \\
Cov_2 = Cov(M_{ijt}^{N^P}, M_{i1\tau}^{N^P}) &= \alpha_{jt}^{N^P} \sigma_{t,\tau}^{N^P} \qquad \text{for } t \neq \tau \\
Cov_3 = Cov(M_{ijt}^{N^P}, M_{i1t}^{N^P}) &= \alpha_{jt}^{N^P} \sigma_{t}^{N^P} \\
Cov_4 = Cov(M_{i1t}^{N^P}, C_i^P) &= \sigma_t^{N^P C^P} \\
Cov_5 = Cov(C_i^P, C_i^P) &= \sigma^{C^P} \\
Cov_6 = Cov(M_{i1t}^{C}, M_{i1\tau}^{C}) &= \sigma_{t,\tau}^{C} \qquad \text{for } t \neq \tau \\
Cov_7 = Cov(M_{ijt}^{C}, M_{i1\tau}^{C}) &= \alpha_{jt}^{C} \sigma_{t,\tau}^{C} \qquad \text{for } t \neq \tau \\
Cov_8 = Cov(M_{ijt}^{C}, M_{i1t}^{C}) &= \alpha_{jt}^{C} \sigma_{t}^{C} \\
Cov_9 = Cov(M_{i1t}^{C}, M_{i1\tau}^{N^P}) &= \sigma_{t,\tau}^{C N^P} \\
Cov_{10} = Cov(M_{i1t}^{C}, C_i^P) &= \sigma_{t}^{C C^P} \\
Cov_{11} = Cov(M_{i1t}^{N^{P_1}}, M_{i1\tau}^{N^{P_2}}) &= \sigma_{t\tau}^{N^{P_1} N^{P_2}} \\
Cov_{12} = Cov(C_i^{P_1}, C_i^{P_2}) &= \sigma^{C^{P_1} C^{P_2}} \\
Cov_{13} = Cov(M_{i1t}^{N^{P_1}}, C^{P_2}) &= \sigma_t^{N^{P_1} C^{P_2}}
\end{aligned}
$$

where $j \neq 1$, and $P_2 = M$ if $P_1 = F$ or $P_2 = F$ if $P_1 = M$. Cov_1 directly identifies $\sigma_{t,\tau}^{N^P}$. From the ratio of Cov_2 and Cov_1 we can identify $\alpha_{jt}^{N^P}$. With this knowledge, we can use Cov_3 to identify $\sigma_t^{N^P}$. Cov_4 directly identifies $\sigma_t^{N^P C^P}$. Cov_5, which truly is a variance, identifies σ^{C_P} (remember our assumption that parental cognitive skills are observed). Cov_6 directly identifies $\sigma_{t,\tau}^{C}$. From the ratio of Cov_7 and Cov_6 we can identify α_{jt}^{C} which can then be used in Cov_8 to identify σ_t^C. Finally using Cov_9 to Cov_{13}, all second moments relative to C_t, C^P and N_t^P can be identified. Using the variance of a measurement M_{ijt}^Y for $Y \in \{C, N^M, N^F\}$ will identify the variance of the measurement error ϵ_{ijt}^Y.

To identify the second moments related to child non-cognitive skill, consider first the teacher reported measures and the following covariances:

$$\begin{aligned}
Cov(M^N_{T,i1t}, M^N_{T,i1\tau}) &= \sigma^N_{t,\tau} \qquad \text{for } t \neq \tau \\
Cov(M^N_{T,ijt}, M^N_{T,i1\tau}) &= \alpha^N_{T,1jt}\sigma^N_{t,\tau} \qquad \text{for } t \neq \tau.
\end{aligned}$$

These two observable covariances identify the covariance of non-cognitive skills across different time periods and the loading factor relative to the teacher measure. Using the covariance of M^N_{Ti1t} with the child cognitive and parental cognitive and non-cognitive measures utilized previously, we can also identify the covariances between child non-cognitive skills and these other skills. Unfortunately, using just teacher reported measures of child non-cognitive skill we cannot identify $\sigma^N_{t,t}$, $\alpha^N_{T,2jt}$, and the variance of T_{it}. Consider now the parent reported measures of child non-cognitive skill and the following covariances

$$\begin{aligned}
Cov(M^N_{P,ijt}, M^C_{i1t}) &= \alpha^N_{P,1jt}\sigma^{NC}_{t,t} + \alpha^N_{P,2jt}\sigma^{C^PC}_t + \alpha^N_{P,3jt}\sigma^{N^PC}_{t,t} \\
Cov(M^N_{P,ijt}, C^P_i) &= \alpha^N_{P,1jt}\sigma^{NC^P}_t + \alpha^N_{P,2jt}\sigma^{C^P} + \alpha^N_{P,3jt}\sigma^{N^PC^P}_t \\
Cov(M^N_{P,ijt}, M^{N^P}_{i1t}) &= \alpha^N_{P,1jt}\sigma^{NN^P}_{t,t} + \alpha^N_{P,2jt}\sigma^{C^PN^P}_t + \alpha^N_{P,3jt}\sigma^{N^P}_{tt}.
\end{aligned}$$

All of the covariances in the above system have been previously identified. Thus, as long as the matrix

$$\begin{bmatrix}
\sigma^{NC}_{t,t} & \sigma^{C^PC}_t & \sigma^{N^PC}_{t,t} \\
\sigma^{NC^P}_t & \sigma^{C^P} & \sigma^{N^PC^P}_t \\
\sigma^{NN^P}_{t,t} & \sigma^{C^PN^P}_t & \sigma^{N^P}_{tt}
\end{bmatrix}$$

has full rank, we can also identify the loading factors in the parental measures of child non-cognitive skill related to the child and parental skills. The covariance

$$Cov(M^N_{T,i1t}, M^N_{P,ijt}) = \alpha^N_{P,1jt}\sigma^N_{t,t} + \alpha^N_{P,2jt}\sigma^{C^PN}_t + \alpha^N_{P,3jt}\sigma^{N^PN}_{t,t}$$

then identifies $\sigma^N_{t,t}$. Assuming we have at least 3 teacher reported measures, we can identify the parameters relative to the teacher random effects and the measurement errors for these measures using

$$
\begin{aligned}
Cov(M^N_{T,ijt}, M^N_{T,i1t}) - \alpha^N_{T,1jt}\sigma^N_{t,t} &= \alpha^N_{T,2jt}\sigma^T_{t,t} \\
Cov(M^N_{T,ijt}, M^N_{T,ikt}) - \alpha^N_{T,1jt}\alpha^N_{T,1kt}\sigma^N_{t,t} &= \alpha^N_{T,2kt}\alpha^N_{T,2jt}\sigma^T_{t,t}.
\end{aligned}
$$

The ratio of the first and second equations identifies $\alpha^N_{T,2kt}$, which in turns leads to the identification of $\sigma^T_{t,t}$. The variance of the measures will also identify the variance of $\epsilon^N_{T,ijt}$. An analogous procedure will also identify $\alpha^N_{P,4jt}$, $\sigma^\theta_{t,t}$, and the variance of $\epsilon^N_{P,ijt}$, again assuming at least three measures are available.

The first two moments of the joint density of S_i and all parameters related to the measurement equations are now identified. However, unless we want to assume S_i is jointly normal, we have not identified the full distribution of S_i. It is important to go beyond normality if we want to allow for non-linearity in the production function f_t. For this part of the identification, we rely on Theorem 1 of Cunha et al. (2010). Consider first W_j and ω_j for $j = \{1, 2\}$ where:

$$
\begin{aligned}
\tilde{W}_j &= \left(\left\{ \frac{M^C_{jt}}{\alpha^C_{jt}} \right\}^T_{t=1}, C^M_i, \left\{ \frac{M^{N^M}_{jt}}{\alpha^{N^M}_{jt}} \right\}^T_{t=1}, C^F_i, \left\{ \frac{M^{N^F}_{jt}}{\alpha^{N^F}_{jt}} \right\}^T_{t=1}, \left\{ M^N_{P,1t} \right\}^T_{t=1} \right) \\
\tilde{\omega}_j &= \left(\left\{ \frac{\epsilon^C_{jt}}{\alpha^C_{jt}} \right\}^T_{t=1}, 0, \left\{ \frac{\epsilon^{N^M}_{jt}}{\alpha^{N^M}_{jt}} \right\}^T_{t=1}, 0, \left\{ \frac{\epsilon^{N^F}_{jt}}{\alpha^{N^F}_{jt}} \right\}^T_{t=1}, \left\{ 0 \right\}^T_{t=1} \right).
\end{aligned}
$$

Notice that the second, foruth and last elements of $\tilde{W}_j$ do not have a j subscript (the last element is the first measurement of the parent reported children non-cognitive skills), which means that we will consider the same vector of measurements for $j = 1$ and $j = 2$. The reason for including $M^N_{P,1t}$ in $\tilde{W}_j$ will be clear in a moment. C^P_i are observed and therefore we can think of $C^P_i + 0$ as the measurement. Trivially the error term for these two measurements will be independent between $j = 1$ and $j = 2$. Using these definitions we can write the system of measures as

$$
\tilde{W}_j = \tilde{S} + \tilde{\omega}_j
$$

where $\tilde{S}$ is a vector of unobserved child cognitive skills, parental skills, and the components of the kth parental measure of child non-cognitive skill. Under the assumption that $E[\tilde{\omega}_1|\tilde{u},\tilde{\omega}_2]=0$, Theorem 1 of Cunha et al. (2010) shows that with the knowledge of the distribution of $\tilde{W}_j$ for $j=1$ and $j=2$, we can identify the joint density of $\tilde{S}$. Of course the joint density of $\tilde{S}$ is not equivalent to the joint density of S since the former does not contain the child's non-cognitive skill. More specifically we have just shown identification of the joint density of children cognitive skills C, maternal skills C^M and N^M, paternal skills C^F and N^F, and the first measurement of the parent reported children non-cognitive skills $M^N_{P,1}$.

Using the fact that we have different people reporting child non-cognitive skill we can adapt Theorem 1 to show identification of the joint density including child non-cognitive skill. For illustrative purposes, assume for a moment that $\alpha^N_{P,21t}=\alpha^N_{P,31t}=0$ so that

$$\begin{aligned} M^N_{P,i1t} &= \alpha^N_{P,11t}N_{it}+\alpha^N_{P,41t}\theta_i+\epsilon^N_{P,i1t} \\ M^N_{T,i1t} &= N_{it}+\alpha^N_{T,21t}T_{it}+\epsilon^N_{T,i1t}. \end{aligned}$$

Notice how the measurement errors, $\alpha^N_{P,41t}\theta_i+\epsilon^N_{P,i1t}$ and $\alpha^N_{T,21t}T_{it}+\epsilon^N_{T,i1t}$, are independent of each other. If we define

$$\begin{aligned} W_1 &= \left(\left\{\frac{M^C_{1t}}{\alpha^C_{1t}}\right\}^T_{t=1}, C^M_i, \left\{\frac{M^{N^M}_{1t}}{\alpha^{N^M}_{1t}}\right\}^T_{t=1}, C^F_i, \left\{\frac{M^{N^F}_{1t}}{\alpha^{N^F}_{1t}}\right\}^T_{t=1}, \left\{M^N_{T,1t}\right\}^T_{t=1}\right) \\ W_2 &= \left(\left\{\frac{M^C_{2t}}{\alpha^C_{2t}}\right\}^T_{t=1}, 0, \left\{\frac{M^{N^M}_{2t}}{\alpha^{N^M}_{2t}}\right\}^T_{t=1}, 0, \left\{\frac{M^{N^F}_{2t}}{\alpha^{N^F}_{2t}}\right\}^T_{t=1}, \left\{\frac{M^N_{P,1t}}{\alpha^N_{P,11t}}\right\}^T_{t=1}\right). \end{aligned}$$

and

$$\begin{aligned} \omega_1 &= \left(\left\{\frac{\epsilon^C_{1t}}{\alpha^C_{1t}}\right\}^T_{t=1}, \frac{\epsilon^{C^M}_{1t}}{\alpha^{C^M}_{1t}}, \left\{\frac{\epsilon^{N^M}_{1t}}{\alpha^{N^M}_{1t}}\right\}^T_{t=1}, \frac{\epsilon^{C^F}_{1t}}{\alpha^{C^F}_{1t}}, \left\{\frac{\epsilon^{N^F}_{1t}}{\alpha^{N^F}_{1t}}\right\}^T_{t=1}, \left\{\alpha^N_{T,21t}\theta+\epsilon^N_{T,1t}\right\}^T_{t=1}\right) \\ \omega_2 &= \left(\left\{\frac{\epsilon^C_{2t}}{\alpha^C_{2t}}\right\}^T_{t=1}, \frac{\epsilon^{C^M}_{2t}}{\alpha^{C^M}_{2t}}, \left\{\frac{\epsilon^{N^M}_{2t}}{\alpha^{N^M}_{2t}}\right\}^T_{t=1}, \frac{\epsilon^{C^F}_{2t}}{\alpha^{C^F}_{2t}}, \left\{\frac{\epsilon^{N^F}_{2t}}{\alpha^{N^F}_{2t}}\right\}^T_{t=1}, \left\{\frac{\alpha^N_{P,41t}}{\alpha^N_{P,11t}}T_t+\frac{\epsilon^N_{P,1t}}{\alpha^N_{P,11t}}\right\}^T_{t=1}\right). \end{aligned}$$

we can again apply Theorem 1 of Cunha et al. (2010), where $W_j=u+\omega_j$. Under the assumption that $E[\omega_1|u, \omega_2] = 0$, Theorem 1 shows that we can identify the joint density of S.

The issue with the above derivation is that α^N_{P21t} and α^N_{P31t} are not equal to zero. In order to recover the distribution of S, the final term of W_2 should be

$$\frac{M^N_{P,1t}}{\alpha^N_{P11t}} - \frac{\alpha^N_{P,2jt}}{\alpha^N_{P,11t}} C^P - \frac{\alpha^N_{P,3jt}}{\alpha^N_{P,11t}} N^P_t.$$

While we do not observe C^P and N^P, we have previously identified the joint density of $M^N_{P,i1t}$ with these variables and all other unobserved skills (other than the child non-cognitive). In order to apply the theorem, we do not actually know need to observe each element of W_j. we need instead to be able recover the distribution of W_1 and W_2, which in general is directly derived from the data. While for W_1 this is true, we can derive the corrected W_2 from the incorrect W_2, after we modify it using our knowledge of the joint density of its last element with all other skills, i.e. the distribution of W_2. After this correction, we can then apply Theorem 1.

B Additional Tables

Appendix Table 1: Definition of Non-Cognitive skills

	$t > 2$ SDQ total		$t > 2$ SDQ Intern.		$t > 2$ SDQ Extern.	
N_t reported by:	Mother	Teacher	Mother	Teacher	Mother	Teacher
	(1)	(2)	(3)	(4)	(5)	(6)
Child Cognitive Skill$_t$	0.191**	0.228**	0.124**	0.148**	0.196**	0.224**
	(0.009)	(0.012)	(0.009)	(0.012)	(0.009)	(0.011)
Mother Cognitive Skill$_t$	0.065**	0.019+	0.041**	0.008	0.068**	0.023*
	(0.009)	(0.011)	(0.010)	(0.012)	(0.010)	(0.011)
Mother Non-Cognitive Skill$_t$	0.312**	0.108**	0.276**	0.092**	0.253**	0.088**
	(0.010)	(0.012)	(0.010)	(0.012)	(0.010)	(0.012)
Missing Indicators	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y
N	15,017	9,697	15,056	9,701	15,045	9,699
R^2	0.207	0.120	0.138	0.047	0.176	0.144

Notes: UK Millennium Cohort Study, data from $t = 3$ and $t = 4$ (child age 7 and 11). Estimates are obtained using linear regressions. The dependent variable is obtained by combining all 4 subscales of the SDQ into a common factor in columns 1 and 2; columns 3 and 4 combine the SDQ sub-scores on 'emotional symptoms' and 'peer problems' into a measure of *internalizing* behaviour, and columns 5 and 6 use the SDQ sub-scores on 'hyperactivity/inattention' and 'conduct problems' to derive a measure of *externalizing* behaviour. Demographic controls as in footnote to Table 4. Standard errors are clustered at the individual level. $^+$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Appendix Table 2: Teachers versus Interviewers

	$t = 1$		$t > 1$	
N_t reported by:	Mother	Interviewer	Mother	Teacher
	(1)	(2)	(3)	(4)
Child Cognitive Skill$_t$	0.188**	0.309**	0.194**	0.264**
	(0.011)	(0.013)	(0.008)	(0.009)
Mother Cognitive Skill$_t$	0.132**	-0.018	0.078**	0.036**
	(0.010)	(0.010)	(0.008)	(0.008)
Mother Non-Cognitive Skill$_t$	0.268**	0.021	0.299**	0.091**
	(0.011)	(0.010)	(0.008)	(0.008)
Missing Indicators	Y	Y	Y	Y
Demographics	Y	Y	Y	Y
N	9,391	9,343	24,514	17,475
R^2	0.218	0.130	0.207	0.158

Notes: UK Millennium Cohort Study, data from $t = 1$ (child age 3) in column 1, and for $t = 2$ to $t = 4$ (child age 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Appendix Table 3: Teacher and School Controls

	Teacher Reported Measures, $t = 3$			
	Baseline	Teacher controls	School FE	Teacher FE
	(1)	(2)	(3)	(4)
Child Cognitive Skill$_t$	0.268**	0.262**	0.276**	0.281**
	(0.015)	(0.016)	(0.029)	(0.043)
Mother Cognitive Skill$_t$	-0.003	0.003	0.020	0.021
	(0.013)	(0.015)	(0.029)	(0.040)
Mother Non-Cognitive Skill$_t$	0.094**	0.088**	0.095*	0.101**
	(0.014)	(0.015)	(0.026)	(0.035)
Missing Indicators	Y	Y	Y	Y
Demographics	Y	Y	Y	Y
N	5,536	4,725	5,536	5,536
R^2	0.128	0.127	0.646	0.800
H_0: Teacher Controls = 0				
P-value:		0.835		

Notes: UK Millennium Cohort Study, data from $t = 3$ (child age 7). 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. Teachers' controls include gender, role (class teacher, or other e.g. TA), qualification and experience (in years). Standard errors are clustered at the individual level in column 1, at the school-level in column 3, and at the teacher level in columns 2 and 4. $^+$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table 4: The Production Function of Skills

	$t=2$				$t=3$				$t=4$			
	C_{t+1}	N_{t+1}	N^M_{t+1}	N^F_{t+1}	C_{t+1}	N_{t+1}	N^M_{t+1}	N^F_{t+1}	C_{t+1}	N_{t+1}	N^M_{t+1}	N^F_{t+1}
C_t	0.546*	0.020*	0.017	-0.008	0.927*	0.020*	-0.022	0.043	0.774*	0.054*	-0.025	0.020
	(0.018)	(0.005)	(0.014)	(0.017)	(0.030)	(0.010)	(0.029)	(0.037)	(0.033)	(0.013)	(0.032)	(0.021)
N_t	0.249*	0.951*	0.170*	0.131	0.019	1.139*	0.305*	0.041	0.022	0.721*	0.300*	0.119*
	(0.066)	(0.099)	(0.069)	(0.068)	(0.035)	(0.058)	(0.066)	(0.040)	(0.021)	(0.070)	(0.040)	(0.040)
C^M	0.013*	0.039*	0.012	0.008	0.001	-0.055*	-0.001	-0.001	0.028*	0.009	0.011	0.003
	(0.007)	(0.013)	(0.009)	(0.008)	(0.006)	(0.014)	(0.009)	(0.009)	(0.006)	(0.009)	(0.011)	(0.009)
N^M_t	0.004	0.027*	0.685*	0.037*	0.002	0.009	0.679*	0.062*	-0.006	0.021	0.627*	0.030
	(0.012)	(0.014)	(0.019)	(0.016)	(0.007)	(0.014)	(0.018)	(0.016)	(0.007)	(0.012)	(0.021)	(0.018)
C^F	0.033*	0.005	0.011	0.038*	0.019*	0.002	0.005	-0.002	0.021*	0.006*	0.034*	0.014
	(0.007)	(0.004)	(0.009)	(0.010)	(0.008)	(0.004)	(0.011)	(0.010)	(0.006)	(0.003)	(0.014)	(0.010)
N^F_t	0.008	0.007	0.012	0.722*	0.012	-0.002	0.036*	0.707*	-0.001	0.002	0.067*	0.704*
	(0.010)	(0.004)	(0.013)	(0.020)	(0.009)	(0.006)	(0.015)	(0.020)	(0.006)	(0.006)	(0.019)	(0.019)
Own $\times C_t$	-0.012*	-0.021*	0.004	0.024*	0.018*	-0.008	0.026*	0.017	-0.003	-0.027*	0.011	0.022
	(0.003)	(0.007)	(0.010)	(0.008)	(0.007)	(0.008)	(0.010)	(0.010)	(0.003)	(0.008)	(0.012)	(0.013)
Own $\times N_t$	-0.074*	-0.093*	0.077*	0.016	0.027*	0.010	0.091*	0.107*	0.015	-0.002	0.111*	0.068*
	(0.020)	(0.035)	(0.031)	(0.021)	(0.014)	(0.022)	(0.020)	(0.018)	(0.011)	(0.007)	(0.015)	(0.016)
Own $\times C^M$	-0.002	0.008	-0.008	0.006	0.001	-0.003	-0.015*	-0.010	-0.001	0.010*	-0.010*	0.004
	(0.003)	(0.006)	(0.006)	(0.006)	(0.003)	(0.006)	(0.005)	(0.006)	(0.003)	(0.004)	(0.005)	(0.006)
Own $\times N^M_t$	-0.040*	-0.009	-0.014*	0.061*	0.041*	0.017*	0.003	0.064*	0.003	0.007	0.005	0.055*
	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)
Own $\times C^F$	-0.005*	-0.009	0.003	-0.016*	0.003	0.008	-0.001	0.003	0.003	0.003	-0.003	-0.015*
	(0.002)	(0.005)	(0.006)	(0.006)	(0.004)	(0.005)	(0.006)	(0.006)	(0.003)	(0.005)	(0.010)	(0.005)
Own $\times N^F_t$	-0.034*	0.004	0.070*	0.001	0.036*	0.021*	0.065*	0.006*	0.005	0.008*	0.065*	0.010*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)	(0.005)
R^2	0.649	0.609	0.463	0.478	0.836	0.730	0.459	0.484	0.846	0.713	0.442	0.492

Notes: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in sections 4 and 5. Standard errors are obtained bootstrapping 30 times the original data set at the individual level. $t=2$ refers to the production function of skills from the first to the second period in our data set, i.e. from 3 to 5 years old. $t=3$ and $t=4$ are derived similarly. Each column corresponds to a different production function, defined in equation 15. * $p<0.05$.

Appendix Table 5: Policy Evaluation in the Presence of Distortions

	SDQ emot.	SDQ conduct	SDQ hyp.	SDQ peer	Q. Indep.	Q. Emot.	Q. Coop.
			$+1sd$ in N_1^M				
			Average Increase in Child Non-Cognitive Skill Measures				
$t=1$	0.142	0.150	0.081	0.093	0.023	0.090	
	(0.032)	(0.032)	(0.031)	(0.031)	(0.022)	(0.030)	
$t=2$	0.147	0.138	0.113	0.102	0.065	0.130	
	(0.024)	(0.023)	(0.025)	(0.027)	(0.019)	(0.027)	
$t=3$	0.131	0.117	0.091	0.109	0.059	0.115	0.084
	(0.022)	(0.021)	(0.026)	(0.024)	(0.021)	(0.026)	(0.022)
$t=4$	0.119	0.096	0.096	0.091			
	(0.023)	(0.017)	(0.021)	(0.021)			
			Percent due to Contamination				
$t=1$	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	
$t=2$	66.9%	61.3%	50.6%	48.2%	41.3%	55.5%	
	(15.2%)	(17.4%)	(20.0%)	(20.3%)	(23.8%)	(19.3%)	
$t=3$	60.1%	51.7%	33.7%	48.3%	16.3%	47.7%	36.4%
	(11.4%)	(12.0%)	(11.8%)	(11.8%)	(15.5%)	(11.8%)	(12.3%)
$t=4$	49.0%	41.1%	30.4%	28.5%			
	(8.3%)	(10.0%)	(9.8%)	(8.7%)			

Notes: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in sections 4 and 5. In the first panel, we report the impact on the SDQ measures of children non cognitive skills at different ages when maternal non-cognitive skills are increased by 1sd when the children are 3 years old. In the second panel we report the fraction of the impact that is due to contamination of the measures. Standard errors are obtained bootstrapping 30 times the original data set at the individual level.

Appendix Table 6: Policy Evaluation Motivated by Baker *et al.* (2008)

	SDQ emot.	SDQ conduct	SDQ hyp.	SDQ peer	Q. Indep.	Q. Emot.	Q. Coop.
				$+0.2sd$ in N_1^M and N_1			
				Average Increase in Child Non-Cognitive Skill Measures			
$t = 1$	0.120	0.149	0.132	0.108	0.055	0.139	
	(0.006)	(0.007)	(0.007)	(0.007)	(0.004)	(0.007)	
$t = 2$	0.104	0.109	0.107	0.100	0.071	0.114	
	(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.007)	
$t = 3$	0.095	0.097	0.096	0.095	0.074	0.101	0.086
	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)
$t = 4$	0.085	0.075	0.083	0.081			
	(0.005)	(0.004)	(0.005)	(0.005)			
				% due to contamination			
$t = 1$	23.7%	20.1%	12.4%	17.2%	8.4%	13.0%	
	(4.4%)	(3.6%)	(4.2%)	(5.3%)	(7.7%)	(3.9%)	
$t = 2$	20.3%	16.7%	11.4%	10.5%	8.2%	13.6%	
	(3.8%)	(3.9%)	(3.4%)	(3.1%)	(3.0%)	(4.1%)	
$t = 3$	19.8%	15.0%	7.7%	13.3%	3.1%	13.1%	8.6%
	(3.3%)	(2.9%)	(2.5%)	(2.7%)	(2.7%)	(2.8%)	(2.2%)
$t = 4$	19.3%	14.8%	9.8%	9.1%			
	(2.5%)	(3.0%)	(2.9%)	(2.0%)			

Notes: UK Millennium Cohort Study. Estimates are obtained using the model and estimation method outlined in sections 4 and 5. In the first panel, we report the impact on the SDQ measures of children non cognitive skills at different ages when maternal non-cognitive skills and children non-cognitive skills are both increased by 0.2sd when the children are 3 years old. In the second panel we report the fraction of the impact that is due to contamination of the measures. Standard errors are obtained bootstrapping 30 times the original data set at the individual level.