Child development and the production of noncognitive skills

Marcelo Gelati
FGV - EPGE

July 2024

JEL Classification J13, D1, J24
Keywords: Child development, cognitive skills, noncognitive skills, parental care

Abstract

This article develops a dynamic model of the production of children's cognitive and noncognitive skills, extending the identification method developed by Agostinelli and Wiswall (2023). The production of skills depends on the current stock of skills and on parental investment. Parental investment is influenced by various household characteristics. The model takes into account measurement errors. The stock of skills in the last period is used to determine children's future outcomes. The estimated technologies show that the mother's and child's noncognitive skills, along with family income, are the most important factors for parental investment. For the production of cognitive skills, the previous stock of cognitive skills and the parental investment are relevant factors. For the production of noncognitive skills, only the previous stock of noncognitive skills is relevant. The stock of noncognitive skills in the last period is more important than the stock of cognitive skills as a predictor of adult outcomes, both in terms of wages and educational level.

1 Introduction

There was significant improvement in the literature of children skills' formation in the last 20 years. It is now known that both cognitive and noncognitive skills play a significant part in later adult outcomes. It is also known that parental investments are crucial to fostering those skills. These discoveries are important to build intervention programs targeted to disadvantaged socioeconomic groups that can improve their relations with their children and guarantee a more promising future for them.

This paper contributes to the literature by extending what is done in the article by Agostinelli and Wiswall (2023). Their new identification procedure uses only the cognitive skills' production function, but to follow the literature developed by Cunha and Heckman (2007), Cunha and Heckman (2008), and Cunha et al. (2010), it is important to introduce cognitive and noncognitive skills' production functions in their algorithm. The traditional human capital models in economics treated childhood as a single period, assuming that inputs were perfect substitutes for both periods, for example, Becker and Tomes (1986). The literature on skill formation then proceeded to address the empirical facts discovered by the psychology literature and started postulating dynamic models to capture dynamic complementarity and self-productivity effects.

Is important to note that while the aforementioned literature is primarily focused on estimating the production function, there are some works in the direction of building structural models to understand household decisions on children's investments. They try to answer questions about resource allocation: how parents should spend their time? In the labor market or with their children? How parents should spend their money? In their own consumption or investing in their children? One major paper that studies this is Del Boca et al. (2013). Other works are Caucutt et al. (2020), Thomas (2023) and Agostinelli and Sorrenti (2021), although the last one uses the model only as a motivation for their reduced-form analyses on the rest of the paper.

The review made by Cunha et al. (2021) points to new directions to explore. The main challenges are to deal better with measurement error and the ordinality of scores, and to frame the problem in terms of sensitive and nonsensitive skills, instead of cognitive and noncognitive skills. I don't develop this here, but all problems pointed out by them will be dealt with. The relevant variables are assumed to be measured with error and there will be anchoring of the parameters in cardinal variables to guarantee interpretability of the results.

I use a log-log specification for both production functions and use the age invariance hypothesis for identification, following Agostinelli and Wiswall (2023). With my specification I'll model self-productivity effects, but won't capture dynamic complementarity effects.

Using the public data from the NLSY79 AND NLSY79 Child/YA, I found that mothers with a higher level of socioemotional skills and in a higher socioeconomic status invest more on their child. Also, that children with a larger stock of noncognitive skills tend to receive more investment. My estimates indicate that parental investment have larger effects in the production of cognitive skills at earlier ages, indicating that it is better to target younger children in the design of programs. Both production of cognitive and noncognitive skills display large self-productivity effects, meaning that the stock of skills in the current period are highly important to the production of skills in the next period. I also found that the stock of skills of children before reaching adulthood are highly important to predict adult outcomes. In particular, noncognitive skills have the larger effect, which points that an effective program target at children must teach parents not only to invest on them, but to give them a solid and stable background.

The paper is organized as follows. Sections 2 and 4 describe the models and the identification hypothesis and procedure, respectively. Sections 3 and 3.10 describe the NLSY datasets and some summary statistics. Finally, section 5 shows the results.

2 Models and measurement

My whole framework is composed of two models of child skill production, one law of parental investment, one model of adult outcomes, which is necessary for anchoring (discussed in detail in the following subsections), and a measurement model. The next sections are dedicated into dissecting all of the above.

2.1 Timing

In this class of models, it is essential to design a dynamic framework. This is the case, since some inputs might be more productive at some periods of time than in others, and also because the relative importance of inputs might be different at different stages. These more productive periods are called sensitive periods. In general, we'll have the periods $t = 1, \ldots, T$, where T is the period when the child reaches "adulthood", that is, when we consider that her level of skills can't be accumulated no longer. This is consistent with the literature for the case of cognitive skills, that states that at a certain period of time there won't be an evolution no more, except when people reach the elderly age where we observe a decline in cognitive skills. This poses no threat to our hypothesis, however, because this period isn't considered in my analysis.

The same can't be said about noncognitive skills, however. Although it is quite hard to modify emotional-behavioral qualities, they are in fact flexible - this is the whole purpose of psychotherapy and psychiatric drugs. I'll consider that this doesn't happen to the people in our sample given the age-range considered and the fact of the high rigidity of those skills.

In particular, I'll consider that the children start with 5 years old and reach adulthood at 14 years old. Together with this, I'll divide the cohorts every two years. So the age groups will be $5-6$, $7-8$, $9-10$, $11-12$, and $13-14$ years old, meaning that we will have periods $t = 1, 2, 3, 4, 5$. I'll start at five years old, as the measures I use are for children five or older. Although there are measures of child cognitive and noncognitive skills for younger children, they are not the same as the ones used from five years of age. So, to maintain consistency (and to preserve the age-invariance of measures, discussed below), this will be the initial age of the first period.

The last period ending at fourteen years old is a decision made based on the dataset. It contains a very small number of observations for children assessments for children aged fifteen or more. This doesn't poses a threat though, because cognitive skills start becoming very rigid once the child reaches ten years old. This also occurs to child noncognitive skills (Hampson et al., 2007; Roberts & DelVecchio, 2000), as much of the parental and societal influences have already been introduced to the child. The interactions between genetic characteristics and environmental conditions are already set and her socioemotional traits have been formed.

Finally, the restriction of segmenting the age groups in an interval of two years is due to the structure of the dataset I'll be using. The NLSY79 Child/YA (described in the data section) happens every two years. So, every mother and child is interviewed in this interval, and the separation between those cohorts is designed to capture the evolution of the child in each survey year.

2.2 Child's cognitive and noncognitive skills formation

The simplest case of skills formation function is when it is a single skill with only one type of parental investment, that is

$$\theta_{t+1} = f_t(\theta_t, I_t, \eta_t)$$

where θ_t is the stock of child skills in period t, I_t is the flow of parental investments, and η_t is a shock in that period.

It is worth pointing out that we are always talking about child's *stock* of skills. This is made evident by how the dependence appears in the function. That is, the child's level of skills in period $t + 1$ will be a function of her level of skills in the preceding period t. The literature coined the term self-productivity when this effect is positive. In contrast, the investment variable is *not* a stock measure (this will be shown in the investment function soon). It is a *flow* measure. At each period t the parents invest in the child and this in turn will impact how her level of skills change in the next period.

However, I'll extend this framework to take account of two different types of child skills and parental investments. My framework consists in

$$\begin{aligned}
\theta_{t+1} &= f_{\theta,t}(\theta_t, \kappa_t, I_t, \eta_{\theta,t}) \\
\kappa_{t+1} &= f_{\kappa,t}(\theta_t, \kappa_t, I_t, \eta_{\kappa,t})
\end{aligned}$$

where κ_t is the stock of noncognitive skills at period t. Also, $\eta_{\theta,t}$ and $\eta_{\kappa,t}$ are the contemporaneous shock of each of the functions. In this setup, there are two important differences. Now we must deal with two production functions at the same time. Not only this, but we also have that one type of skill is directly impacting the accumulation of the other skill. This effect is called cross-complementarity.

I assume a parametric functional form for those functions, following Agostinelli and Wiswall (2023). Assuming a log-log format, I'll write

$$\log \theta_{t+1} = \log A_{\theta,t} + \gamma_{1,\theta,t} \log \theta_t + \gamma_{2,\theta,t} \log \kappa_t + \gamma_{3,\theta,t} \log I_t + \eta_{\theta,t} \tag{1}$$

$$\log \kappa_{t+1} = \log A_{\kappa,t} + \gamma_{1,\kappa,t} \log \theta_t + \gamma_{2,\kappa,t} \log \kappa_t + \gamma_{3,\kappa,t} \log I_t + \eta_{\kappa,t} \tag{2}$$

where I'm assuming that (1) the shocks are independent of the latent variables on all periods; (2) the same skills shocks are independent of each other between periods; (3) the cross skills shocks are independent of each other between periods; and that (4) $E[\eta_{\theta,t}] = E[\eta_{\kappa,t}] = 0$.

I chose a log-log format to get a clean interpretation of the coefficients. It is a cleaner form of obtaining the elasticities of each input in the formation of skills. This means that I'll resign the estimation of the dynamic complementarities, which is how investments reacts across different levels of skills. As the question I'm posing does not requires studying those dynamic complementarities, this won't impact the rest of the study.

2.3 Parental investment law

The parental investment model is as follows

$$\log I_t = \alpha_{1,I,t} \log \theta_t + \alpha_{2,I,t} \log \kappa_t + \alpha_{3,I,t} \log M_\theta + \alpha_{4,I,t} \log M_\kappa + \alpha_{5,I,t} \log Y_t + \eta_{I,t} \quad (3)$$

where M_θ and M_κ are the mother's cognitive and noncognitive skills, respectively. Notice that these variables don't have the subscript t due to our hypothesis that once the adulthood is reached, the skills are stable over time. I also assume that the income Y_t impacts the flow of investments. Finally, $\eta_{I,t}$ and $\eta_{E,t}$ are the contemporaneous shocks of each laws of motion.

For these technologies, I'm assuming that (1) both technologies have the property of constant returns of scale ($\sum_{i=1}^{5} \alpha_{i,I,t} = 1$); (2) the shocks are independent of all latent variables over all periods; (3) the same investments shocks are independent of each other between periods; (4) the cross investments shocks are independent of each other between periods; and that (5) $E[\eta_{I,t}] = 0$.

2.4 Adult outcomes

All of the variables described in the preceding subsections (with the exception of earnings) are measured in an ordinal and arbitrary scale. This means that in order to meaningfully interpret our estimations, I need to scale these variables to a common cardinal variable. This process is called anchoring.

The anchor here will be the adult outcome A. The adult outcome model will have a log-log format and it will depend only on the stock of child skills at the last period, i.e.,

$$\log A = \alpha_{1,A} + \alpha_{2,A} \log \theta_5 + \alpha_{3,A} \log \kappa_5 + \pi' X + \eta_A \quad (4)$$

Where as usual I'm assuming that the error term η_A (1) is independent of the regressors; (2) is independent of all the other errors; and (3) $E[\eta_A] = 0$. Here, X is a vector of controls.

With this, the anchored variables are defined as

$$\log \tilde{\theta}_t := \alpha_{1,A} + \alpha_{2,A} \log \theta_t$$
$$\log \tilde{\kappa}_t := \alpha_{1,A} + \alpha_{3,A} \log \kappa_t$$

2.5 Variables measured with error

Finally, we have the measurement models. I assume that all skills and investments are measured with error, following the specifications

$$\begin{aligned}
Z_{\theta,t,m} &= \mu_{\theta,t,m} + \lambda_{\theta,t,m} \log \theta_t + \epsilon_{\theta,t,m} \\
Z_{\kappa,t,m} &= \mu_{\kappa,t,m} + \lambda_{\kappa,t,m} \log \kappa_t + \epsilon_{\kappa,t,m} \\
Z_{I,t,m} &= \mu_{I,t,m} + \lambda_{I,t,m} \log I_t + \epsilon_{I,t,m} \\
Z_{M_\theta,m} &= \mu_{M_\theta,m} + \lambda_{M_\theta,m} \log M_\theta + \epsilon_{M_\theta,m} \\
Z_{M_\kappa,m} &= \mu_{M_\kappa,m} + \lambda_{M_\kappa,m} \log M_\kappa + \epsilon_{M_\kappa,m}
\end{aligned}$$

where m is the subindex for the measure of the relevant latent variable at period t. I omit the subscript t if the variable isn't time-dependent. Notice that we can use as many measures as wanted. As for the assumptions, I assume the errors $\epsilon_{X,t,m}$ (1) are independent of all latent variables X_t; (2) are independent of each other contemporaneously; (3) are independent of each other across time; and (4) that $E[\epsilon_{X,t,m}] = 0$.

3 Data

3.1 Description of the data

I use data from the National Longitudinal Survey of Youth 1979 (NLSY79) and the National Longitudinal Survey of Youth 1979, Child and Young Adults (NLSY79 Child/YA), both conducted by the U.S. Bureau of Labor and Statistics. The NLSY79 is a longitudinal study that started in 1979 and was conducted annually until 1994. After that, it was conducted biennially until 2020, the date of the last interview. It was designed to track the lives of American youth and information about it, such as labor market behavior, family and school background, income, among other things. The first cohort had 12686 respondents with ages that ranged from 15 and 22.

There are three subsamples in the NLSY79. One is the cross-sectional subsample, which is representative of Americans born between January 1, 1957, and December 31, 1964 and has 6111 respondents. The other one is the supplemental subsample, comprised of 5295 Hispanic or Latino, black, and economically disadvantaged nonblack/non-Hispanic respondents. Finally, the last is the military subsample, where they collected 1280 respondents to represent the population serving in one of the four branches of the U.S. military. Since the 1984 interview, 1079 members of the military sample were dropped, and since the 1990 interview, all of the economically disadvantaged nonblack/non-Hispanic members were dropped from the supplemental subsample.

The NLSY79 Child/YA study is a supplementary longitudinal study designed to track the children of the respondents in the original study. It started in 1986 with 8336 children and was collected biennially until 2018, with a total of 11545 children. After 1994, they created a separated survey called Young Adults which focus on children 15 or older. There are several cognitive and socioemotional assessments, as well as medical, schooling and religious questions.

Both mother's and child's cognitive and noncognitive skills need three measures to be estimated, because they are the initial variables. We need at least two measures for parental investment, one as the main measure and the other one as an instrument. If a measure is negatively ordered in the database (meaning that a higher score implies a worse outcome) I'll reverse it so that all measures are positively ordered. The next subsections discuss in detail which variables are used from both NLSY79 and NLSY79 Child/YA.

3.2 Child's cognitive skills

For child's cognitive skills, I'm using the The Peabody Individual Achievement Test (PIAT) Math. It is administered to children of 5 years old or older, and it comprises of 84 math questions of increasing difficulty, each one with four options. The measure m' (see section 4.2) being used is the PIAT Reading Recognition test, which has the same format of the PIAT Math, but this one measures word recognition and pronunciation ability. The third measure is the PIAT Reading Comprehension test. It has 66 items of increasing difficulty, in which the child has to read a sentence and then select one of four pictures that represents the sentence. All of them are positively ordered.

3.3 Child's noncognitive skills

For child's noncognitive skills, I've selected the Behavior Problems Index (BPI). The interviewers ask the mother questions about children behavior in the last three months. They construct a three-item scale with responses (1) Often True; (2) Sometimes True; (3) Not True. There is a recoding process in which answers (1) and (2) are recoded to 1 and answer (3) is recoded to 0. The total score of each scale is just a summation of each question score, meaning that higher scores implies greater behavioral problems. My main measure is the Dependent subscale of BPI, which measures if the child is too much dependent of an adult or demands too much attention. The measure m' is the Hyperactive subscale, which measures if the child is restless, impulsive or has difficulty in concentrating. Finally, my last measure is the Headstrong subscale, which measures if the child is disobedient, stubborn or has strong temper.

3.4 Parental investment

Here, I use the Home Observation Measurement of the Environment (HOME) inventory. In particular, I choose variables from the Cognitive Stimulation subscale to measure parental investment. My main measure is the question "How often has any family member taken or arranged to take your child to any type of museum (children's, scientific, art, historical, etc.) within the past year?", and it has the following response options: (1) Never; (2) Once or twice; (3) Several times; (4) About once a month; (5) Once a week or more. The alternative measure I'm using as an instrument asks "About how many books does your child have?", and it has the responses (1) None; (2) 1 or 2 books; (3) 3 to 9 books; (4) 10 or more books. If the child is older than 10 years, then the alternatives are (1) None; (2) 1 to 9 books; (3) 10 to 19 books; (4) 20 or more books.

3.5 Mother's cognitive skills

For the mother's cognitive skills, I'm using three subtests of the Armed Services Vocational Aptitude Battery (ASVAB), which was taken in 1980. My first measure is the Numerical Operations subtest, which contains 50 items that asks simple arithmetic questions for the respondent. For the fixed measure m', I'm using the

Word Knowledge subtest, which tests the knowledge of the meaning of words of the respondent. It has 35 items. Finally, as my third measure, I'm using the 15 item subtest Paragraph Comprehension, which checks the understanding of the meaning of paragraphs. All of them are positively ordered.

3.6 Mother's noncognitive skills

For the mother's noncognitive skills, I use the Rosenberg Self-Esteem Scale, administered in 1980. It is a four-item Likert scale, where questions have the following response options: (1) Strongly Agree; (2) Agree; (3) Disagree; (4) Strongly Disagree. My main measure is the question "All in all, I am inclined to feel that I am a failure", which is positively ordered. For my measure m', I'm using the question "I am able to do things as well as most other people", which is negatively ordered. Finally, the third measure is the question "I wish I could have more respect for myself", which is positively ordered.

3.7 Income

The income variable is a constructed variable that gives the total net family income based on multiple sources. The interviewers ask questions based on wages, business' profits, inheritance, financial income, government assistance, and others. This is a family income variable, so all members in the household are considered for this composition. All values are in 2014 dollars.

3.8 Adult outcomes

I'm using both wages and salary income in 2014 dollars at age 30 and highest academic degree as of the 2014 interview to measure adult outcomes.

3.9 Data processing

The NLSY79 and NLSY79 Child/YA studies can be merged through the mother ID. In this way, we can discover the mother of each child. Notice, however, that we are only keeping track of the mothers, so the fathers in the original study will be discarded. After getting all children cohorts in the NLSY79 Child/YA from 1986 to 2014 and merging with the NLSY79 dataset, I get 11551 observations.

From that, I get a panel which has all measures by survey year. But as my model has the evolution of children by *age group*, I need to arrange all children in their appropriate age group, so I check their age and put the measures of all variables in each age group. This entails an important decision, since even though as the survey are biennially and my age groups are divided into slots of two years, sometimes a children might fall in the same age group between two surveys. This might happen, because a children can have her first interview at 5 years old and her second interview at 6 years and 10 months old, meaning that she will be placed in the 5-6 age group twice. This creates an issue, as the child had a cognitive development but is treated as remaining in the same spot.

There are two ways of solving this. One is by making all cohorts follow the rounds of interviews. So if a child is in one cohort at one interview, she will be in the next cohort in the next interview. This generates two greater problems: (1) there will be children under-aged in cohorts, as a children might be 8 years old an be in the 11-12 year cohort and (2) there will be children over-aged in cohorts. This last issue is specially problematic, because then there will be children with 16 years in the 13-14 year cohort and they won't have the measures of some assessments.

The other option is to drop the observation when a child repeats the cohort in two consecutive interviews. So, suppose a child is on the cohort 5-6 years old on two interviews. In this case, I'd drop the observations of the second interview. Nevertheless, she would appear on some higher cohort in the third interview, and then those measures will appear in the dataset. This is the path that I chose to follow. As the dropping is purely random - the child's age and the time interval of both interviews happens by chance -, there is no harm in doing so. Summing the observations of all cohorts gives us a total of 43964 children. The number of children that repeats cohorts in two consecutive interviews is 1637, approximately 3.7% of the total number of observations.

The last step of data processing is adjusting the measures that are negatively ordered. As the measures are ordinal, any mapping that change the order can be applied. Additionally, I drop variables that have a 999 score, which are safe to assume that were imputed incorrectly as a missing value (in the NLSY datasets missing values have negative scores).

3.10 Summary statistics

The summary statistics reported are of the treated dataset, as described in the previous subsection.

The dataset contains at least 5000 observations for each age group in the PIAT tests. We notice that the scores are increasingly monotonic by age, just as we would expect from the PIAT scores. It also contains at least 5000 observations for each age group in the BPI subscales. The means are somewhat high, meaning that there are more "well-behaved" children than not (at least for what is measured by the subscale).

Notice that the parental investment variables are somewhat stable over the age groups. The museum variable is in average low, meaning that parents don't take their children frequently to the museum. In contrast, the average of the book variable is almost at the ceiling, meaning that parents in general buy a lot of books to their children.

We can tell that the mothers have a high score in noncognitive skills, meaning that they are reporting that they see themselves in a good light. The average drops a little in the question "I wish I had more self-respect".

4 Identification and estimation

I use the age-invariance hypothesis developed by Agostinelli and Wiswall (2023) to identify my technologies. I prove here that it is possible to use this hypothesis on both cognitive and noncognitive skills and identify two skills production functions at the same time, something not done by them. With this, I can merge the old literature in child development with their new method of identification. The following subsections describe in detail this procedure.

4.1 Identification: age invariance

In order to identify all the parameters, I assume that both cognitive and noncognitive skills' measures are age invariant.

We say that a measure $Z_{X,t}$ is age invariant for the latent variable X_t in period t if for all $x \in \mathbb{R}$ we have that $E[Z_{X,t}|X_t = x] = E[Z_{X,t+1}|X_{t+1} = x]$. Intuitively, this means that the average value of the measure between periods must be the same conditional to the fact that the latent measure has the same value in those periods. Applying it to child's cognitive skills, this means that, in average, she will perform the same in the test if she did not experience an increase (or a decrease) in her cognitive skills.

Using our assumptions from the measurement model, we have that

Table 1: Children's cognitive and noncognitive skills measures summary

Cognitive skills					
Measure	Mean	Std	Min	Max	N
PIAT Math (ages 5-6)	15.770	6.855	0	60	6878
PIAT Recognition	17.549	7.224	0	84	6760
PIAT Comprehension	16.831	6.394	0	56	6520
PIAT Math (ages 7-8)	30.682	10.756	0	84	7049
PIAT Recognition	33.593	10.975	0	84	7041
PIAT Comprehension	31.274	10.170	0	78	6779
PIAT Math (ages 9-10)	43.798	10.695	0	84	7014
PIAT Recognition	46.060	12.767	0	84	7012
PIAT Comprehension	41.791	10.864	0	84	6934
PIAT Math (ages 11-12)	51.117	10.931	0	84	6735
PIAT Recognition	55.040	13.896	3	84	6726
PIAT Comprehension	48.775	11.714	0	84	6671
PIAT Math (ages 13-14)	55.142	11.810	0	84	4994
PIAT Recognition	60.731	14.237	0	84	5003
PIAT Comprehension	52.753	12.176	0	84	4969
Noncognitive skills					
Dependent Subscale (ages 5-6)	4.537	1.269	2	6	7282
Hyperactive Subscale	4.181	1.538	1	6	7228
Headstrong Subscale	3.812	1.644	1	6	7244
Dependent Subscale (ages 7-8)	4.663	1.268	2	6	7511
Hyperactive Subscale	4.119	1.594	1	6	7467
Headstrong Subscale	3.708	1.675	1	6	7465
Dependent Subscale (ages 9-10)	4.851	1.202	2	6	7488
Hyperactive Subscale	4.212	1.619	1	6	7449
Headstrong Subscale	3.772	1.676	1	6	7459
Dependent Subscale (ages 11-12)	4.993	1.154	2	6	7229
Hyperactive Subscale	4.319	1.598	1	6	7203
Headstrong Subscale	3.755	1.693	1	6	7200
Dependent Subscale (ages 13-14)	5.166	1.060	2	6	5525
Hyperactive Subscale	4.401	1.595	1	6	5495
Headstrong Subscale	3.762	1.696	1	6	5508

Table 2: Parental investments summary

Measure	Mean	Std	Min	Max	N
How often child was taken to museum (ages 5-6)	2.20	0.99	1	5	7302
How many books child has	3.73	0.63	1	4	7331
How often child was taken to museum (ages 7-8)	2.26	0.97	1	5	7540
How many books child has	3.77	0.57	0	4	7546
How often child was taken to museum (ages 9-10)	2.24	0.95	1	5	7510
How many books child has	3.64	0.69	1	4	7533
How often child was taken to museum (ages 11-12)	2.17	0.93	1	5	7237
How many books child has	3.45	0.82	1	4	7239
How often child was taken to museum (ages 13-14)	2.05	0.91	1	5	5539
How many books child has	3.35	0.87	1	4	5533

Table 3: Mother's cognitive and noncognitive skills summary

Measure	Mean	Std	Min	Max	N
Cognitive skills					
Numerical operations	31.416	11.832	0	50	10978
Word knowledge	22.122	8.605	0	35	10978
Paragraph comprehension	9.722	3.778	0	15	10978
Noncognitive skills					
I am inclined to feel that I am a failure	3.394	0.616	1	4	11121
I am as capable as others	3.303	0.570	1	4	11132
I wish I had more self-respect	2.826	0.814	1	4	11116

$$
\begin{aligned}
E[Z_{X,t}|X_t = x] &= \mu_{X,t} + \lambda_{X,t} \log x + E[\epsilon_{X,t}|X_t = x] \\
&= \mu_{X,t} + \lambda_{X,t} \log x + E[\epsilon_{X,t}] \\
&= \mu_{X,t} + \lambda_{X,t} \log x,
\end{aligned}
$$

where the equality in the second line comes from independence of errors, and the third line comes from the assumption of the zero average of errors.

Now, suppose that $Z_{X,t}$ is an age invariant measure. This implies that

$$
\begin{gathered}
E[Z_{X,t}|X_t = x] = E[Z_{X,t+1}|X_{t+1} = x] \implies \\
\mu_{X,t} + \lambda_{X,t} \log x = \mu_{X,t+1} + \lambda_{X,t+1} \log x
\end{gathered}
$$

This equality holds for all x if, and only if, we have that $\mu_{X,t} = \mu_{X,t+1}$ and $\lambda_{X,t} = \lambda_{X,t+1}$, implying that the measurement error models must have the same scale and location parameters for all periods.

As I'm going to assume that the measures for child's cognitive and noncognitive skills are age invariant for all periods, once I estimate the measurement error's parameters for the first period, I'll have found the values for all periods.

The hypothesis of age invariance is one about latent skills. It mainly says that what matters for the average performance in a test isn't a child's age, but the level of the latent skill that the child has. Why should we believe that this happens in here? Think about it in this way: suppose a child has done a math test designed by some professor. Now let's suppose that some other professor designed a math test. We won't expect that the child will perform the same in both tests in average, because it is a complete different test. The questions in one might be more difficult, or the test might be longer, or something entirely different. We have evidences to believe that in this case we cannot apply the age-invariance hypothesis.

We also won't be able to apply the age-invariance hypothesis if the child does the second math test within an interval of a single month. In this case, the child knows what are the correct answers and we will expect that she performs better because of it, even if she has the same level of latent skill. However, if the child performs the same test one year later (and doesn't have access to the answers), we will expect that her performance will depend only on her latent skill level. So if she has exactly the same latent skills level, she will perform, in average, the same. As this applies to both PIAT Math and BPI measures, I can assume this in my work.

4.2 Estimation: the algorithm

First, we need the initial conditions. We can normalize the mean of the latent skills in the initial period. That is, $(E[\log\theta_1], E[\log\kappa_1], E[\log\theta_M], E[\log\kappa_M]) = (0, 0, 0, 0)$.

Now, we can estimate the parameters from the skills measurement models. Notice that if we have two measures $Z_{X,t,m}$ and $Z_{X,t,m'}$ of latent variable X_t, then

$$\begin{aligned}Cov\left(Z_{X,t,m}, Z_{X,t,m'}\right) = Cov(&\mu_{X,t,m} + \lambda_{X,t,m}\log X_t + \epsilon_{X,t,m},\\ &\mu_{X,t,m'} + \lambda_{X,t,m'}\log X_t + \epsilon_{X,t,m'})\\ &= Cov\left(\lambda_{X,t,m}\log X_t + \epsilon_{X,t,m},\ \lambda_{X,t,m'}\log X_t + \epsilon_{X,t,m'}\right)\\ &= Cov\left(\lambda_{X,t,m}\log X_t,\ \lambda_{X,t,m'}\log X_t\right)\\ &= \lambda_{X,t,m}\lambda_{X,t,m'}Var\left(\log X_t\right),\end{aligned}$$

where the second line comes from the fact that μ is a constant and the third line comes from the hypothesis of the independence of the errors.

So, let's choose three measures for each skill θ_t, κ_t, M_θ, and M_κ: measure 1, measure m, and measure m'. In the first period, we will have that

$$\frac{Cov\left(Z_{X,1,m}, Z_{X,1,m'}\right)}{Cov\left(Z_{X,1,1}, Z_{X,1,m'}\right)} = \frac{\lambda_{X,1,m}}{\lambda_{X,1,1}}$$

Normalizing $\lambda_{X,1,1} = 1$, we identify the first period scale parameters.

$$\begin{aligned}\lambda_{\theta,1,m} &= \frac{Cov\left(Z_{\theta,1,m}, Z_{\theta,1,m'}\right)}{Cov\left(Z_{\theta,1,1}, Z_{\theta,1,m'}\right)}\\ \lambda_{\kappa,1,m} &= \frac{Cov\left(Z_{\kappa,1,m}, Z_{\kappa,1,m'}\right)}{Cov\left(Z_{\kappa,1,1}, Z_{\kappa,1,m'}\right)}\\ \lambda_{M_\theta,m} &= \frac{Cov\left(Z_{M_\theta,m}, Z_{M_\theta,m'}\right)}{Cov\left(Z_{M_\theta,1}, Z_{M_\theta,m'}\right)}\\ \lambda_{M_\kappa,m} &= \frac{Cov\left(Z_{M_\kappa,m}, Z_{M_\kappa,m'}\right)}{Cov\left(Z_{M_\kappa,1}, Z_{M_\kappa,m'}\right)}\end{aligned} \tag{5}$$

Notice that $E[Z_{X,t,m}] = \mu_{X,t,m} + \lambda_{X,t,m}E[\log X_t]$. As we normalized the average of the skills in the initial period to 0, the first period location parameters are identified:

$$
\begin{aligned}
\mu_{\theta,1,m} &= E[Z_{\theta,1,m}] \\
\mu_{\kappa,1,m} &= E[Z_{\kappa,1,m}] \\
\mu_{M_\theta,m} &= E[Z_{M_\theta,m}] \\
\mu_{M_\kappa,m} &= E[Z_{M_\kappa,m}]
\end{aligned} \tag{6}
$$

First, I'll show how to estimate the parental investment model. Define the following residual variables

$$
\begin{aligned}
\tilde{Z}_{X,1,m} &:= \frac{Z_{X,1,m} - \mu_{X,1,m}}{\lambda_{X,1,m}} \\
\tilde{\epsilon}_{X,1,m} &:= \frac{\epsilon_{X,1,m}}{\lambda_{X,1,m}}
\end{aligned} \tag{7}
$$

With this we can write all variables that are measured with error as $\log X_t = \tilde{Z}_{X,1,m} - \tilde{\epsilon}_{X,1,m}$. Using this, we rewrite the parental investment model as

$$
\begin{aligned}
\tilde{Z}_{I,1,m} - \tilde{\epsilon}_{I,1,m} = {} & \alpha_{1,I,1}(\tilde{Z}_{\theta,1,m} - \tilde{\epsilon}_{\theta,1,m}) + \alpha_{2,I,1}(\tilde{Z}_{\kappa,1,m} - \tilde{\epsilon}_{\kappa,1,m}) \\
& + \alpha_{3,I,1}(\tilde{Z}_{M_\theta,m} - \tilde{\epsilon}_{M_\theta,m}) + \alpha_{4,I,1}(\tilde{Z}_{M_\kappa,m} - \tilde{\epsilon}_{M_\kappa,m}) \\
& + \alpha_{5,I,1} \log Y_1 + \eta_{I,1}
\end{aligned}
$$

Isolating $Z_{I,1,m}$, we get that

$$
\begin{aligned}
Z_{I,1,m} = {} & \mu_{I,1,m} + \lambda_{I,1,m}\alpha_{1,I,1}\tilde{Z}_{\theta,1,m} + \lambda_{I,1,m}\alpha_{2,I,1}\tilde{Z}_{\kappa,1,m} + \lambda_{I,1,m}\alpha_{3,I,1}\tilde{Z}_{M_\theta,m} \\
& + \lambda_{I,1,m}\alpha_{4,I,1}\tilde{Z}_{M_\kappa,m} + \lambda_{I,1,m}\alpha_{5,I,1} \log Y_1 \\
& + \tilde{\epsilon}_{I,1,m} + \lambda_{I,1,m}(\eta_{I,1} - \alpha_{1,I,1}\tilde{\epsilon}_{\theta,1,m} - \alpha_{2,I,1}\tilde{\epsilon}_{\kappa,1,m} - \\
& \alpha_{3,I,1}\tilde{\epsilon}_{M_\theta,m} - \alpha_{4,I,1}\tilde{\epsilon}_{M_\kappa,m})
\end{aligned}
$$

First, we define $\beta_{0,I,1,m} := \mu_{I,1,m}$, $\beta_{j,I,1,m} := \lambda_{I,1,m}\alpha_{j,I,1}$, and $\pi_{I,1,m} := \tilde{\epsilon}_{I,1,m} + \lambda_{I,1,m}(\eta_{I,1} - \alpha_{1,I,1}\tilde{\epsilon}_{\theta,1,m} - \alpha_{2,I,1}\tilde{\epsilon}_{\kappa,1,m} - \alpha_{3,I,1}\tilde{\epsilon}_{M_\theta,m} - \alpha_{4,I,1}\tilde{\epsilon}_{M_\kappa,m})$, then rewrite the preceding equation as

$$
\begin{aligned}
Z_{I,1,m} = {} & \beta_{0,I,1,m} + \beta_{1,I,1,m}\tilde{Z}_{\theta,1,m} + \beta_{2,I,1,m}\tilde{Z}_{\kappa,1,m} + \beta_{3,I,1,m}\tilde{Z}_{M_\theta,m} \\
& + \beta_{4,I,1,m}\tilde{Z}_{M_\kappa,m} + \beta_{5,I,1,m} \log Y_1 \\
& + \pi_{I,1,m}
\end{aligned} \tag{8}
$$

As each $\epsilon_{X,1,m}$ is correlated with its measure $Z_{X,1,m}$, the new error term $\pi_{I,1,m}$ will be correlated with the independent variables. This implies that the least squares estimator will be inconsistent.

With this in mind, we can use a vector of alternative measures $Z_{1,m'} := (Z_{\theta,1,m'}, Z_{M_\theta,m'}, Z_{M_\kappa,m'})$ as an instrument and apply an instrumental variables approach to consistently estimate the parameters.

To see that $E[\pi_{I,1,m}Z_{1,m'}] = 0$, we just need to check independence of these variables, because then $E[\pi_{I,1,m}Z_{1,m'}] = E[\pi_{I,1,m}]E[Z_{1,m'}] = 0$ (given that $E[\pi_{I,1,m}] = 0$). Well, using all the assumptions we made that all errors are independent of measures (in particulars that measure errors are cross-independent), we have that

$$
\begin{aligned}
E[\tilde{\epsilon}_{I,1,m}|Z_{1,m'}] &= E[\tilde{\epsilon}_{I,m}] = 0 \\
E[\eta_{I,1}|Z_{1,m'}] &= E[\eta_{I,1}] = 0 \\
E[\tilde{\epsilon}_{\theta,1,m}|Z_{1,m'}] &= E[\tilde{\epsilon}_{\theta,1,m}] = 0 \\
E[\tilde{\epsilon}_{\kappa,1,m}|Z_{1,m'}] &= E[\tilde{\epsilon}_{\kappa,1,m}] = 0 \\
E[\tilde{\epsilon}_{M_\theta,m}|Z_{1,m'}] &= E[\tilde{\epsilon}_{M_\theta,m}] = 0 \\
E[\tilde{\epsilon}_{M_\kappa,m}|Z_{1,m'}] &= E[\tilde{\epsilon}_{M_\kappa,m}] = 0
\end{aligned}
$$

And this shows that the error $\pi_{I,1,m}$ is in fact independent of our instrument vector $Z_{1,m'}$. Estimating the equation above using two-stage least squares, we get all the coefficients, which will be used to recover the structural parameters. As $\lambda_{I,1,m}\alpha_{j,I,1} = \beta_{j,I,1,m}$ for $j = 1, 2, 3, 4$, we can sum on j and get that (as the investment function have constant returns of scale) $\lambda_{I,1,m} = \sum_{j=1}^{5} \beta_{j,I,1,m}$. Therefore, we have

$$
\alpha_{j,I,1} = \frac{\beta_{j,I,1,m}}{\sum_{j=1}^{5} \beta_{j,I,1,m}} \tag{9}
$$

And as $\mu_{I,1,m} = \beta_{0,I,1,m}$ by definition, we have identified all parameters related to the parental investment function.

Now we have to deal with the identification of both skill formation technologies. Let's first do it to the child's cognitive skills production function. We can rewrite it as

$$
\begin{aligned}
\tilde{Z}_{\theta,2,m} - \tilde{\epsilon}_{\theta,2,m} = \log A_1 \\
&+ \gamma_{1,\theta,1}(\tilde{Z}_{\theta,1,m} - \tilde{\epsilon}_{\theta,1,m}) + \gamma_{2,\theta,1}(\tilde{Z}_{\kappa,1,m} - \tilde{\epsilon}_{\theta,1,m}) \\
&+ \gamma_{3,\theta,1}(\tilde{Z}_{I,1,m} - \tilde{\epsilon}_{I,1,m}) + \eta_{\theta,1}
\end{aligned}
$$

Isolating $Z_{\theta,2,m}$, we get

$$\begin{aligned} Z_{\theta,2,m} = &\lambda_{\theta,2,m} \log A_{\theta,1} + \mu_{\theta,2,m} \\ &+ \lambda_{\theta,2,m}\gamma_{1,\theta,1}\tilde{Z}_{\theta,1,m} + \lambda_{\theta,2,m}\gamma_{2,\theta,1}\tilde{Z}_{\kappa,1,m} \\ &+ \lambda_{\theta,2,m}\gamma_{3,\theta,1}\tilde{Z}_{I,1,m} + \lambda_{\theta,2,m} \\ &+ \tilde{\epsilon}_{\theta,2,m} + \lambda_{\theta,2,m}(\eta_{\theta,1} - \gamma_{1,\theta,1}\tilde{\epsilon}_{\theta,1,m} - \gamma_{2,\theta,1}\tilde{\epsilon}_{\kappa,1,m} - \gamma_{3,\theta,1}\tilde{\epsilon}_{I,1,m}) \end{aligned}$$

Now, define $\delta_{0,\theta,1,m} := \lambda_{\theta,2,m} \log A_{\theta,1} + \mu_{\theta,2,m}$, $\delta_{j,\theta,1,m} := \lambda_{\theta,2,m}\gamma_{j,\theta,1}$, and $\pi_{\theta,1,m} := \tilde{\epsilon}_{\theta,2,m} + \lambda_{\theta,2,m}(\eta_{\theta,1} - \gamma_{1,\theta,1}\tilde{\epsilon}_{\theta,1,m} - \gamma_{2,\theta,1}\tilde{\epsilon}_{\kappa,1,m} - \gamma_{3,\theta,1}\tilde{\epsilon}_{I,1,m})$

$$\begin{aligned} Z_{\theta,2,m} = &\delta_{0,\theta,1,m} \\ &+ \delta_{1,\theta,1,m}\tilde{Z}_{\theta,1,m} + \delta_{2,\theta,1,m}\tilde{Z}_{\kappa,1,m} \\ &+ \delta_{3,\theta,1,m}\tilde{Z}_{I,1,m} \\ &+ \pi_{\theta,1,m} \end{aligned} \tag{10}$$

As before, estimating this via OLS will give inconsistent estimates. So we'll be using the instrument vector $W_{1,m'} := (Z_{\theta,1,m'}, Z_{\kappa,1,m'}, Z_{I,1,m'})$. We must have that $E[\pi_{\theta,1,m}W_{1,m'}] = 0$, and we can test this by checking the independence of those terms, because then $E[\pi_{\theta,1,m}W_{1,m'}] = E[\pi_{\theta,1,m}]E[W_{1,m'}] = 0$. Independence is assured by noticing that

$$\begin{aligned} E[\tilde{\epsilon}_{\theta,2,m}|W_{1,m'}] &= E[\tilde{\epsilon}_{\theta,2,m}] = 0 \\ E[\eta_{\theta,1}|W_{1,m'}] &= E[\eta_{\theta,1}] = 0 \\ E[\tilde{\epsilon}_{\theta,1,m}|W_{1,m'}] &= E[\tilde{\epsilon}_{\theta,1,m}] = 0 \\ E[\tilde{\epsilon}_{\kappa,1,m}|W_{1,m'}] &= E[\tilde{\epsilon}_{\kappa,1,m}] = 0 \\ E[\tilde{\epsilon}_{I,1,m}|W_{1,m'}] &= E[\tilde{\epsilon}_{I,1,m}] = 0 \end{aligned}$$

Therefore, after estimation of the reduced form parameters, we recover the structural parameters by

$$\begin{aligned} \log A_1 &= \frac{\delta_{0,1,m} - \mu_{\theta,2,m}}{\lambda_{\theta,2,m}} \\ \gamma_{j,1} &= \frac{\delta_{j,1,m}}{\lambda_{\theta,2,m}}, \quad j = 1, 2, 3 \end{aligned} \tag{11}$$

This means that if we assume that $Z_{\theta,1,m}$ is an age invariant measure for θ_1, then $\mu_{\theta,2,m} = \mu_{\theta,1,m}$ and $\lambda_{\theta,2,m} = \lambda_{\theta,1,m}$, which we have already estimated. Thus, the parameters from the cognitive skills production function can be recovered. As the cognitive and noncognitive skills production function are identical, we can estimate the parameters from the last one exactly the same way.

With this, the estimation for the first period is complete. It only remains to repeat the steps outlined before until the final period.

When all periods are estimated, what remains to be done is the estimation of the adult outcome function. Notice that we have

$$\begin{aligned}\log A &= \alpha_{1,A} + \alpha_{2,A} \log \theta_5 + \alpha_{3,A} \log \kappa_5 + \eta_A \\ &= \alpha_{1,A} + \alpha_{2,A} \tilde{Z}_{\theta,5,m} + \alpha_{3,A} \tilde{Z}_{\kappa,5,m} + (\eta_A - \alpha_{2,A} \tilde{\epsilon}_{\theta,5,m} - \alpha_{3,A} \tilde{\epsilon}_{\kappa,5,m})\end{aligned}$$

so we can use measures of cognitive and noncognitive skills and estimate it via 2SLS as before.

5 Results

The results are reported in the following tables. I'm using 100 bootstraps clustered at the family level to compute standard errors. The bootstrap at family level is adopted to take care of intra-family correlations and guarantee that we have i.i.d groups. Standard errors are reported in parentheses.

5.1 Investment function

Looking at the table below, we have significance in both noncognitive skills and on income. The child cognitive skills parameter is significant only on the first age group. The mother cognitive skills parameter isn't significant.

We observe that there are two main forces driving the investment on children: socioemotional skills (both from the mother and the child) and income. For example, in the first age group. a 100% increasing on mother socioemotional skills generates approximately a 45% increase in the parental investment. A 100% increase in the child socioemotional skills generates approximately a 30% increase on the parental investment. This goes in line with the intuition that stabler mothers invest more on children and well-behaved children receive more attention from the mothers.

Table 4: Estimates for investment function

	Parental investment function			
	Ages 5-6	Ages 7-8	Ages 9-10	Ages 11-12
Log cognitive	.028	.008	.012	-.002
skills (θ_t)	(.01)	(.006)	(.009)	(.007)
Log noncognitive	.315	.188	.233	.186
skills (κ_t)	(.091)	(.082)	(.084)	(.089)
Log mother	.006	.014	.009	.003
cognitive skills (M_θ)	(.011)	(.015)	(.013)	(.011)
Log mother	.439	.572	.592	.644
noncognitive skills (M_κ)	(.123)	(.116)	(.125)	(.106)
Log income (Y_t)	.212	.219	.154	.169
	(.055)	(.065)	(.071)	(.048)
n	5058	5187	5296	5070

5.2 Production functions

For the cognitive skills production function, we have significance throughout the estimates, except on the noncognitive skills. The effect is small for all age groups, except on the second one, with a 31.6% variation.

Except from that, the results are exactly what have been found on the literature. We have a high self-productivity, meaning that a higher stock of skills in one period have a large impact on the stock of skills of the next period, and that parental investments are really important for the development of cognitive skills, specially on early stages. The effect is decreasing with age, but nevertheless high (even on the last age group).

For the noncognitive skills production function, the effects are a lot smaller. Both cognitive skills and parental investment don't are a great contributor to the accumulation of noncognitive skills in the next period. Nevertheless, the self-productivity effect is seen here too.

Table 5: Estimates for cognitive skills production function

Cognitive skills production function				
	Ages 5-6	Ages 7-8	Ages 9-10	Ages 11-12
Log cognitive skills (θ_t)	1.241	.769	.897	1.045
	(.041)	(.023)	(.020)	(.025)
Log noncognitive skills (κ_t)	.059	.286	-.095	.049
	(.278)	(.221)	(.247)	(.289)
Log parental investment (I_t)	1.964	1.713	1.372	.49
	(.673)	(.450)	(.440)	(.347)
Log TFP (A_t)	11.264	12.765	7.783	2.124
	(1.525)	(1.336)	(1.053)	(.989)
n	5121	5372	5439	4019

Table 6: Estimates for noncognitive skills production function

Noncognitive skills production function				
	Ages 5-6	Ages 7-8	Ages 9-10	Ages 11-12
Log cognitive skills (θ_t)	.007	.003	.01	.013
	(.004)	(.002)	(.003)	(.003)
Log noncognitive skills (κ_t)	.678	.653	.656	.621
	(.039)	(.030)	(.032)	(.029)
Log parental investment (I_t)	.078	.062	.114	-.072
	(.063)	(.045)	(.054)	(.046)
Log TFP (A_t)	-.039	.035	-.25	.031
	(.146)	(.098)	(.123)	(.102)
n	5257	5524	5623	4214

5.3 Adult outcomes

This regression was made with the following controls: if the family lives in urban area, the marital status of the mother, the highest grade the mother achieved, and if the mother is in the workforce or not.

The most compelling findings come from the adult outcomes function. We can see that both stock of cognitive skills highly significant for how well the child performed in the future, but the most relevant one is the stock of noncognitive skills. For example, a doubling of the stock of noncognitive skills at ages $13-14$ is responsible for an approximately 25% increase of their wages at age 30, and it is responsible for a 12% variation in the highest academic degree achieved.

Although the cognitive skills are significant for those outcomes, they have a smaller impact. For example, we only observe a 3% and 1.5% variation in wages and academic degree, respectively, as the stock of cognitive skills doubles at ages $13-14$. This implies that the so called "soft" skills play an important role in your success later in life.

Table 7: Estimates for adult outcome function

	Log wages at 30		Log highest academic degree	
Log cognitive skills (θ_5)	0.030	0.028	0.014	0.011
	(0.004)	(0.005)	(0.001)	(0.001)
Log noncognitive skill (κ_5)		0.197		0.101
		(0.057)		(0.017)
n	1617	1554	3305	3195

6 Conclusion

The direction of future research this paper points is to identify which are the most important aspects of children education that can lead them to develop firm bases as stable and functional adults, as it is seen that the stock of noncognitive skills are the most relevant predictors of adult success. This could be done trying to find which investment variables capture this.

In this way, it is really important to find how home environment impacts the effect in the production of child noncognitive skills. If quality parental interaction with the children boosts the stock of their socioemotional skills, we can use this information to create programs that teach parents how to effectively raise their children, showing them how this affects their future outcomes.

Finally, one major point is try to solve the identification problem for other error hypotheses. The assumptions on errors made here might be too stringent, as during the assessment a interview for one measure might affect the measure of the other, so correlations between them should be addressed.

7 Bibliography

Agostinelli, F., & Sorrenti, G. (2021). *Money vs. time: Family income, maternal labor supply, and child development* (ECON - Working Papers No. 273). Department of Economics - University of Zurich. https://EconPapers.repec.org/RePEc:zur:econwp:273

Agostinelli, F., & Wiswall, M. (2023, June). *Estimating the technology of children's skill formation* (Working Paper No. 22442). National Bureau of Economic Research. https://doi.org/10.3386/w22442

Becker, G. S., & Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics*, *4*(3), S1–S39. Retrieved February 20, 2024, from http://www.jstor.org/stable/2534952

Caucutt, E. M., Lochner, L., Mullins, J., & Park, Y. (2020, September). *Child skill production: Accounting for parental and market-based time and goods investments* (Working Paper No. 27838). National Bureau of Economic Research. https://doi.org/10.3386/w27838

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31–47. https://doi.org/10.1257/aer.97.2.31

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, *43*(4), 738–782. https://doi.org/10.3368/jhr.43.4.738

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, *78*(3), 883–931. https://doi.org/https://doi.org/10.3982/ECTA6551

Cunha, F., Nielsen, E., & Williams, B. (2021). The econometrics of early childhood human capital and investments. *Annual Review of Economics*, *13*(1), 487–513. https://doi.org/10.1146/annurev-economics-080217-053409

Del Boca, D., Flinn, C., & Wiswall, M. (2013). Household Choices and Child Development. *The Review of Economic Studies*, *81*(1), 137–185. https://doi.org/10.1093/restud/rdt026

Hampson, S. E., Andrews, J. A., Barckley, M., & Peterson, M. (2007). Trait stability and continuity in childhood: Relating sociability and hostility to the five-factor model of personality. *Journal of research in personality*, *41*(3), 507–523.
Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological bulletin*, *126*(1), 3.
Thomas, K. (2023, April). *How to close the skill gap? parental background and children's skill development in indonesia* (tech. rep.). https://github.com/KatherinaThomas/Documents

8 Appendix

8.1 Adding controls in the investment function

In this section, instead of having the controls in the outcome function, the controls are in the investment function. This means that the parental investment law is given by

$$\begin{aligned}\log I_t = \alpha_{1,I,t} \log \theta_t + \alpha_{2,I,t} \log \kappa_t + \alpha_{3,I,t} \log M_\theta \\ + \alpha_{4,I,t} \log M_\kappa + \alpha_{5,I,t} \log Y_t + \pi' X_t + \eta_{I,t}\end{aligned}$$

Here, X_t is the vector of controls. In particular, I'm using if at period t the mother is employed, the highest grade she achieved, her marital status, and if the family lives in a urban area.

The interesting part of this exercise is that those variables are relevant to parental investment, and as these estimations impact the values of the variables used in the other models, the exogeneity will be transferred to them.

The results in the parental investment, cognitive and noncognitive skills, and adult outcomes models is as follows:

Table 8: Estimates for investment function

	Parental investment function			
	Ages 5-6	Ages 7-8	Ages 9-10	Ages 11-12
Log cognitive	.017	.009	.011	-.004
skills (θ_t)	(.011)	(.007)	(.011)	(.009)
Log noncognitive	.375	.189	.260	.194
skills (κ_t)	(.124)	(.097)	(.115)	(.116)
Log mother	-.008	-.001	-.007	-.012
cognitive skills (M_θ)	(.014)	(.018)	(.016)	(.014)
Log mother	.280	.502	.513	.645
noncognitive skills (M_κ)	(.195)	(.178)	(.183)	(.144)
Log income (Y_t)	.337	.302	.224	.177
	(.094)	(.105)	(.099)	(.065)
n	4918	5075	5199	4997

Table 9: Estimates for cognitive skills production function

	Cognitive skills production function			
	Ages 5-6	Ages 7-8	Ages 9-10	Ages 11-12
Log cognitive skills (θ_t)	1.241	.769	.897	1.045
	(.041)	(.023)	(.020)	(.025)
Log noncognitive skills (κ_t)	.059	.286	-.095	.049
	(.278)	(.221)	(.247)	(.289)
Log parental investment (I_t)	1.565	1.381	1.108	.430
	(.569)	(.415)	(.400)	(.312)
Log TFP (A_t)	11.413	12.962	11.279	1.757
	(2.291)	(1.873)	(2.397)	(1.203)
n	5121	5372	5439	4019

Table 10: Estimates for noncognitive skills production function

Noncognitive skills production function				
	Ages 5-6	Ages 7-8	Ages 9-10	Ages 11-12
Log cognitive skills (θ_t)	.007	.003	.01	.013
	(.004)	(.002)	(.003)	(.003)
Log noncognitive skills (κ_t)	.678	.653	.656	.621
	(.039)	(.030)	(.032)	(.029)
Log parental investment (I_t)	.062	.050	.092	-.063
	(.052)	(.037)	(.045)	(.043)
Log TFP (A_t)	-.033	.042	.039	.086
	(.210)	(.128)	(.243)	(.121)
n	5257	5524	5623	4214

Table 11: Estimates for adult outcome function

	Log wages at 30		Log highest academic degree	
Log cognitive skills (θ_5)	0.030	0.028	0.014	0.011
	(0.004)	(0.005)	(0.001)	(0.001)
Log noncognitive skill (κ_5)		0.197		0.101
		(0.057)		(0.017)
n	1617	1554	3305	3195