IDB WORKING PAPER SERIES No. IDB-WP-527

The Socio-Economic Gradient of Child Development

Cross-Sectional Evidence from Children 6-42 Months in Bogota

Marta Rubio-Codina
Orazio Attanasio
Costas Meghir
Natalia Varela
Sally Grantham-McGregor

January 2013

Inter-American Development Bank
Social Protection and Health Division

The Socio-Economic Gradient of Child Development

Cross-Sectional Evidence from Children 6-42 Months in Bogota

Marta Rubio-Codina
Orazio Attanasio
Costas Meghir
Natalia Varela
Sally Grantham-McGregor

Inter-American Development Bank

2013

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

The socio-economic gradient of child development: cross-sectional evidence from children 6-42 months in Bogotá / Marta Rubio-Codina, Orazio Attanasio, Costas Meghir, Natalia Varela, Sally Grantham-McGregor.
p. cm. — (IDB Working Paper Series ; 527)
Includes bibliographic references.
1. Child development—Social aspects—Colombia. 2. Child development— Economic aspects—Colombia. I. Rubio-Codina, Marta. II. Attanasio, Orazio. III. Meghir, Costas. IV. Varela, Natalia. V. Grantham-McGregor, Sally. VI. Inter-American Development Bank. Social Sector. VII. Series.
IDB-WP-527

http://www.iadb.org

scl-sph@iadb.org

www.iadb.org/SocialProtection

The Socio-Economic Gradient of Child Development:

Cross-Sectional Evidence from Children 6-42 Months In Bogota

Marta Rubio-Codina, Institute for Fiscal Studies
Orazio Attanasio, University College London and Institute for Fiscal Studies
Costas Meghir, Yale University and Institute for Fiscal Studies
Natalia Varela, Université Laval
Sally Grantham-McGregor, Institute of Child Health, University College London[†]

Abstract: We study the socio-economic gradient of child development on a representative sample of low- and middle-income children aged 6-42 months in Bogota, using the Bayley Scales of Infant Development, a high quality test based on direct observation of the child's abilities. We find a statistically significant difference between children in the 90th and 10th percentile of the wealth distribution in our sample of 0.33 standard deviations (SD) in cognition, 0.29 SD in receptive language and 0.38 SD in expressive language at 14 months. The socio-economic gap increases substantially with age to 1 SD (cognition), 0.80 SD (receptive language) and 0.69 SD (expressive language) by 42 months. While the gap persists after controlling for mediating factors such as parental and biomedical characteristics, the level of stimulation in the home, and the quality of the institutional care setting; its size is significantly reduced by variables related to the home environment – i.e. parental investments in care quantity and quality. These findings have important implications for the design of well-targeted, effective and timely interventions that promote early childhood development.

JEL Codes: I0

Keywords: Early Childhood Development, Socio-Economic Gradient, Colombia

---

[†]The opinions expressed herein are those of the authors and not necessarily of the institutions they represent. This research would not have been possible without the outstanding work of Pablo Muñoz, Mara Minski, Belén Gómez, Juan Fernando Trujillo and Hanner Sánchez on the field; and funding from the Inter-American Development Bank for data collection. We also thank Camila Fernández and researchers at EdePo@IFS for useful comments and suggestions; and José Guerra for research assistance. Rubio-Codina's research time was financed by the Leverhulme Trust Early Career Fellowship ECF/2008/0170 on "CCTs, Child Care and Child Development". Attanasio's research time was partially financed by the European Research Council Advanced Grants 249612 on "Exiting Long Run Poverty: The determinants of asset accumulation in developing countries" and by the ESRC Professorial Fellowship ES/K010700/1 "The accumulation of human capital in developing countries". All errors are our own.

I. Introduction

In low and middle-income countries, an estimated 219 million (or 39% of) children under five years fail to reach their developmental potential due to exposure to risk factors such as illness, nutritional deficiencies, inadequate learning opportunities, less responsive parents, etc.—all of which are associated to poverty (Grantham-McGregor et al. 2007). These factors affect cognitive abilities beyond the effect of genetics (Hackman and Farah 2009) and generate developmental delays that are difficult to compensate later on in life as early childhood is a particularly sensitive period for brain development (Shonkoff 2010).

An increasing number of empirical studies have provided evidence on the long-term consequences of the exposure to risk and protective factors in early childhood. Low levels of early childhood development (ECD) negatively influence school readiness and performance (Hoddinott et al. 2008; Walker et al. 2005, 2006), employability and earnings potential (Maluccio et al. 2009), and adult health, competence and overall well-being (Walker et al. 2011).[1] As these are likely to translate into higher fertility rates and lower resources for the next generation, poor ECD outcomes also play an important part in the intergenerational transmission of poverty (Grantham-McGregor et al. 2007; Heckman and Masterov 2007).

This evidence has resulted in increased consensus—both amongst academics and policy makers—on the importance of the early years for human capital accumulation and individual performance throughout life. The negative consequences of low ECD levels undermine the expected social and economic benefits of public (governmental) and private (parental) investments in children's health and education at later stages in life. Moreover, they are also likely to affect the economy in the aggregate as the quality of the human resources available in the labor market and their ability to effectively contribute to the country's economy are challenged. Given the high private and public returns in the short, medium and long run of better ECD outcomes, understanding *when* and *why* disadvantages in child development start is critical in order to design well-targeted, effective and timely interventions that take into account the role of mediating factors, that can remediate them.

The association of poverty and socio-economic status (SES)—as measured by income, wealth or parental education—on health outcomes and cognitive achievement has been well documented in developed countries. Overall, the evidence indicates a positive association between SES and child development (Blau 1999; Bradley and Corwyn 2002; Aughinbaugh and Gittleman 2003), with the income gradient widening as children get older, at least in terms of children's health (Case et al.

[1] See also the review by Almond and Currie (2010).

2002). While less is known in developing countries about the link between SES and ECD outcomes, an increasing number of studies are documenting patterns similar to those observed in western economies across the world. We briefly discuss some of the papers most relevant to our work next.

One of the first studies in this area is that by Ghuman et al. (2005) on a sample of low-income children younger than 3 years in the Philippines. The authors find evidence that children in households with fewer assets have lower performance on a test of receptive language skills, purposively developed for that population. More recently, Paxson and Schady (2007) analyze a sample of children 3 to 6 years in low-income households in rural Ecuador. They find that children from families in the 90th percentile of the wealth distribution in their sample perform almost 2 Standard Deviations (SD) higher on receptive language, as measured on the Spanish version of the *Peabody Picture Vocabulary Test* (PPVT), compared to children from families in the 10th percentile. Macours et al. (2012) also find a positive SES gradient of child development on the PPVT, amongst other tests, for children aged 3 to 7 living in highly disadvantaged communities in rural Nicaragua. In Colombia, Bernal and Van Der Werf (2011) obtain similar results on the PPVT when comparing children 3 to 10 years in the lowest and highest third of the country's wealth distribution, using cross-sectional data from a nationally representative sample. The gap is largest amongst children in urban areas, and more so for children 4.5 and 8.5 years.

Similarly, Fernald et al. (2011) assess a nationally representative sample of children 3 to 6 years in Madagascar. They find that children from families in the top wealth quintile or whose mothers had secondary education perform significantly better across a range of measures of cognitive and language development compared with children of women in the lowest wealth quintile or women with no education. These differences are largest for receptive language, working memory and memory of phrases, and double between age 3 and age 6. Findings from two recent World Bank studies on children 3 to 5 years in rural areas in Mozambique (Bruns et al. 2010) and Cambodia (Filmer and Nadeau 2010) are in line with this evidence.

Despite these studies, however, very little is known about the gap before 3 years of age. In addition, with the exception of Bernal and Van Der Werf (2011) and Fernald et al. (2011), the studies above assess children from very disadvantaged families in rural areas and hence lack national representativeness. Moreover, they often focus on a limited number of specific functions that proxy for cognitive development—such as receptive language using the PPVT, short term memory, etc.—as opposed to using a global measure that includes a comprehensive range of cognitive and language functions.

Two very recent exceptions to the above are recent works by Hamadani et al. (2012) and Fernald et al. (2012). The former collects high quality measures of cognitive development—the *Bayley Scales of Infant Development* (BSID-II) and the *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI-II)—child nutritional status and the home environment for a panel of poor children in rural Bangladesh over a period of five years starting at birth, which allows controlling for initial conditions in the analysis. The authors find a cognitive gap associated with poverty as early as 7 months of age. Fernald et al. (2012) investigate the SES gap in language, gross motor and personal-social development for children 3-23 months in rural India, Indonesia, Peru and Senegal, using the *Extended Ages and Stages Questionnaire* (EASQ), which is based on maternal reports. They find differences in EASQ total scores of 0.27 to 0.48 SD between children in the first and fifth quintile of the within-country household wealth distribution. These wealth gradients are increasing with age in most countries.

In this paper, we study a representative sample of children aged 6 to 42 months living in the lowest 3 (out of 6) SES strata in the city of Bogota, Colombia. These 3 strata represent 85% of the city's population. We measure SES status by an index of household wealth and proceed to estimate the SES gradient in child development and to describe how it changes with age. We assess child development using five of the six sub-scales of the latest version of the *Bayley Scales of Infant and Toddler Development* (Bayley-III): cognitive, receptive language, expressive language, fine motor, gross motor and socio-emotional. The Bayley offers a comprehensive assessment of the child's performance in a set of tasks and activities by direct observation and is the most commonly used test of development in the world for children up to 36 months.

Hence, our study fills the gaps in the existing literature for developing countries in a number of ways. First, we quantify the size of the SES gap of child development *at very early ages—from 6 to 42 months*. Second, our sample is *representative of low- and middle-income groups* in an *urban* environment. Third, we examine the SES gradient across the age range in order to establish when it becomes statistically significant and how it evolves with age. Fourth, we study the SES gradient across developmental domains (cognition, receptive and expressive language, fine and gross motor, and socio-emotional development) to determine whether there are differences in the formation and evolution of the gap across these areas, hence offering a comprehensive assessment. Finally, we exploit rich individual and household information to examine the roles of other factors—such as maternal education, nutritional status of the child, the quality of the home environment and preschool attendance—to the SES gap in child development.

We find a sizeable gap in child development that is statistically significant since very early ages. We estimate a significant average gap between children in households in the 90th and the 10th percentile of a wealth index distribution of 0.33 SD of a z-score at 14 months for cognitive development, and of 0.25 SD and 0.35 SD at 12 months for receptive and expressive language, respectively. Moreover, in line with the evidence above, the gap increases substantially and monotonically with age, to levels of 1.00 SD for cognition, 0.80 SD for receptive language and 0.69 SD for expressive language by 42 months. The gaps in fine motor and socio-emotional development become statistically significant at 22 months, when they achieve magnitudes of 0.23 SD and 0.29 SD, respectively. These differences peak at 0.53 SD and 0.64 SD by 42 months. There are no significant differences in gross motor skills between children of different SES backgrounds. Interestingly, the SES gap persists after controlling for other variables, including maternal education, although it is significantly reduced by parental factors and by factors related to the quality of the home environment—i.e. by investments in the quality and quantity of childcare. This leaves some room for policy intervention.

The remainder of the paper is organized as follows. The next section discusses the study design, the sample and the data. Section III lays out the empirical strategy and presents the SES (wealth) effects by age and developmental domain. In Section IV, we study the contribution of other variables to the SES gap. Section V discusses our main findings and concludes.

II. Study Design: Sample and Data

The population we analyze in this paper are children 6 to 42 months living in families in Bogota. To estimate the SES gradient, it would be desirable to study a sample representative of the entire population of the city. Constructing such sample is made easier by the socio-economic stratification of the city, which was first introduced in the 1980s as a mechanism to cross-subsidize basic public services (drinking water, electricity, sewage, gas and telephone). Following the 1994 law of Public Services, the Department for National Planning classified entire urban blocks into "*estratos*" according to their location and quality of infrastructure and housing.[2] This scheme divides the city in 6 SES strata and, in principle, is revised every five years. For our sampling exercise, we combined data from the 2005 Census and the 2001 Cadastre, which classified 12.6% of all residential blocks in Bogota as *estrato* 1 (E1), 38.9% as *estrato* 2 (E2), 36.6% as *estrato* 3 (E3), 6.9% as *estrato* 4 (E4), 2.9%

[2] *Ley 142 de Servicios Públicos*.

as *estrato* 5 (E5), and the remaining 2.2% residential blocks as estrato 6 (E6).[3] More recently, the distribution seems to have shifted slightly to the right.[4]

While in theory it would be very useful to construct a representative sample of the entire city, in practice it is very difficult to contact the higher strata and obtain their consent to participate in a study of this nature.[5] For this reason, at the beginning of the project, we decided to exclude the wealthiest two sectors (E5 and E6) and focus on the first 4 *estratos*, which represent about 95% of the population of the city. We designed the original study sample to be balanced across E1 to E4 and across 8-month range age groups (6 to 14, 15 to 23, 24 to 32, and 33 to 41 months), with 90 children in each stratum-age cell for a total of 1,440 children in 240 blocks.

Data were collected in three stages. First, once neighborhoods and blocks within them were selected, we visited all households in a block in order to identify those with children 6-42 months. Next, trained interviewers carried out a household survey on a random sub-sample of these children.[6] This survey collected basic socio-economic information on the household (demographic composition, education level and employment status for household members, dwelling characteristics, assets, etc.); as well as information on the child's nutritional status (birth weight, gestational age, breastfeeding, weaning, etc.) and care arrangements, both formal and informal. We also administered a slightly modified *Family Care Indicator* (FCI), which collects, by direct observation, the number of newspapers, magazines and books for adults in the household; the number of different types of toys the child usually plays with; and by maternal/caregiver report, the number of different types of play activities the child engaged in with an adult over the 7 days before the interview.[7] We will use these variables in the analysis of mediation in the acquisition of skills. In a third and final stage, the Bayley-III test was administered in the presence of the mother by trained psychologists ('testers') in the library or public child care center closest to the child's home. Height and weight on both mother and child were also collected. The Data Appendix provides more details on the sampling strategy, the data collection procedures and the instruments used.

[3] See DNP (2008), Uribe-Mallarino (2008), http://institutodeestudiosurbanos.info/endatos/0200/02-030-vivienda/02.03.01.htm and http://www.dane.gov.co/index.php?option=com_content&view=article&id=354&Itemid=114 for more details.

[4] Currently, 8.1% of all residential blocks in Bogota are classified as E1, 37.3% as E2, 38.2% as E3, 11.3% as E4, 2.1% as E5, and the remaining 2.7% as E6 (DANE 2011).

[5] Higher income families typically live in apartment blocks ("*propiedad horizontal*") or in residential compounds ("conjuntos residenciales"), where access is restricted and heavily controlled by doormen.

[6] Note that we assume that unobservables can be spatially correlated within neighbourhoods but not across and will hence cluster standard errors (SE) at the neighbourhood level throughout the analysis.

[7] The FCI is a short and easy to administer test derived by UNICEF (Frongillo et al. 2013) from the much longer Home Observations for Measurement of the Environment (HOME, Caldwell and Bradley 2001), a well-known measure of the level of home stimuli. Hamadani et al. (2010) validate the FCI against the HOME in a poor environment and assess how well it correlates with the BSID-II.

IIA. Analysis Sample

Data collection took place between March 7 and August 20 of 2011 on a sample of 1,533 children 6 to 42 months in 497 blocks mostly in *estratos* 1 to 3 (including 12 children in E4). The final sample spans geographically all Bogota (see Figure A2 in the Appendix) and is representative of approximately the 85% of the city population that lives in these *estratos*. Table 1 presents the distribution of the sample. For each *estrato* (in columns) and for each age group (in rows), we report the number of household surveys completed and the number of children for whom we administered the Bayley-III test.

Table 1: Sample Sizes Household Survey and Bayley-III by *Estrato* and Age

	Estrato 1		Estrato 2		Estrato 3		Estrato 4		TOTAL	
Age (months)	**Survey**	**Bayley-III**	**Survey**	**Bayley-III**	**Survey**	**Bayley-III**	**Survey**	**Bayley-III**	**Survey**	**Bayley-III**
6 to 14	120	108	115	104	117	100	3	2	355	314
15 to 23	106	87	139	112	158	133	0	0	403	332
24 to 32	112	103	139	125	137	117	7	7	395	352
33 to 41	118	105	133	118	127	107	2	2	380	332
Total Number of Children	456	403	526	459	539	457	12	11	**1533**	**1330**
Children per Block	3.4	3.0	3.3	2.9	2.7	2.3	2.4	2.2	3.1	2.7
Total Number of Blocks	134		159		199		5		**497**	
Sample Loss: Survey to Bayley-III	11.62%		12.74%		15.21%		8.33%		**13.24%**	

A few considerations are in order. Firstly, the largest group is children from E3 and we have very few children from E4. As we discuss in Section AI in the Data Appendix, the reason for this is that we replaced children from E4 that could not be reached with children in E3 (50%), E2 (25%) and E1 (25%). Indeed, soon after the start of field activities, it was clear that households living in E4 were extremely reluctant to participate in the survey, mostly because of apparent mistrust. Secondly, we have fewer children from E1 than E2, although the target number was the same (450 children in each stratum). This is because we experienced a larger number of rejections from E1, possibly also due to higher mistrust amongst those most vulnerable. Finally, note that we had to visit more than double the number of blocks we had originally planned so as to achieve our target sample size.

The Bayley-III test was administered to 1,330 (86.8%) of the children for whom we have a household survey. These children constitute our sample of analysis. Sample loss between the survey and the Bayley-III increases with SES strata: 11.62% amongst E1 households, 12.74% in E2, and 15.21% in E3, the difference being almost statistically significant for E3. In line with this, a Chi-square test of goodness of fit rejects the null that the proportion of children not showing up for the Bayley-III is equally distributed across *estratos* (p-value = 0.044).

If we model the probability of sample loss between the household survey and the Bayley-III test with a probit, we find that having a younger mother, a larger number of children aged 5 to 7 and a lower number of elders (older than 55) living in the household, and attending a childcare center (specially for the younger in the sample), are all factors that increase the probability of a surveyed child not attending the test (see Table 2).[8] These results possibly suggest that mothers without alternative forms of care faced difficulties affording the time to take the child to the test. In addition, interviewer dummies (not reported in the Table) are jointly significant (Chi-square test with 7 degrees of freedom =34.68, p-value =0.000), suggesting that the interviewer played a major role in motivating the mother to take the child to the test—both through her ability in the administration of the household survey and through direct encouragement.

Table 2: Determinants of Sample Loss between the Household Survey and the Bayley-III

	Loss Household Survey to Bayley-III Test =1 (n =1,533)
Estrato 2 =1	0.111
	(0.119)
Estrato 3 =1	0.284*
	(0.115)
Estrato 4 =1	0.190
	(0.442)
Age Child (in months)	0.023
	(0.051)
Care Center Attendance =1	0.554*
	(0.233)
Care Center Attendance =1 * Age	-0.250*
	(0.115)
Diarrhea Last 15 Days=1	0.155
	(0.097)
Age Mother	-0.014+
	(0.007)
Number Children 5 to 7 in Household	0.180*
	(0.073)
Number Elder Older than 55 in Household	-0.148+
	(0.083)
Chi-Sq (7) of Joint Significance of Interviewer Dummies (p-value)	34.68 (0.000)

Notes: +significant at the 10%; *significant at the 5%. SE clustered at the neighbourhood level (primary sampling unit or *sección*) in parantheses.

However, although the interviewer affects sample loss between the household survey and the Bayley-III (as shown), their identity is independent of the child's ability and performance in the test for two main reasons: (i) households are assigned to interviewers randomly, and (ii) the interviewers administer the household survey but not the test. Taking advantage of these factors, we have used the identity of the interviewer to correct for potential selection bias into the test using a Heckman selection model. Results indicate that we cannot reject the null of no selection bias—i.e. selection into the Bayley-III is uncorrelated with unobserved child ability. Not surprisingly, the SES effects we

[8] We also investigated the contribution of other variables but kept this more parsimonious model since these other variables were not significant in explaining sample loss. Results are available upon request.

estimate are virtually the same with and without the selection correction. Nonetheless, as a further robustness test, we will check whether our results are affected by the inclusion of those variables that are significant determinants of sample loss.

Table 3: Mean Sample Characteristics by *Estrato*

		ESTRATO 1 (n = 403)		**ESTRATO 2** (n = 459)		**ESTRATO 3** (n = 457)		**ESTRATO 4** (n =11)	
I	**CHILD CHARACTERISTICS**								
	5 - 18 months of age =1	0.352		0.346		0.324		0.182	
	19 - 30 months of age =1	0.328		0.309		0.398		0.364	
	31 - 42 months of age =1	0.320		0.344		0.278		0.455	
	Female =1	0.514		0.479		0.490		0.364	
	Premature (gestational age < 37 weeks) =1	0.166		0.146		0.151		0.273	
	Birth Weight in gr	3004.1	(538.6)	3045.8	(462.0)	3065.6	(536.6)	2791.0	(756.7)
	Stunted (z-height for age < -2 SD) =1	0.214		0.179		0.138		0.273	
	Firstborn =1	0.471		0.488		0.549		0.455	
II	**PARENTAL CHARACTERISTICS**								
	Age Mother	25.397	(6.485)	26.673	(6.476)	28.425	(6.603)	33.300	(6.325)
	Education Years Mother	9.009	(3.175)	10.298	(3.038)	11.624	(3.069)	14.900	(1.729)
	Mother has more than Secondary Education =1	0.137		0.306		0.433		0.900	
	Mother Works (paid or unpaid) =1	0.430		0.524		0.577		0.700	
	Mother Gave Birth Before Age 18 =1	0.201		0.135		0.074		0.000	
	Education Years Father	8.300	(3.087)	9.657	(3.228)	11.366	(3.426)	14.571	(2.507)
	Father has more than Secondary Education=1	0.071		0.208		0.427		0.714	
	Father Deceased or No Longer Living with Child =1	0.303		0.331		0.324		0.364	
III	**HOUSEHOLD CHARACTERISTICS**								
	Household Size	4.864	(1.695)	4.680	(1.719)	4.460	(1.378)	4.636	(1.027)
	Gradmother Lives in Household =1	0.295		0.309		0.328		0.273	
	Crowding (people per room)*	2.146	(1.204)	1.850	(1.047)	1.581	(1.093)	0.807	(0.122)
	Quality Floors (tiles, carpet, wood)* =1	0.437		0.725		0.891		1.000	
	External Windows* =1	0.849		0.889		0.884		1.000	
	Shared Kitchen* =1	0.218		0.214		0.160		0.091	
	Shared Bathroom* =1	0.280		0.264		0.188		0.000	
	More than One Bathroom* =1	0.102		0.137		0.289		1.000	
	Car* =1	0.042		0.092		0.210		0.909	
	Fridge* =1	0.620		0.756		0.827		1.000	
	Microwave* =1	0.132		0.194		0.333		0.727	
	Washing Machine* =1	0.454		0.542		0.735		1.000	
	Boiler* =1	0.268		0.338		0.530		0.909	
	Computer* =1	0.216		0.368		0.608		1.000	
	Smartphone* =1	0.022		0.072		0.195		0.455	
	Flat TV* =1	0.176		0.198		0.333		0.636	
	Home Theatre* =1	0.055		0.070		0.158		0.364	
	DVD* =1	0.648		0.688		0.814		0.909	
	Stereo* =1	0.536		0.612		0.689		0.727	
	Games Console* =1	0.074		0.122		0.164		0.364	
	Internet* =1	0.186		0.303		0.479		1.000	
IV	**LEVEL HOME STIMULATION**								
	Books and Newpapers (FCI Score)	2.211	(1.918)	2.532	(2.008)	3.116	(2.074)	4.909	(1.514)
	Play Materials (FCI Score)	5.945	(2.479)	6.575	(2.527)	7.221	(2.630)	8.364	(2.873)
	Play Activities (FCI Score)	4.010	(1.732)	4.331	(1.727)	4.740	(1.611)	5.091	(1.578)
V	**CHILD CARE ARRENGMENTS**								
	Child Care Centre Attendance =1	0.323		0.309		0.298		0.545	
	Care Minder =1	0.469		0.556		0.530		0.636	

*Variables used to construct wealth index. Data are means. SD reported in parantheses for continuous variables.

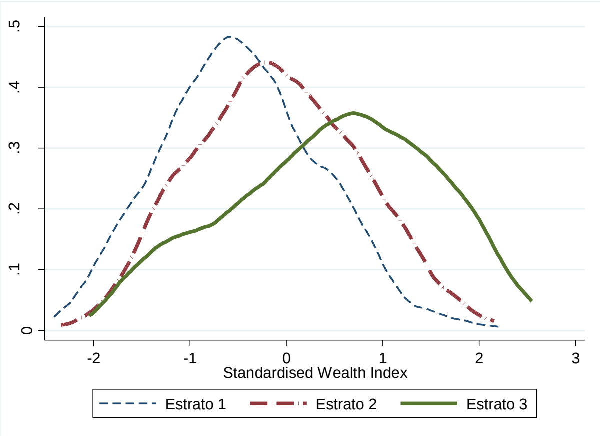
Table 3 reports summary statistics for the final study sample (n = 1,330) by *estrato*. As shown, the sample of children is well balanced across *estrato* by age and gender. On average, 15.49% of the children in the sample are premature (birth before 37 weeks of gestation), 17.63% are stunted (z-score height for age below 2SD deviations of the median of the 2006 World Health Organization International Reference Population, WHO 2006), and 50.38% are firstborns. The mother is missing for 2.25% of the children (3 have passed away and 27 mothers are not living with the child) and the father for 32.85% (11 deceased and 426 absent). As expected, parental education and employment status increase with *estrato* (Panel II), as do the number of assets in the household and the quality of the dwelling (Panel III). Panel IV reports the mean scores for the quality of the home environment variables discussed above (FCI), which are also positively correlated with SES—as measured by *estrato*—as is attendance to a formal child care center or having a child minder.

IIB. <u>Construction of a Household Wealth Index</u>

While it is clear from Table 3 that there is a strong correlation between *estrato* of residence and household economic well-being, there is a substantial amount of heterogeneity within *estratos.* This is illustrated, to an extent, by the ample geographical dispersion of E3 across the city, from more peripheral areas neighboring E2 to very central areas surrounding E4 and E5, with many more services and utilities (see Figure A1 in the Appendix) as well as from the variance in the distribution of observable characteristics within the *estratos* (shown in Table 3 above and in Figure 1 below). Using data on asset ownership and dwelling characteristics collected in the household survey to construct a household wealth index by principal components (Filmer and Scott 2008). We use this index to first assess the amount of heterogeneity within *estratos* and then, to perform the analysis of the SES gap in child development. By using a variable directly related to household economic well-being, as opposed to *estrato* of residence, we obtain as precise a match as possible between SES and child development.

To construct the wealth index, we first identified those variables that had enough variability in the sample and across *estratos* and that are ex-ante reasonable measures of wealth and standard of living, such as crowding (people per room), lack of external windows, shared bathroom, absence of a fridge, ownership of flat TV, etc. All our indicators, except for the crowding index are categorical. Hence, to allow the comparability of the factor loads, we performed a polychoric principal component analysis. Our final wealth index is the first principal component of the following variables: car, fridge, microwave, washing machine, boiler, computer, smartphone, flat TV, home theatre, DVD, stereo, games console, internet, garage, whether the household residence has any external windows, whether the household shares the kitchen with other households, whether the household shares the bathroom with other households, whether the household has more than one bathroom, whether the household has quality floors (tiles, carpet or wood as opposed to gravel, cement or dirt), and the household crowding index. The mean values of these variables by *estrato* are reported in Table 3.[9]

Figure 1 and Table 4: Distribution and Descriptive Statistics of the Household Wealth Index by *Estrato*



	Standardised Wealth Index			
	N	Mean	Median	SD
Estrato 1	403	-0.469	-0.464	0.815
Estrato 2	459	-0.101	-0.138	0.849
Estrato 3	457	0.467	0.561	1.042
Estrato 4	11	1.960	1.960	0.306

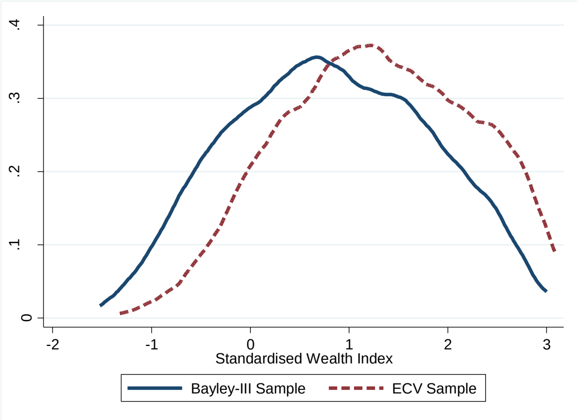
Figure 1 and Table 4 show the distribution of our constructed wealth index by *estrato*. As suspected, there is a remarkable range of variation in household wealth, especially within E3. This indicates that, even if the correlation between household wealth and *estrato* is real—as demonstrated by the fact that the distribution of the wealth index moves to the right the higher the *estrato*—socio-economic block stratification masks a lot of variability within household wealth. The simple correlation coefficient amongst the wealth index and *estrato* is r =0.403.

Based on our wealth index, we can also check the representativeness of our sample, in terms of wealth, relative to the overall population of Bogota. To do so, we use a nationally representative socio-demographic survey collected by the Statistical Agency (DANE), the 2010 *Encuesta de Calidad de Vida* (ECV). Unfortunately, the household level variables available in the ECV that can be used to compute a wealth index are not exactly the same as those available in our data. We therefore estimate a new wealth index using households in the Bogota ECV sample and a set of wealth related variables available in both data sets.[10] We then use the estimated factor loadings to compute a new wealth index both in the ECV and in our sample. Figure 2 plots the density of this index in both samples. As our sample excludes households living in E4 to E6, the density of the wealth index is shifted to the left relative to that in the ECV. However, and perhaps surprisingly, the support of our

[9] The classification of households in wealth quartiles is robust to alternative definitions of the index, including the inclusion of variables other than assets, such as total household income and father's education.
[10] The set of common variables used is: car, fridge, microwave, washer, boiler, computer, DVD, stereo, games console, internet, shared kitchen, shared bathroom, whether there is more than one bathroom in the household, quality floors, and the crowding index.

sample is almost as large as that of the ECV: the largest value of the wealth index in our sample corresponds roughly to the 98th percentile of the ECV.

Figure 2: Distribution of Household Wealth Index in Our Sample (Bayley-III) and in the ECV Sample



IIC. Child Development Outcomes: Use of Externally vs. Internally Standardised Bayley-III Scores

Our analysis is based on the third version of the *Bayley Scales of Infant and Toddler Development* (Bayley-III), which assesses child development by direct observation of child performance. The Data Appendix provides more details on the test and its administration.

The cognitive, receptive language, expressive language, fine motor and gross motor scales are composed of a series of tasks (items) of increasing difficulty that the child has to perform. The child scores 1 for each item correctly performed and 0 otherwise. The test start point is established on the basis of the age of the child (in months and days). To go forward, the child must score 1 on the first three consecutive items given the age-specific start point, or else go back to the previous start point.[11] Once on the test, items are administered until the child scores five consecutive zeros or until the scale is complete. The socio-emotional scale consists of a maximum of 35 5-item response questions responded by the mother or the main caregiver of the child.[12] The number of questions depends on the age of the child.

For each scale, the **raw score** is the sum of all responses. Because items are arranged in increasing level of difficulty, higher raw scores correspond to higher achievement. Raw scores are routinely

[11] We did not adjust for prematurity at the time of the test because of concerns about recall bias in gestational age. Indeed, we observe over 9% mismatches in prematurity reports between the household survey and the Bayley-III test. Results are robust to controlling for prematurity in the analysis.

[12] It also includes a "cannot tell" option for those situations in which the child has not displayed the behaviour the item inquires about.

adjusted by month-of-age and by difficulty level in a non-linear fashion to compute the **composite scores**, which are constructed to have a mean of 100 and a SD of 15 at every month-of-age. The norms (weights) used to construct composite scores are based on the reference population on which the test was standardized; in this case, a representative sample of 1,700 children in the US. Working with externally standardized scores is useful as it allows comparing scores within and across populations, as well as across scales. However, some anomalies in the distribution of the composite scores and its evolution with age, that we describe next, suggest that the external norms used to standardize the Bayley-III may not be appropriate for our sample.

Figure 3: Bayley-III Composite Scores over Age — Non-Parametric Regressions

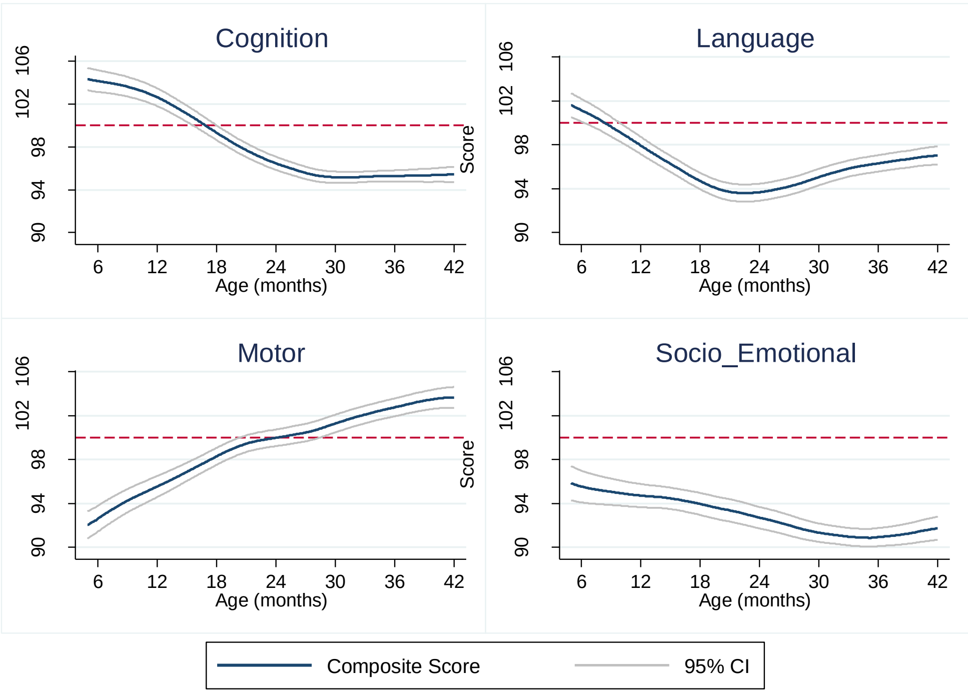


Figure 3 plots average cognitive, language, motor and socio-emotional composite scores against children's age in months estimated in our sample by kernel regressions, along with 95% confidence intervals.[13] Given that composite scores are constructed to have a mean of 100 points at every month-of-age, one would expect the non-parametric regression line to be flat around 100. Instead, we observe a non-linear relationship between the scores and age. The graph in the top left hand side of Figure 3 shows that, while children aged 6-7 months in the sample have a composite score in cognition of over a 100 the score drops by about 10 points, stabilizing at 95 points at 30 months of age. This drop is statistically significant (as the confidence intervals show) and 5 points below the mean of the population of reference. A similar drop, of approximately 8 points—from 102 to 94—is

[13] Note that the composite scores combine the two language (receptive and expressive) and the two motor (fine and gross) sub-scales to produce a single language and motor scale, respectively.

observed in language composite scores for children ages 6 to 22 months. From then onwards, language scores increase again to an average level of 96-97 at around 42 months of age, which continues to be significantly lower than the reference average score.

The two bottom graphs in Figure 3 plot composite scores for the motor and socio-emotional scales. Motor scores increase by 12 points in a relatively linear fashion over the entire age range. Hence, if anything, children in our sample seem to show signs of a delay in motor development with respect to the reference population in the first 18 months of life only, on average. Socio-emotional scores vary non-linearly with age within a 5 point range, from a maximum of 96 at 6 months to a minimum of 91 at 36 months.

To a certain extent, the negative age gradients—i.e. the *presumed* delay in development with respect to the population of reference as the child ages—observed for cognitive and language development and to a lesser extent for socio-emotional development are to be expected given that we are working with a more deprived sample (representative of low- and middle-income groups in Bogota) than the sample for which the test was standardized (representative of all socio-economic groups in the US). However, the 12-point increase in motor development over the age range and the 4-point increase in language scores from 22 to 42 months are unusual. This could imply either that there is a "catch-up" in language development to the standards of reference starting at around 2 years of age and/or, more plausibly, that the norms used for the standardization of the test are not fully applicable to our sample. In other words, that the test *as translated* fails to provide internationally comparable measures of age-appropriate skills for children in low- and middle-income households in Bogota, at least for some age groups.[14] Hence, our analysis will relate to within country comparisons as the comparison with international standards is likely to be tenuous. We discuss this more below showing how we standardize the scores for internally valid comparisons.

In Table 5, we report the means and SDs of the composite scores for each developmental domain by quartile of the household wealth index described above (in columns) and by 12-month age range groups (in rows). Looking at the evolution of the scores by age, we note, for each of the wealth quartiles, patterns similar to those observed in Figure 3 for the entire population. This further suggests the potential inappropriateness of using external norms to our sample. Average cognitive composite scores decrease from the lowest to the middle age group to then stabilize. This trend is common to all wealth quartiles. For language scores, we find that the dip for the middle age group is common across wealth groups, although the increase in language scores between the middle and

[14] We observe a very similar pattern for the composite scores over age of the "wealthiest" in our sample—i.e. households in the top quartile of the household distribution of the wealth index and hence likely to be more similar to the reference population. Results are available upon request.

the oldest age groups is somewhat larger the higher the household SES. While motor development scores increase with age of the child, socio-emotional scores tend to decrease with age, although very mildly. These trends in motor and socio-emotional scores are similar across wealth quartiles. Note also that the SDs reported in Table 5 are significantly smaller than 15 and decreasing with age of the child for all developmental domains, and for cognition in particular. This highlights the need to adjust scores for age in a flexible fashion in the analysis—different than the external standardization norms—so as to allow comparisons across the age distribution. Table 5 also shows that there is substantial variability in the data.

Table 5: Bayley-III Composite Scores by Wealth Quartile and Age Categories

	ALL			WEALTH QUARTILE 1			WEALTH QUARTILE 2			WEALTH QUARTILE 3			WEALTH QUARTILE 4		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
I. Cognitive															
5 - 18 months	451	103.492	9.885	112	102.009	10.790	120	103.458	9.456	108	103.519	9.254	111	105.000	9.886
19 - 30 months	460	95.543	8.180	120	93.667	7.268	107	95.187	8.549	116	94.828	7.800	117	98.504	8.389
31 - 42 months	419	95.334	6.283	101	93.119	6.239	108	93.704	5.895	106	96.226	5.292	104	98.269	6.377
All	1330	98.173	9.119	333	96.306	9.283	335	97.672	9.235	330	98.121	8.521	332	100.602	8.919
II. Language															
5 - 18 months	451	99.022	10.248	112	97.732	10.590	120	98.492	9.724	108	98.324	10.191	111	101.577	10.202
19 - 30 months	460	93.046	10.481	120	91.575	9.358	107	91.224	10.945	116	92.664	10.062	117	96.598	10.810
31 - 42 months	419	96.845	8.832	101	94.079	8.336	108	94.824	8.357	106	97.425	8.378	104	101.038	8.666
All	1330	96.269	10.213	333	94.405	9.822	335	94.988	10.151	330	96.045	9.896	332	99.654	10.201
III. Motor															
5 - 18 months	451	95.244	12.281	112	95.089	12.177	120	93.933	12.707	108	96.176	11.581	111	95.910	12.609
19 - 30 months	458	99.288	10.268	120	98.458	9.586	107	99.449	10.340	115	99.122	11.077	116	100.164	10.106
31 - 42 months	417	103.513	9.445	100	101.880	9.385	108	101.556	8.747	105	105.076	9.429	104	105.538	9.652
All	1326	99.241	11.259	332	98.352	10.792	335	98.152	11.254	328	100.058	11.327	331	100.426	11.523
IV. Socio-Emotional															
5 - 18 months	451	94.956	13.275	112	92.857	13.993	120	95.000	14.595	108	95.694	12.852	111	96.306	11.195
19 - 30 months	460	92.674	13.211	120	91.250	13.367	107	91.449	12.977	116	93.578	12.680	117	94.359	13.670
31 - 42 months	419	91.241	9.820	101	90.000	9.513	108	90.370	10.450	106	91.509	9.516	104	93.077	9.586
All	1330	92.996	12.353	333	91.411	12.573	335	92.373	12.976	330	93.606	11.902	332	94.608	11.724

A common alternative to the use of composite scores, when working with data from developing countries for which external norms may not be appropriate, is to standardize scores internally (Fernald et al. 2011; Schady 2011). The purpose of this method is to construct a score that is valid for within sample comparisons across children of different ages. Often the standardization is done by dividing the sample into the smallest possible age groups (ideally monthly given how sensitive developmental milestones are to age at these very young ages) but guaranteeing enough observations per group, and computing z-scores within age groups (i.e. subtract the months-of-age-specific mean of the raw score and divide by the months-of-age-specific SD).

We followed this approach but computed internal z-scores slightly differently from what is typically done in the literature to allow for more flexibility and to replicate as closely as possible the way in which the Bayley-III constructs the composite scores, while taking into account our limited sample size.[15] Instead of using months-of-age-specific means and SDs, we parameterize the conditional mean and SD as a function of age and estimate it using regression methods. In particular, for each sub-scale *k* = {cognition, receptive language, expressive language, fine motor, gross motor, socio-emotional} of the Bayley-III, we compute the conditional mean using the fitted values of the following regression:

$$\mathrm{Y_i} = \alpha + \beta X_i + \varepsilon_\mathrm{i} \qquad \forall\, i \tag{1}$$

where $\mathrm{Y_i}$ is the raw score of children *i* in each sub-scale *k* and X_i is a polynomial in age of varying order depending on the sub-scale. For the age-conditional SD, we square the residuals of the previous regression, $(\mathrm{Y_i} - \widehat{\beta} X_i)$, and regress them on another flexible age polynomial (D_i) that can, but need not, have the same order as X_i:

$$(\mathrm{Y_i} - \widehat{\beta} X_i)^2 = \gamma + \delta D_i + v_\mathrm{i} \qquad \forall\, i \tag{2}$$

Our estimate of the conditional SD is the square root of the fitted value of the regression in equation (2). Finally, for each sub-scale *k*, we compute the internally age-adjusted z-score, $\mathrm{ZY_i}$, by subtracting from the raw score the within sample age-conditional mean of the score estimated in (1) and dividing by the within sample age-conditional SD obtained from (2). More specifically:

$$\mathrm{ZY_i} = \frac{\mathrm{Y_i} - \widehat{\beta} X_i}{\sqrt{\widehat{\delta} D_i}} \qquad \forall\, i \tag{3}$$

Results are qualitatively similar to internally standardizing scores using the more traditional method described above and to using composite scores. The advantage of the method we follow is that it is less sensitive to outliers and/or a small number of observations within age category. As shown in the Appendix, this procedure worked well and resulted in smooth normally distributed internally standardized scores (Figure A3), with mean zero across the age range (Figure A4).

III. SES Gap by Developmental Domain: Size, Timing and Evolution

In this section, we estimate the SES gap in child development, quantify its size, and establish the ages at which it first becomes statistically significant and how it evolves with age. If we had a very large sample size, we could specify the conditional mean of the outcome of interest as a very flexible function of age and wealth. In particular, for each age, we could compute the mean development at different levels of the household wealth index, possibly for different values of certain control variables, such as gender. Such an approach is not feasible given our sample size. We want, however, to preserve a certain degree of flexibility to avoid the risk that the results we obtain are

[15] To compute composite scores, the Bayley-III standardises raw scores in a non-linear fashion by (i) giving more weight to items in the younger ages, and (ii) lumping months together until the age of 36 months, and in larger intervals thereafter.

affected by the specific functional form assumption we make to the conditional means. To this end, we specify the conditional mean of each sub-scale *k* = {cognition, receptive language, expressive language, fine motor, gross motor, socio-emotional} of the Bayley-III as a cubic spline in the wealth index with two nodes, interacted with a polynomial in age. In addition, we allow for additive gender and testers effects to account for any unobserved differences in the administration or the scoring of the test by testers.[16] In particular, we estimate the following regression using OLS:

$$\mathrm{ZY_i} = \propto_0 + f(age_i, W_i) + \lambda * sex_i + \sum_{t=2}^{6} \zeta_t * tester_i + \varepsilon_\mathrm{i}\,, \quad \forall\, i \qquad (4)$$

where $\mathrm{ZY_i}$ is the age-adjusted z-score, internally standardized as described above, obtained by child *i* on developmental domain *k*, and $f(age_i, W_i)$ is a function of child's age and the household wealth index constructed as follows:

$$f(age_i, W_i) = g_1(age_i) + [1 + g_2(age_i)] * [\alpha_1 W_i + \alpha_2 W_i^2 + \alpha_3 W_i^3 +$$

$$+ \alpha_{41}(W_i - p_1)^3 \boldsymbol{I}(W_i > p_1) + \alpha_{42}(W_i - p_2)^3 \boldsymbol{I}(W_i > p_2)], \qquad \forall\, i \qquad (5)$$

where ***I(.)*** is the indicator function, p_1 and p_2 are, respectively, the 25th and 75th percentile of the wealth index distribution, and $g_1()$ and $g_2()$ are linear polynomials in age (expressed in months). It should be remembered that we internally standardized scores over the entire age range. ε_i is the error term and includes other individual and household characteristics correlated with developmental outcomes. Part of the variability of the error term can be modeled as a function of observables, as we do in section IV below to study the contribution of these factors to the process of acquisition of skills. As noted earlier, we cluster standard errors at the neighborhood level (primary sampling unit, n = 128) to allow for spatial correlation within neighborhoods, but assuming unobservables are not correlated across neighborhoods.

The expression in equation (5) is a cubic spline in wealth with two nodes interacted with a polynomial in age. Whilst the value of the individual parameters might be difficult to interpret, this specification fits the data well while offering a reasonable degree of smoothing, allowing us to characterize the patterns in the data. In what follows, we will compute the SES gap evaluating the conditional mean function on the right hand side of equation (4) at different ages and different levels of the wealth index. We have experimented with different specifications of the conditional means in equation (4), including higher order age polynomials in $g_1()$ and $g_2()$, and our results are robust to the details of the implementation.

[16] Note that testers were distributed randomly across *estratos* and age categories.

In Tables 6 and 7, we report the differences between the conditional mean z-score in sub-scale k on the left hand side of equation (4) evaluated at the 90th and 10th percentiles of the household wealth index distribution, and at the 80th and 20th percentile, respectively. These differences are evaluated at various ages. As the scores are internally standardized, the figures in the Tables can be interpreted in terms of standard deviations within our sample. Each column refers to a different outcome, or developmental domain, as measured by the different sub-scales in the Bayley-III. The standard errors are computed by bootstrapping and clustering at the neighborhood level.

Table 6: Estimated Gap in Child Development between the 90th and 10th Percentile of the Wealth Index Distribution by Age

Age (months)	Cognition	Receptive Language	Expressive Language	Fine Motor	Gross Motor	Socio - Emotional
6	0.144	0.144	0.284	-0.014	-0.333+	0.002
	(0.241)	(0.172)	(0.240)	(0.203)	(0.195)	(0.198)
8	0.192	0.181	0.307	0.016	-0.295	0.038
	(0.221)	(0.156)	(0.220)	(0.186)	(0.181)	(0.182)
10	0.239	0.217	0.330	0.046	-0.256	0.074
	(0.201)	(0.142)	(0.201)	(0.170)	(0.169)	(0.167)
12	0.287	0.253*	0.352+	0.076	-0.218	0.109
	(0.182)	(0.128)	(0.183)	(0.154)	(0.157)	(0.152)
14	0.334*	0.290*	0.375*	0.106	-0.180	0.145
	(0.164)	(0.115)	(0.165)	(0.140)	(0.145)	(0.137)
18	0.429**	0.363**	0.421**	0.166	-0.103	0.216+
	(0.130)	(0.094)	(0.134)	(0.115)	(0.127)	(0.113)
22	0.525**	0.435**	0.466**	0.226*	-0.026	0.288**
	(0.106)	(0.084)	(0.109)	(0.100)	(0.116)	(0.096)
26	0.620**	0.508**	0.512**	0.286**	0.050	0.359**
	(0.096)	(0.088)	(0.098)	(0.100)	(0.115)	(0.091)
30	0.715**	0.581**	0.557**	0.346**	0.127	0.430**
	(0.106)	(0.104)	(0.105)	(0.114)	(0.123)	(0.100)
34	0.810**	0.653**	0.602**	0.406**	0.203	0.502**
	(0.131)	(0.129)	(0.127)	(0.138)	(0.139)	(0.120)
38	0.905**	0.726**	0.648**	0.467**	0.280+	0.573**
	(0.164)	(0.157)	(0.157)	(0.168)	(0.161)	(0.147)
42	1.000**	0.799**	0.693**	0.527*	0.356+	0.644**
	(0.201)	(0.188)	(0.192)	(0.201)	(0.186)	(0.177)

Notes: +significant at 10%, *significant at 5%, **significant at 1%. For each age, we report the difference in the predicted values at the 90th and 10th percentile of the household wealth distribution from a regression that fits a cubic spline in wealth with two nodes interacted with age linearly, and controlling for child's sex and tester dummies. SE are clustered at the neighbourhood level (primary sampling unit or sección). Bootstrapped SE of the difference using 500 replications are reported in parantheses.

In Table 6, we notice that the gap in cognitive development between children in households in the 90th and 10th percentiles of the wealth index distribution increases monotonically with children's age, becoming significantly different from zero at the 5% at 14 months. At this age, the difference between a child in the 90th percentile and one in the 10th percentile of the wealth distribution is about a third of a SD of a cognitive development z-score. This difference increases to 0.72 of a SD at 30 months and is as high as one SD at 42 months.

A similar pattern emerges for receptive and expressive language, except that the differences are statistically significant for younger children, at 12 months (at the 5% for receptive and at the 10% for expressive language). The magnitude of the gap is slightly smaller for receptive language, measuring 0.29 of a SD at 14 months and reaching 0.80 of a SD at 42 months. For expressive language, we observe a larger gap, similar to that in cognition for children around 18 months. However, the gap increases at a slower pace with age, reaching 0.69 SD by 42 months.

There are no large differences in gross motor skills between the 10th and 90th percentiles of the wealth distribution. For fine motor skills, which are known to be linked to cognition, we find sizeable gaps that increase with age and that become statistically significant at 22 months. The size of the gap reaches over half of a standard deviation at 42 months.

Finally, for the socio-emotional scale, we find again that the gap is increasing with age and that it becomes statistically significant at the 10% at 18 months and at the 5% at 22 months. The size of the gap is smaller than the gap in cognition or language, peaking at 0.64 at 42 months.

We have formally tested whether the gap increases with age by comparing the size of the gap at 42 months with that at 6 months and find that the gap significantly widens as children get older for cognitive development (estimated mean gap difference =0.87, 95% confidence interval (CI) =(0.13, 1.64)), receptive language (mean =0.67, 95% CI =(0.04, 1.31), motor development (mean =0.66, 95% CI =(0.02, 1.28) and socio-emotional development (mean =0.63, 95% CI =(0.01, 1.29). The standard errors for these tests are computed by simulating the distribution of the difference in the gap using 1000 random draws of the sample (sampling with replacement) and clustering SE at the neighborhood level.

Not surprisingly, Table 7 shows that the differences between the 80th and 20th percentiles are a scaled down version of the differences between the 90th and 10th percentiles reported in Table 6. In the case of cognition, they become statistically significant at 18 months, when they measure 0.19 of a SD, and then increase up to 0.67 of a SD by age 42 months. The pattern is substantially similar for receptive language, where, however, at 42 months the gap measures 0.54 of a SD. Interestingly, in the case of expressive language, the gaps are larger at younger ages and are statistically different from zero at the 5% at age 14 months, when it measures 0.25 of a SD. Interestingly, in the case of gross motor skills, at very early ages, the gap between children in the 80th and 20th wealth percentiles is actually negative (in that poorer children seem to have better gross motor skills) and statistically different from zero. However, by 42 months of age, the gap is positive and equal to 0.31. The pattern for the gaps in fine motor and socio-emotional skills follows that shown in Table 6.

For children between the 80th and the 20th wealth percentiles, we find that the only gaps that significantly increase with age are that in cognition (mean =0.73, 95% CI =(0.23, 1.24)) and in motor development (mean =0.64, 95% CI =(0.21, 1.03)).

Table 7: Estimated Gap in Child Development between the 80th and 20th Percentile of the Wealth Index Distribution by Age

Age (months)	Cognition	Receptive Language	Expressive Language	Fine Motor	Gross Motor	Socio - Emotional
6	-0.052	0.030	0.175	0.036	-0.350*	0.127
	(0.154)	(0.153)	(0.174)	(0.143)	(0.136)	(0.150)
8	-0.012	0.059	0.195	0.051	-0.312*	0.132
	(0.141)	(0.140)	(0.158)	(0.131)	(0.127)	(0.139)
10	0.029	0.087	0.215	0.065	-0.276*	0.137
	(0.128)	(0.127)	(0.143)	(0.119)	(0.118)	(0.128)
12	0.069	0.116	0.234+	0.080	-0.240*	0.142
	(0.117)	(0.115)	(0.128)	(0.107)	(0.111)	(0.117)
14	0.109	0.144	0.254*	0.094	-0.203*	0.147
	(0.106)	(0.104)	(0.114)	(0.097)	(0.103)	(0.107)
18	0.190**	0.201*	0.294**	0.124	-0.130	0.157+
	(0.087)	(0.085)	(0.089)	(0.078)	(0.092)	(0.091)
22	0.270**	0.257**	0.334**	0.153*	-0.057	0.166*
	(0.076)	(0.074)	(0.074)	(0.066)	(0.086)	(0.080)
26	0.351**	0.314**	0.373**	0.182**	0.015	0.176*
	(0.077)	(0.075)	(0.074)	(0.065)	(0.086)	(0.078)
30	0.432**	0.371**	0.413**	0.211**	0.088	0.186*
	(0.087)	(0.087)	(0.090)	(0.075)	(0.092)	(0.085)
34	0.521**	0.427**	0.453**	0.240*	0.161	0.196*
	(0.106)	(0.107)	(0.114)	(0.093)	(0.103)	(0.099)
38	0.593**	0.484**	0.492**	0.269*	0.234*	0.206+
	(0.129)	(0.131)	(0.144)	(0.115)	(0.118)	(0.118)
42	0.673**	0.541**	0.532**	0.298*	0.307**	0.215
	(0.154)	(0.157)	(0.175)	(0.139)	(0.136)	(0.139)

Notes: +significant at 10%, *significant at 5%, **significant at 1%. For each age, we report the difference in the predicted values at the 80th and 20th percentiles of the household wealth index distribution from a regression that fits a cubic spline in wealth with two nodes interacted with age linearly, and controlling for child's sex and tester dummies. SE are clustered at the neighbourhood level (primary sampling unit or sección). SE of the difference using 500 replications are reported in parantheses.

The findings in Tables 6 and 7 are robust. We find gaps of similar sizes if we estimate a modified version of equation (4) where, instead of the cubic spline, we include dummies for the quartile of the wealth distribution the household child *i* lives in belongs to. Estimating this variant of equation (4) separately by 12-month age range categories (6 to 18, 19 to 30, and 31 to 42 months) also offers a comparable picture regarding the evolution of the gap with age and its timing. Similarly, we observe a comparable pattern if we use the level of maternal education dummies instead of wealth

Figure 4: Internally Standardized Scores by Wealth Percentile and Age &

Gap between the 90th and the 10th Percentiles by Age, by Developmental Domain

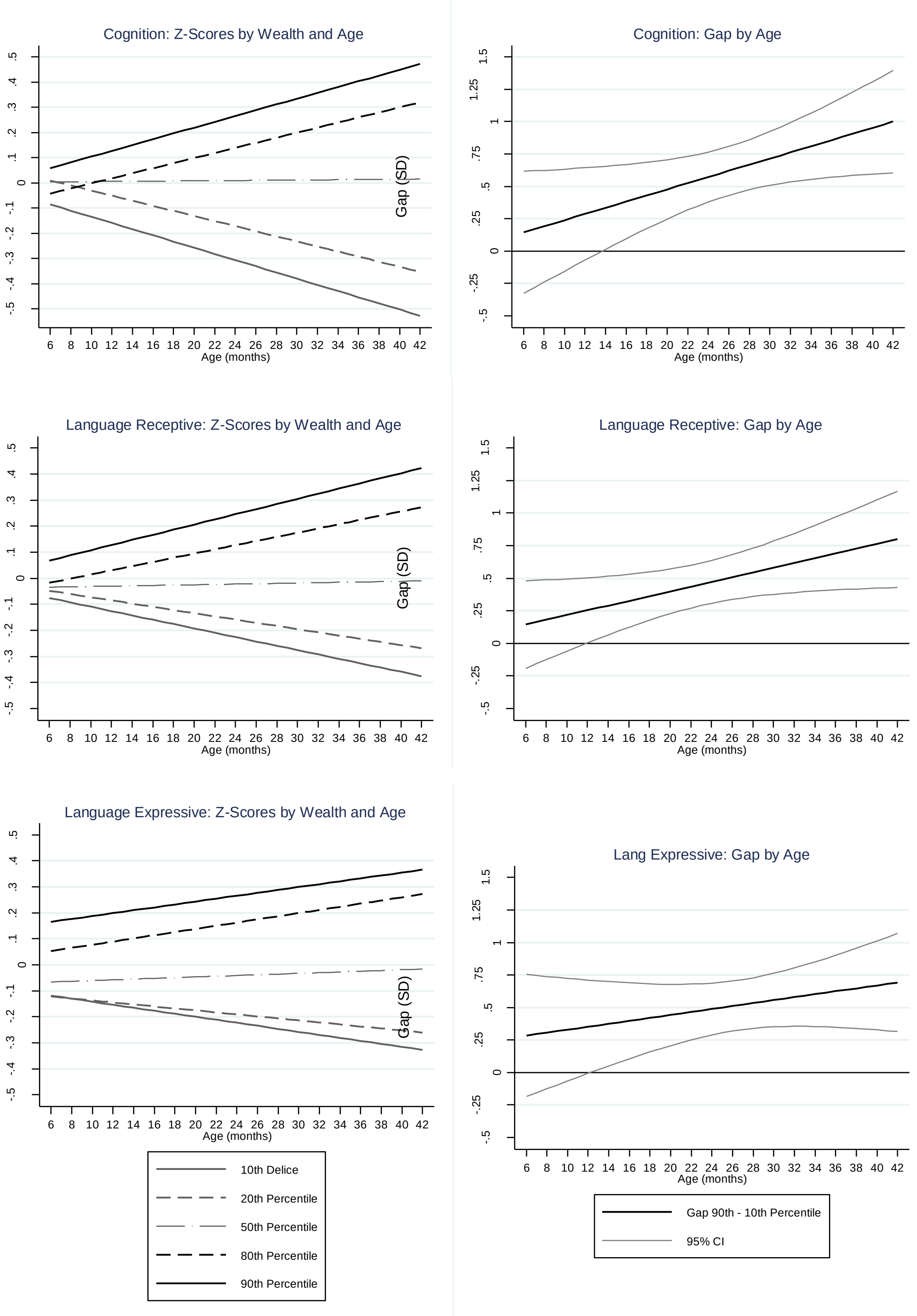
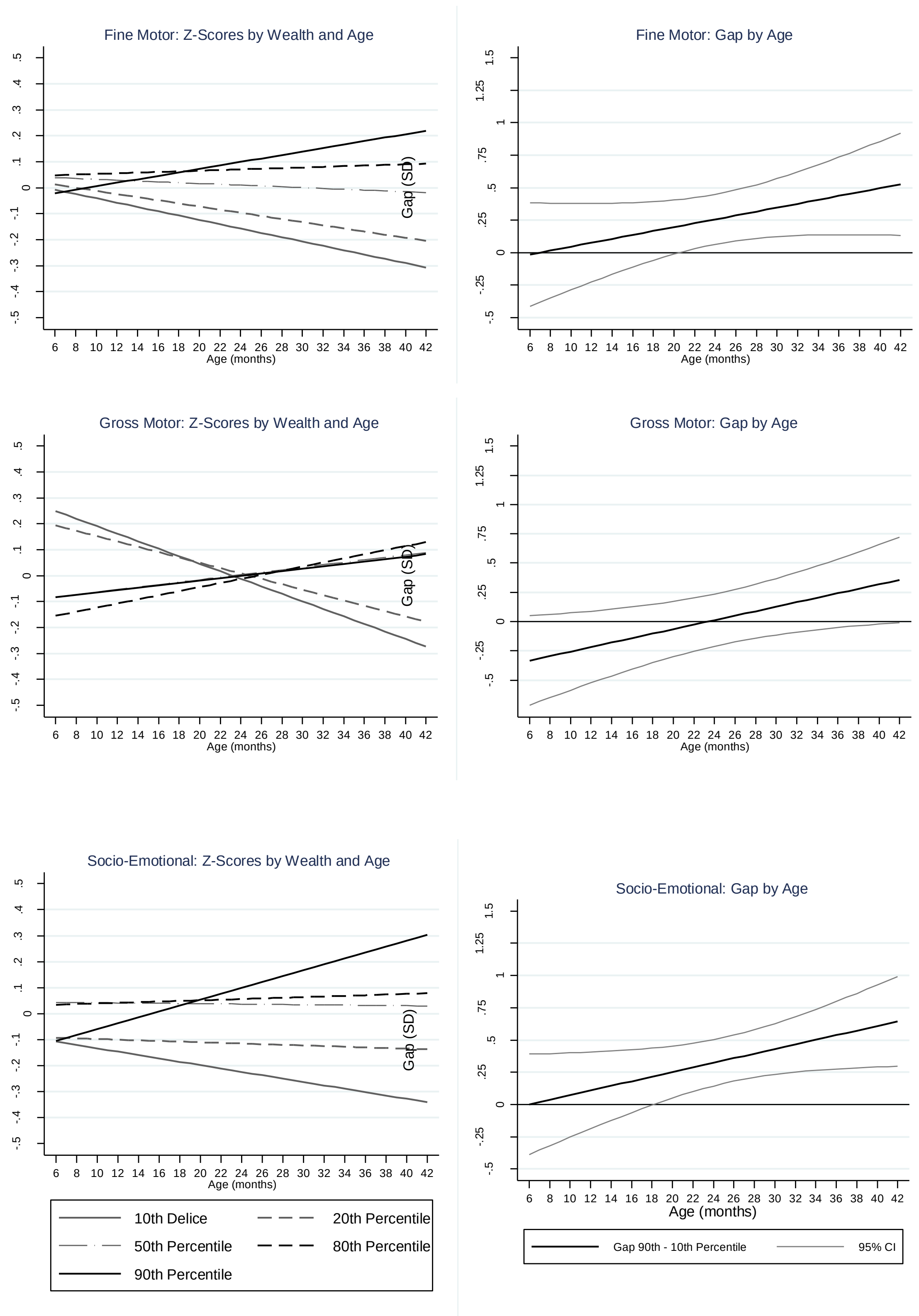


Figure 4 (continued): Internally Standardized Scores by Wealth Percentile and Age & Gap between the 90th and the 10th Percentiles by Age, by Developmental Domain



quartiles, and compare children whose mothers have completed secondary education or more with children whose mothers have at most primary education completed.

In addition, our findings are robust to: (i) working with composite scores and with raw scores internally standardized following the traditional approach (i.e. computing z-scores using within sample means and SDs); (ii) including those variables that were significant determinants of sample loss between the household survey and the Bayley-III test as additional covariates—namely, age of the mother, number of children 5 to 7 and number of elders (older than 55) living in the household, and child's attendance to a child care center (see Section IIB above); (iii) controlling for prematurity; and (iv) excluding children extremely advanced or extremely delayed in each domain—i.e. children 2.5 SD above or below the sample mean.[17] These results are available upon request.

Figure 4 effectively summarizes the above results. In particular, in the top left panel of Figure 4, we plot the conditional mean of the internally standardized cognitive score for the 10th, 20th, 50th, 80th and 90th percentiles of the household wealth index against age, as predicted by the right hand side of equation (4). In the top right-hand side panel, we plot the differences between the 90th and 10th percentile with confidence intervals. This plot corresponds to the estimated SES gap sizes reported in Table 6. The following panels are analogous to the top two except that they plot the corresponding figures for receptive and expressive language, fine and gross motor, and socio-emotional development, respectively.[18]

In addition to the analysis we have just presented, we also investigated whether the SES evolves differently for boys than girls, and found no major gender differences. Nonetheless, the gap in cognition, language expressive and fine motor skills becomes statistically significant at slightly earlier ages for girls. In addition, the overall observed gap in socio-emotional development is concentrated amongst boys mostly. Further details are available upon request.

IV. <u>Association of Other Factors with the SES Gap</u>

Until now, we have quantified the SES gap on child development controlling for tester effects, gender and children's age in months. We now consider a series of potential mediating factors, by

---

[17] This amounts to dropping 4 children in the language sub-scale, 9 children in the motor sub-scale and 5 children in the socio-emotional sub-scale.

[18] Plots of the differences between the 80th and 20th percentiles are available upon request. Overlapping the differences between the 90th and 10th percentiles and those between the 80th and 20th percentiles in one same graph, it is apparent that they are roughly parallel, with the exception of the socio-emotional development and (possibly) the fine motor skills.

adding different sets of regressors to the benchmark specification in equation (4). These include: parental factors, child biomedical factors, factors related to the level of stimulation in the home (home environment), and factors related to institutional child care arrangements. We are interested in understanding both the contribution of these factors to the Bayley-III scores and the extent to which they are associated with the SES gap and hence reduce the effect of wealth that we have estimated in the previous section. Of course this analysis is not causal but documents the observed correlations in the data.

Table 8 presents results for the cognitive scale. We use this Table to illustrate the procedure we follow and discuss our main findings. Next, in Table 9 we present results for the remaining sub-scales, except for gross motor, as the SES gap in this area only becomes significant at 38 months. We do not report the estimates of the coefficients on the control variables (factors) in Table 8 or Table 9 but report them in Table A2 in the Appendix, although only for the most complete model. The coefficients on the regressors in the more parsimonious specifications are available upon request.

Each column in Table 8 reports results obtained with a different set of regressors. The first column, labeled 'Model 1', reports the estimated SES gap between children in the 90th and the 10th percentile of the household wealth distribution using our benchmark specification at 4 points in the distribution of age (in rows): 12, 18, 30, and 42 months. In other words, we reproduce the estimates reported in Table 6 for these ages. 'Model 2' reports the same gap after we add to the regression equation parental factors. These include: two dummies for maternal education (equal to one if the mother has some secondary education and if she has more than secondary education completed—the omitted category being primary education completed or less); a dummy for first child; and two dummies for whether the mother and the father are no longer living in the household, either because they have left the household or because they are deceased. For absent mothers, we have replaced missing education levels with the median value of the variable in the quartile of the wealth distribution the household belongs to and have accounted for the replacement with a dummy.

Maternal education has a strong association with child development in particular for receptive language (see Table A2). On average, children of mothers with at least secondary education completed score 0.17 of a SD higher on cognition, 0.31 of a SD higher on receptive language, and 0.21 of a SD higher on fine motor development. Maternal absence is strongly and negatively correlated with all dimensions of child development, while birth order is significantly correlated with language development. These variables jointly contribute to explaining child development as the p-value of the F-test of joint significance reported in Table A2 in the Appendix shows (p-value = [0.00 – 0.08] depending on the outcome). Maternal years of education are highly correlated with variables related to the quality of the home environment (r = [0.31 - 0.39] depending on the outcome, p < 0.000): if the later variables are not included in the regression, the association between a ‘highly’ educated mother (secondary education completed or higher) and child development increases to 0.35 SD for cognition, 0.48 SD for receptive language, 0.29 SD for expressive language and 0.32 SD for fine motor development (results available upon request).

Table 8: Contribution of Other Factors to the SES Gap in Cognitive Development

	Model 1	Model 2	p-value (Gap Model 1 = Gap Model 2)[1]	Model 3	Model 4	p-value (Gap Model 3 = Gap Model 4)[1]	Model 5
Age							
12 months	0.287+	0.184	0.013	0.176	0.101	0.038	0.111
	(0.182)	(0.181)		(0.179)	(0.182)		(0.182)
18 months	0.429**	0.321*	0.005	0.310*	0.224+	0.011	0.220+
	(0.130)	(0.132)		(0.130)	(0.132)		(0.132)
30 months	0.715**	0.596**	0.004	0.578**	0.470**	0.003	0.438**
	(0.106)	(0.113)		(0.114)	(0.114)		(0.114)
42 months	1.000**	0.871**	0.017	0.846**	0.715**	0.006	0.656**
	(0.201)	(0.206)		(0.213)	(0.210)		(0.211)
Parental Factors							
maternal education, absent mother, absent father, first child	No	Yes		Yes	Yes		Yes
Biomedical Factors							
prematurity, prematurity*age, birthweight, stunting	No	No		Yes	Yes		Yes
Home Environment Factors							
books, newspapers and magazines for adults, number of play materials, number of play activities	No	No		No	Yes		Yes
Institutional Care Factors							
attendance to public child care center, private center, or "hogar comunitario"	No	No		No	No		Yes

Notes: +significant at 10%, *significant at 5%, **significant at 1%. For each age, Model 1 reports the gap in scores (in SD) between the 90th and 10th percentile of the household wealth distribution as the difference in the predicted values from a regression that fits a cubic spline in wealth with two nodes interacted with age linearly, and controlling for child's sex and tester dummies. Subsequent models control for additional factors as indicated in the table. SE are clustered at the neighbourhood level (primary sampling unit or sección). Bootstrapped SE of the difference using 500 replications are reported in parantheses.

[1]p-values >0.10 not reported. p-values of the difference in the gaps between Models constructed using bootstrapping methods (500 replications).

Paternal education is not statistically significant if maternal education is included, which is probably due to the fact that both variables are highly correlated (r = 0.51, p <0.001) since partners often match within similar education levels and environments and since education is also strongly correlated with wealth. We also examined the inclusion of maternal age, a dummy for having given birth before 18 (13.6%), the interaction between being a young mother (birth before 18) with no grandmother in the house, and a dummy for maternal employment (whether paid or unpaid).[19] Similarly, we considered the effect of maternal anthropometrics (height, weight, body mass index and dummies for overweight and obesity) on child outcomes. For parsimony, we excluded these variables from the analysis since they were not significantly associated with child development.

[19] We did not consider paternal employment since 98% of the fathers in the sample are employed.

As shown in Table 8, adding parental characteristics 'reduces' the estimated SES gap, perhaps not surprisingly, given the association between education and wealth. The estimated gap in cognitive development between the 90^{th} and 10^{th} percentile is reduced from 0.43 to 0.32 of a SD at 18 months, and from 1 to 0.87 SDs at 42 months. As indicated in the third column, labeled 'p-value(Gap Model 1 = Gap Model 2)', the change in the estimated gap after controlling for parental background variables is statistically different from zero at all ages considered.[20]

In 'Model 3', we include a set of variables related to the child biomedical and nutritional status: a dummy for prematurity (birth before 37 weeks of gestation), the interaction between prematurity and age, birth weight (in Kg), and a dummy that equals one if the child is stunted (height-for-age <-2SD of the median of the 2006 World Health Organization International Reference Population, WHO 2006).[21] Birth weight was missing for 8.38% of the children in the sample. We imputed missing values with the predicted value from a regression of birth weight on sex, gestational age and height-for-age, and have accounted for the replacement with a dummy.

Prematurity is negatively correlated with cognitive development though the effect is largely mitigated by age, and disappears once other factors are controlled for (Table A2). Similarly, birth weight is positively associated with cognitive and fine motor development. Stunting is significantly associated with lower receptive language development at the 5% (0.11 SD) and with lower fine motor development at the 10% (0.18 SD). While we are not able to identify a significant association between low height-for-age and cognitive development in our data, we do observe a positive and significant, albeit small, raw correlation between height-for-age and cognitive z-scores (r =0.05, p <0.06), receptive and expressive language (r =0.08, p <0.00), fine motor (r =0.06, p <0.02) and gross motor skills (r =0.1, p <0.00). This is consistent with the findings in Fernald et al. (2006), who report that the association identified between height-for-age z-scores and the BSID-II mental development index amongst poor children in semi-urban Mexico is no longer significant once household related variables are controlled for.[22] As the p-values of the F-tests of joint significance for the various dependent variables reported in Table A2 show, when considered jointly, all biomedical factors contribute to explaining child development (except socio-emotional development).

[20] We have used bootstrapping methods to construct this test in order to take into account the fact that the coefficients are estimated on the same sample.

[21] Whereas 16.85% of the children in the sample are stunted, only 2.78% are underweight and 0.75% wasted.

[22] Note however that most studies have found a relationship with stunting even after controlling for socio-economic background variables (Grantham-McGregor et al. 2007; Walker et al. 2007).

After including these biomedical factors, the SES gap in cognitive development, as measured by the difference between the 90th and 10th percentile, is marginally reduced relative to the figure from 'Model 2' for all age groups. However, this reduction is not statistically significant.

In a fourth step, in 'Model 4', we add variables related to the quality of the home environment: the FCI scores for 'books, newspapers and magazines for adults', 'play activities' and 'play materials'.[23] As shown in Table A2, both play activities and play materials have a strong association with child development, the coefficient of play activities (the number of quality interactions between the child and an adult) being larger for language and socio-emotional development and the coefficient of play materials being larger for cognitive and fine motor development. The score for books for adults and newspapers in the home is never statistically significant.

As shown in the sixth column of Table 8, labeled 'p-value(Gap Model 3 = Gap Model 4)' the SES gap in cognitive development is significantly reduced when we consider these additional variables. For instance, at 42 months, it goes from 0.85 in Model 3 to 0.72 in Model 4, with the difference being statistically significant at 1%. Analogously, at 30 months, the estimated gap is reduced from 0.58 to 0.47.

Finally, in 'Model 5', we include three dummies for public, private and *hogares communitarios* (small pseudo-nurseries run by community women) preschool attendance. These variables are jointly significantly correlated with better cognitive, receptive and fine motor development but do not explain expressive language nor socio-emotional development, as shown by the p-values of the F-tests of joint significance reported in the final row of Table A2 (p-values =[<0.00 - 0.01] depending on the outcome),

This final column in Table 8, 'Model 5' shows that the SES gap in cognitive development is further reduced relative to that reported for 'Model 4' after controlling for factors related to the type of preschool attended, albeit not significantly so.

Table 9 replicates the exercise in Table 8 for the other sub-scales of the Bayley-III. In the top two panels, we report the analysis for receptive and expressive language; and that for fine motor skills and socio-emotional development in the bottom two panels of the Table. As noted, we do not report this analysis for gross motor development as we do not observe large gaps in this dimension (results available upon request).

[23] We examined several interactions between age and the FCI scores but they were not statistically significant. Section AII in the Data Appendix provides more details on the construction of these variables.

Table 9: Contribution of Other Factors to the SES Gap in Receptive and Expressive Language, Fine Motor and Socio-Emotional Development

	Mod 1	Mod 2	p-value (Gap Mod1 = Gap Mod2)[1]	Mod 3	p-value (Gap Mod2 = Gap Mod3)[1]	Mod 4	p-value (Gap Mod4 = Gap Mod3)[1]	Mod 5	Mod 1	Mod 2	p-value (Gap Mod1 = Gap Mod2)[1]	Mod 3	Mod 4	p-value (Gap Mod4 = Gap Mod3)[1]	Mod 5
	RECEPTIVE LANGUAGE								EXPRESSIVE LANGUAGE						
Age															
12 months	0.253*	0.109	0.003	0.115		0.032	0.024	0.044	0.352+	0.260	0.053	0.267	0.174	0.008	0.183
	(0.128)	(0.128)		(0.130)		(0.130)		(0.128)	(0.183)	(0.179)		(0.177)	(0.176)		(0.177)
18 months	0.363**	0.213*	0.000	0.211*		0.115	0.004	0.113	0.421**	0.320*	0.022	0.318*	0.214+	0.001	0.210
	(0.094)	(0.098)		(0.099)		(0.101)		(0.099)	(0.134)	(0.132)		(0.128)	(0.129)		(0.129)
30 months	0.581**	0.421**	0.001	0.402**		0.279*	0.000	0.249*	0.557**	0.439**	0.006	0.421**	0.293**	0.000	0.264*
	(0.104)	(0.112)		(0.112)		(0.113)		(0.112)	(0.105)	(0.110)		(0.107)	(0.110)		(0.110)
42 months	0.799**	0.630**	0.006	0.593**		0.443*	0.003	0.386*	0.693**	0.559**	0.009	0.524**	0.372+	0.000	0.318
	(0.188)	(0.194)		(0.196)		(0.192)		(0.192)	(0.192)	(0.198)		(0.198)	(0.200)		(0.201)
	FINE MOTOR								SOCIO-EMOTIONAL						
Age															
12 months	0.076	-0.007	0.078	-0.025		-0.070		-0.059	0.109	0.077		0.077	0.034		0.034
	(0.154)	(0.168)		(0.168)		(0.166)		(0.165)	(0.152)	(0.159)		(0.160)	(0.161)		(0.161)
18 months	0.166	0.083	0.056	0.059		0.009		0.012	0.216+	0.178		0.179	0.125	0.084	0.125
	(0.115)	(0.132)		(0.131)		(0.133)		(0.132)	(0.113)	(0.122)		(0.122)	(0.126)		(0.126)
30 months	0.346**	0.262*	0.045	0.228+	0.074	0.166		0.153	0.430**	0.380**		0.381**	0.307*	0.041	0.307*
	(0.114)	(0.128)		(0.125)		(0.136)		(0.138)	(0.100)	(0.110)		(0.111)	(0.119)		(0.120)
42 months	0.527**	0.442*	0.076	0.396+		0.323		0.295	0.644**	0.582**		0.584**	0.489*	0.051	0.490*
	(0.201)	(0.206)		(0.204)		(0.215)		(0.219)	(0.177)	(0.184)		(0.187)	(0.192)		(0.193)
Parental Factors															
maternal education, absent mother, absent father, first child	No	Yes		Yes		Yes		Yes	No	Yes		Yes	Yes		Yes
Biomedical Factors															
prematurity, prematurity*age, birthweight, stunting	No	No		Yes		Yes		Yes	No	No		Yes	Yes		Yes
Home Environment Factors															
books, newspapers and magazines for adults, number of play materials, number of play activities	No	No		No		Yes		Yes	No	No		No	Yes		Yes
Institutional Care Factors															
attendance to public child care center, private center, or "hogar comunitario"	No	No		No		No		Yes	No	No		No	No		Yes

Notes: +significant at 10%, *significant at 5%, **significant at 1%. For each age, Model 1 reports the gap in scores (in SD) between the 90th and 10th percentile of the household wealth distribution as the difference in the predicted values from a regression that fits a cubic spline in wealth with two nodes interacted with age linearly, and controlling for child's sex and tester dummies. Subsequent models control for additional factors as indicated in the table. SE are clustered at the neighbourhood level (primary sampling unit or sección). Bootstrapped SE of the difference using 500 replications are reported in parantheses.

[1]p-values >0.10 not reported. p-values of the difference in the gaps between Models constructed using bootstrapping methods (500 replications).

The size of the SES gap in receptive and expressive language development, as measured by the difference between the 90th and 10th percentile, follows a very similar pattern to that in cognition after controlling for parental, biomedical, home environment and institutional factors. Indeed, while the SES gap continues to persist, it is reduced after the inclusion of each set of variables. However, this reduction is only statistically significant after the inclusion of parental background variables (as shown in columns labeled 'p-value(Gap Model 1 = Gap Model 2)') and variables related to the level of stimulation in the home environment (see columns labeled 'p-value(Gap Model 3 = Gap Model 4)'). The SES gap in fine motor and socio-emotional development is also reduced with the inclusion of these factors, although by a smaller amount. Moreover, these differences in the size of the gap are occasionally significant from each other statistically speaking.

It is worth noting that even after the inclusion of all these variables, the SES gap in cognition remains statistically significant and substantially large, at 0.44 of a SD by 30 months and 0.66 of a SD by 42 months for cognition and at 0.25 SD by 30 months and 0.39 SD by 42 months for receptive language. For expressive language, while the gap remains significant by 30 months at 0.26 SD, the difference by 42 months, at 0.32 SD, is less precisely estimated and hence not significant. Notably, the gap in fine motor development disappears after the inclusion of all these factors. On the other hand, the gap in socio-emotional development remains quite large, at 0.31 SD at 30 months and 0.49 SD at 42 months, being the developmental dimension less affected by parental, biomedical, environmental and institutional factors.

V. Discussion and Conclusion

In this paper we examine the developmental profile by age, household wealth and developmental domain on a large sample of children 6 to 42 months of age from low- and middle-income households in the city of Bogota. We assess child development using the *Bayley Scales of Infant and Toddler Development*. We quantify the size, timing and evolution of the SES gap by developmental area, and additionally study the contribution of other factors to this gap.

We find evidence of a sizeable and statistically significant SES gap in child development since very early ages. The average gap between children in households in the 90th and the 10th percentile of the wealth distribution in our sample is significant starting at 14 months for cognitive development (gap size of 0.33 of a SD), and at 12 months for receptive and expressive language (gap sizes of 0.25 and 0.35 of a SD), respectively. Moreover, and in line with previous studies in the literature—most on older children—the gap increases substantially, monotonically and significantly with age, to levels of 1.00 SD for cognition, of 0.80 SD for receptive language and of 0.69 SD for expressive language by

42 months. The gap in fine motor development becomes statistically significant at 22 months, reaching a maximum of 0.53 SD by 42 months. Similarly, the gap in socio-emotional development becomes statistically significant at the 5% at 22 months, peaking at 0.64 SD by 42 months. There are no large differences in gross motor skills between children of different SES backgrounds.

Another important finding is that the estimated SES gap in cognitive, receptive language, expressive language and fine motor development for the older children in the sample persists after controlling for other factors, including maternal education. However, it is generally significantly reduced after the inclusion of parental factors and factors related to the quality of the home environment in the estimation. On the other hand, the contribution of biomedical factors and institutional care to child development is marginal.

Parental and biomedical factors (excluding stunting) could be considered as initial endowments, whereas the level of stimulation in the home and the type of institutional care could be thought of as choice variables related to parental investments in their children (i.e. the quantity and quality of child care provided), and hence possibly modifiable through public policy. It should be stressed, however, that the evidence on the mediating factors we report in Section IV is not causal. Parental investments are determined jointly with child development and may react to its evolution, either in a ‘compensatory’ (parents invest more in children with lower perceived ability to close gaps) or a ‘complementary’ (parents devote more resources to children with higher perceived ability because of the highest expected returns) fashion. In addition, it is possible that the associations reported are to a certain extent driven by unobserved variables that influence both development and the contributing factors analyzed, such as parental cognitive abilities (proxied partly with parental educational attainment).

Despite these caveats, our findings are important for stimulating research in understanding further the determinants of child development and the intergenerational transmission of poverty. Moreover, they may contribute to guiding the eventual development and timing of supportive ECD interventions that can contribute to reducing unequal opportunities earlier in life and income disparities in the longer run.

References

Almond, D. and J. Currie. 2010. "Human Capital Development Before Age Five." *Handook* of Labor Economics, O Ashfentler and D Card (eds), 4b(15): 1315-1486.

Aughinbaugh, A., and M. Gittleman. 2003. "Does Money Matter: A Comparison of the Effect of Income on Child Development in the United States and Great Britain." *Journal of Human Resources* 38 (2): 416–40.

Bayley, N. 1969. *Manual for the Bayley Scales of Infant Development*, New York: The Psychological Corporation.

Bayley, N. 2006. *Bayley Scales of Infant and Toddler Development – Third Edition*, San Antonio, TX: Harcourt Assessment.

Bernal, R. and C. Van Der Werf. 2011. "Situación de la Infancia en Colombia" in *Colombia en Movimiento: Un análisis descriptivo basado en la Encuesta Longitudinal Colombiana de la Universidad de los Andes*", Universidad de Los Andes, Bogota, Colombia.

Blau, D. M. 1999. "The Effect of Income on Child Development." *Review of Economics and Statistics* 81 (2): 261–76.

Bradley, R. H., and R. F. Corwyn. 2002. "Socioeconomic Status and Child Development." *Annual Review of Psychology* 53: 371–99.

Bruns, B., S. Martinez, S. Naudeau, and V. Pereira. 2010. "Impact Evaluation of Save the Children Early Childhood Development Program in Mozambique. Baseline Results." World Bank, Washington, DC.

Caldwell, B. and R. Bradley. 2001. *HOME inventory and administration manual* (3rd edition), University of Arkansas for Medical Sciences and University of Arkansas at Little Rock.

Case, A., D. Lubotsky and C. Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *The American Economic Review* 92 (5): 1308-1334.

DANE . 2011. *Encuesta Multipropósito de Bogotá*: http://190.25.231.249/pad/index.php/catalog/160/variable/V1505

DNP. 2008. *Evaluación de la Estratificación socio-económica como instrumento de clasificación de los usuarios y herramienta de asignación de subsidios y contribuciones a los servicios públicos domiciliarios*, published by Departamento Nacional de Planeación, Bogota, Colombia.

Fernald, L.C., P. Kariger, M. Hidrobo and P.J. Gertler. 2012. "Socioeconomic gradients in child development in very young children: evidence from India, Indonesia, Peru, and Senegal." *Proceedings of the National Academy of Science* 109 (2): 17273-80.

Fernald, L., A. Weber, E. Galasso and L. Ratsifandrihamanana. 2011. "Socioeconomic gradients and child development in a very low income population: evidence from Madagascar." *Developmental Science* 14(4): 832–847.

Fernald L.C., P. Kariger, P. Engle and A. Raikes. 2009. "Examining Early Child Development in Low-Income Countries: A Toolkit for the Assessment of Children in the First Five Years of Life." World Bank, Washington D.C.

Fernald, L. C., L. M Neufeuld, L. R Barton, L. Schnaas, J. Rivera and P. Gertler. 2006. "Parallel deficits in linear growth and mental development in low-income Mexican infants in the second year of life." *Public Health Nutrition* 9(2): 178–186.

Filmer, D., and S. Naudeau. (2010). "Impact Evaluation on Early Childhood Care and Development (ECCD Programs in Cambodia, Baseline Report)." World Bank, Washington, DC.

Filmer, D. and K. Scott. 2008. "Assessing Asset Indices." World Bank Policy Research Working Paper No. 4605, World Bank, Washington, DC.

Frongillo, A. Edward, Sara M. Sywulka, and Patricia Kariger. 2003. "UNICEF Psychosocial Care Indicators Project. Final report to UNICEF." Mimeo, Cornell University.

Ghuman, S. J. R. Behrman, J. B. Borja, S. Gultiano, and E. M. King. 2005. "Family Background, Service Providers and Early Childhood Development in the Philippines." *Economic Development and Cultural Change* 54(1): 129-64.

Grantham-McGregor S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, B. Strupp and the International Child Development Steering Group. 2007. "Developmental potential in the first 5 years for children in developing countries." *The Lancet* 369(9555):60-70.

Hackman, D. A., and M. J. Farah. 2009. “Socioeconomic Status and the Developing Brain.” *Trends in Cognitive Science* 13 (2): 65–73.

Hamadani J. F. Tofail, S. El Arifeen, S. Huda, D. Ridout, O. Attanasio and S. Grantham-McGregor. 2012. “The Early Determinants of the Cognitive Gap Associated with Poverty in Rural Bangladeshi Children”, mimeo ICDDR, Bangladesh.

Hamadani, J., F. Tofail, A. Hilaly, S. Huda, P. Engel and S. Grantham-McGregor. 2010. “Use of Family Care Indicators and Their Relationship with Child Development in Bangladesh.” *Journal of Health Population and Nutrition* 28(1):23-33.

Heckman J. J. and D.V. Masterov. 2007. “The Productivity Argument for Investing in Young Children.” *Review of Agricultural Economics* 29(3):446-493.

Hoddinott J., J.A. Maluccio, J.R. Behrman, R. Flores and R. Martorell. 2008. “The Impact of Nutrition During Early Childhood on Income, Hours Worked, and Wages of Guatemalan Adults.” *The Lancet* 371(February):411-416.

Macours, K., N. Schady, and R. Vakis. 2012. “Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment.” *American Economic Journal: Applied Economics* 4(2):247-73.

Maluccio J. A., J. Hoddinott, J.R. Behrman, A. Quisumbing, R. Martorell and A.D. Stein. 2009. “The Impact of Nutrition During Early Childhood on Education among Guatemalan Adults.” *Economic Journal* 119(537):734–763.

Paxson, C., and N. Schady. 2007. “Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health, and Parenting.” *Journal of Human Resources* 42 (1): 49–84.

Schady, N. 2011. “Parents’ Education, Mothers’ Vocabulary, and Cognitive Development in Early Childhood: Longitudinal Evidence from Ecuador.” *American Journal of Public Health* 101(12): 2299-2307.

Shonkoff J. 2010. “Building a New Biodevelopmental Framework to Guide the Future of Early Childhood Policy.” *Child Development* 81(1):357–367.

Uribe-Mallarino. 2008. “Estratificación social en Bogotá: de la política pública a la dinámica de la segregación social”, manuscript Universidad Javeriana, Bogota, Colombia.

Walker S., S. Chang-Lopez, C. Powell, E. Simonoff and S. Grantham-McGregor. 2005. “Effects of Early Childhood Psychosocial Stimulation and Nutritional Supplementation on Cognition and Education in Growth-stunted Jamaican Children: Prospective Cohort Study.” *The Lancet* 366(9499):1804-1807.

Walker S.P., S.M. Chang, C. Powell, E. Simonoff and S.M. Grantham-McGregor. 2006. “Effects of Psychosocial Stimulation and Dietary Supplementation in Early Childhood on Psychosocial Functioning in Late Adolescence: Follow-up of Randomized Controlled Trial.” *British Medical Journal* 333(7566): 472.

Walker S.P., T.D. Wachs, J. Meeks Gardner, B. Lozzoff, G.A. Wasserman, E. Pollitt, J.A. Carter and the International Child Development Steering Group. 2007. “Child developmental: risk factors for adverse outcomes in developing countries.” *The Lancet* 369(9556):145-157.

Walker S.P., S.M. Chang, C. M. Vera-Hernandez and S.M. Grantham-McGregor. 2011. “Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behaviour”, *Pediatrics* 127(5): 849-857.

DATA APPENDIX

AI. Sampling Strategy

Residential blocks in Bogota are classified in six *estratos* according to their location (industrial, commercial, residential, marginalized areas, etc.), quality of streets and pavements, accessibility to households, and housing quality (materials of roofs and front walls, size of the façade, parking and garden). In practice, *estratos* are very concentrated amongst them and separate from each other, as shown in Figure A1 in the Appendix. Data from the census indicate that both family size and the number of young children per household vary substantially by *estrato*, with households in higher *estratos* having fewer children and lower family sizes on average (see Table A1 in the Appendix). Given these considerations, it was deemed important to sample households in each of the first four *estratos* separately and independently, in order to obtain a sample representative of the low- and middle-income levels.

We followed a three stage sampling process. First, we stratified the city by *estrato* and randomly sampled neighborhoods (*secciones*, the primary sampling unit) within them, weighting by the proportion of women in fertile age (13-49 years). Second, within each neighborhood, we randomly sampled 3 blocks (*manzanas*, the secondary sampling unit), also weighting by the proportion of women in fertile age.[24] Third, in each selected block, we carried out a mini-census (door-to-door sampling) to identify households with children (tertiary sampling unit) in the eligible age ranges. To ensure there would be enough children of a given month-age in the sample, we stratified eligible children in a block by age categories: 6 to 14, 15 to 23, 24 to 32, and 33 to 41 months. Based on the 2005 Census, we estimated that there would be 10 eligible children per block and age group on average. Of these, 8 were included in the study by random draw (potential sample) and we expected a rejection rate of 25%.[25] In anticipation of the fact that there would be less eligible children, as well as higher rejection rates amongst E4 families, we decided to include *all* children satisfying the inclusion criteria in the blocks in this *estrato*.

The original sample design was perfectly balanced across age groups and *estratos*, with 90 children in each stratum-age cell for a total of 1,440 children in 240 blocks. These sample sizes would allow detecting differences of 0.415 SD of a z-score within stratum-age group at 80% power and 5%

---

[24] Because of budget and logistical constraints, we opted for a proportional probability design whereby neighbourhoods (and within them, blocks) with a higher proportion of women in fertile age had a higher probability of being included in the sample.

[25] We excluded from the potential sample one child with mental disabilities and a couple of twins. In addition, in the 4 households where there was more than one child satisfying the inclusion criteria, we randomly selected one to be included in the potential sample.

significance. However, as soon as field operations started, it was clear that households living in E4 were extremely reluctant to participate in the survey, mostly because of apparent mistrust. Moreover, relative to the data in the 2005 Census, we found a much reduced number of children per block. This situation induced us to modify the sample structure in two ways. Firstly, we increased the number of blocks sampled. Secondly, we replaced the 360 children we had originally planned for E4 with children in the other 3 *estratos* as follows: 90 children from E1, another 90 from E2 and the remaining 180 children from E3. The distribution of our final sample is presented in Table 1 in the main text.

AII. <u>Data Collection Strategy and Measurements</u>

Data was collected in three stages:

(i) <u>*Door-to-Door Sampling*</u>: after selecting the neighborhoods and blocks to be included in the sample, two enumerators visited *all* households in each block in order to identify *all* children 6 to 42 months living in it. These children were stratified by age category and 80% (per block and age group) were randomly drawn to be included in the sample. All included children in a block were randomly assigned to one of eight interviewers (all female), ensuring that the number of children per *estrato* and age group was balanced across interviewers.

(ii) <u>*Household Survey*</u>: the interviewer contacted her assigned households by phone and arranged a convenient time for a visit to apply the household survey by direct interview with the biological mother of the target child. The survey collected basic socio-economic information on the household such as demographic composition, education level and employment status for all household members, income, expenditures, dwelling characteristics and assets; as well as other more specific information regarding the child's nutritional status (birth weight, gestational age, breastfeeding and weaning, etc.) and informal and formal care arrangements (main carer and nursery/pre-school attendance by type of center).

We also administered a slightly modified *Family Care Indicator* (FCI), which collects by direct observation the number of newspapers, magazines and books for adults in the household, as well as the number of toys the child usually plays with classified by type—bought or home-made—and by use/purpose: toys to play music, materials to color and paint (including books), picture books, toys to play pretend games, toys that involve movement such as balls, construction toys such as blocks, and toys for shapes and colors. The FCI also asks the mother or main caregiver about the type of play activities an adult in the household engaged with the child over the 7 days before the interview, as well as the number of times child and adult performed the activity together.[26] Play activities included are: play with the child's toys; go for a walk; paint, color and scribble; sing songs; name objects, color or count; look at picture books; and tell stories. We use these data to construct three scores 'books, newspapers and magazines for adults', 'play activities' and 'play materials', which together compose what we call the 'level of stimulation in the house'.

(iii) *Administration of the Bayley*-III *and Anthropometric Measurements*: upon completion of the household survey, mother and interviewer set an appointment to assess the developmental level of the child using the Bayley-III, and to collect height and weight on both mother and child. These measures were collected by six qualified psychologists ('testers') in the library or public child care center closest to the child's home.

To avoid biases, the Bayley-III must be administered to all children under the same conditions. These involve: (i) a quiet environment, different from the child's home, where the child can focus on the test without being distracted by his own toys, other siblings, the TV, etc.; and (ii) a large enough room (about 3 m^2) that can comfortably fit a medium-sized table, three chairs (one for the child, one for the mother and one for the tester) and the materials required for some of the items in the gross motor scale (a set of six steps, one meter of measurement tape, etc.). To avoid differential sample loss between the survey and the Bayley-III test by *estrato*, it was crucial we fit out spaces satisfying these requirements in all neighborhoods in the sample. To this end, we arranged testing sites in all libraries in the local public library network in Bogota (*BibloRed*) and in public child care centers (*Jardines Sociales*) as they are well-spread all over the city. In return for lending us their facilities, we offered workshops on parenting and child rearing practices to the centers'/libraries' staff and parents. Tested children were given a set of picture books and nutritional supplements (vitamins and minerals) for daily consumption over 3 months as a present. The mother received feedback on her child's performance in the test, brochures on parenting and $10,000 pesos (about $5.6 US) to compensate for travel costs to the testing site.

Door-to-door sampling activities were scheduled two weeks ahead of the household survey to ensure that there would be enough children to be interviewed in a block. Similarly, we aimed to administer the Bayley-III within a week from the household survey. We strictly monitored the ages and *estrato* of all tested children throughout the process to guarantee a final well-balanced sample. Moreover, data collection was organized such that all interviewers/testers interviewed/tested a similar number of children in each *estrato* and age group, *and* that these were equally distributed over the 6 months of field activities. This is important in order to minimize measurement biases due

[26] In the original FCI, this question refers to the past 3 days.

to: (i) child *estrato* (e.g. the tester scores children from different backgrounds differently to compensate for perceived disadvantages); (ii) child age (e.g. the tester may find it easier to test older children); and (iii) seasonality or time effects (e.g. measurements are less accurate because they are assessed faster in periods of heavy rains where travelling around the city is harder or testers may become stale/careless over time).

AIII. The Bayley Scales of Infant and Toddler Development

The *Bayley Scales of Infant Development* (BSID) were developed by Nancy Bayley in 1969 to assess child development from birth to 30 months. In 1993, a second version (BSID-II) expanded the age range to 42 months. A third version (Bayley-III) was published in 2006. This is the version we use in our study. The BSID are largely used internationally, have been well-validated in the US and have shown good predictive ability of later development and academic achievement. While there is not a lot of information yet on the use of the Bayley-III, it is anticipated that this version will keep its assessment and predictive properties (Fernald et al. 2009). The test consists of five scales:

(i) The *Cognitive Scale* measures learning processes, processing speed, problem solving, memory, attention to novelty and habituation, counting and classification, and playing skills (from solitary non-relational play to social fantasy play). This scale primarily requires non-verbal responses from the child.

(ii) The *Language Scale* is in turn composed by the language receptive and language expressive sub-scales. The first measures the child's ability to respond to stimulus in the environment (localize a sound and respond to it, discriminate between sounds, etc.) and to comprehend and respond appropriately to words and requests. The latter assesses the ability to imitate sounds and words, to vocalize and verbalize wants and needs, to name pictures of objects and actions, and language production, amongst others.

(iii) The *Motor Scale* is also divided in two sub-scales: fine and gross motor. The fine motor sub-scale measures hand and fingers coordination and hand and eye coordination, including the child's ability to grasp, reach and produce specific and controlled movements with their hands. The gross motor sub-scale measures the child's control of their body, equilibrium, posture and movement of the torso and extremities.

(iv) The *Socio-Emotional Scale* measures social and emotional milestones, such as self-regulation, communication needs, how the child relates and interacts with familiar and non-familiar people, attention, and other temperament and social behavior aspects.

(v) The *Adaptive Behavior Questionnaires* measure daily functional abilities of the child in nine different areas: communication, community use, functional pre-academics, home living, health and safety, leisure, self-care, self-direction, social, and motor.

The scales can be administered and scored independently so that domain-specific assessments can be made. The first three are assessed by direct observation of the child's performance in a set of tasks; the socio-emotional and adaptive behavior questionnaires are based on maternal/caregiver reports. We administered all scales except the adaptive behavior questionnaire, which was excluded because of time constraints. Furthermore, some items appeared to be culturally inappropriate.

Administration times depend on the child characteristics (age, temperament, level of development, etc.) but typically vary between 50 and 90 minutes. In our case, administration times ranged from 55 to 110 minutes, on average. We found that the cognitive scale was the most difficult to train and administer, as well as the longest to give, at over 20 minutes on average. Each language and motor sub-scales took less than 20 minutes to administer and the socio-emotional questionnaire took about 10 minutes to complete.

We translated the test to Spanish and then back-translated into English to ensure linguistic and functional equivalence. We piloted the translation intensively and made minor modifications (wording and phrasing) to guarantee the items would be well-understood. 20 children across the entire age range were tested a second time between 6 and 19 days after the first test (median=8 days) by either the trainer or one of three testers to compute test-retest reliabilities. Test-retest intra-class correlations vary from r=[0.95-0.98] for the cognitive, language and motor scales and r=0.87 for the socio-emotional questionnaire, indicating that the translated versions offered stable measurements over time.

For the test administration, we hired six recent female Psychology graduates and trained them intensively during six weeks, including 20 to 25 practice administrations per tester. Some of the testers had previous experience in testing. The testers did the practices in couples and inter-rater reliabilities were computed. Practice testing continued until an intra-class correlation of over 0.9 was obtained on each scale, between each pair of testers and between the tester and the trainer, who supervised the process throughout. Furthermore, 5% of the measurements during field activities were supervised by the trainer and corrective feedback was given when appropriate. The intra-class correlations between tester and trainer scores during these tests were all well above 0.9, which guarantees high quality data.

APPENDIX TABLES AND FIGURES

Table A1: Demographic Composition by *Estrato* in Bogota

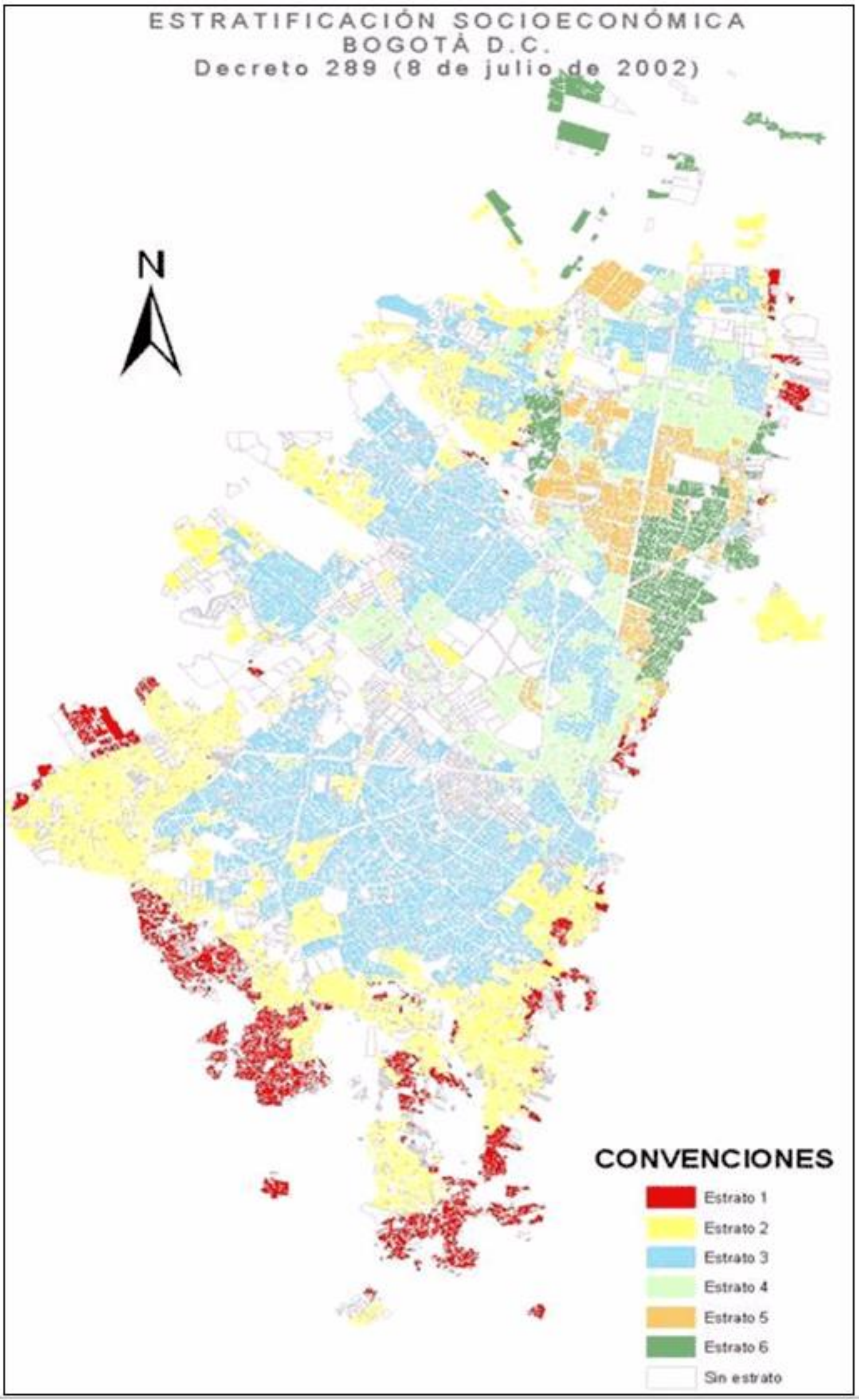
	Income Level	Blocks	Proportion Blocks	Households	Households per Block	Average Household Size	Prob Child in Household	Average Number Women 13-49
E1	Low-low	3530	0.1262	102464	29.0	3.9	0.135	62
E2	Low	10870	0.3887	553257	50.9	3.6	0.117	111
E3	Middle-low	10223	0.3656	589859	57.7	3.4	0.089	129
E4	Middle	1923	0.0688	118259	61.5	3	0.056	214
E5	Middle-high	805	0.0288	43467	54.0	3	0.054	N/A
E6	High	612	0.0219	35826	58.5	2.8	0.054	N/A
TOTAL		27963		1443132				

Table A2: Contribution of Other Factors to Child Development

	Cognition	Receptive Language	Expressive Language	Fine Motor	Socio-Emotional
Female =1	0.172**	0.144**	0.279**	0.285**	0.082
	(0.050)	(0.052)	(0.057)	(0.053)	(0.054)
Parental Factors					
Mother has Some Secondary Education =1	0.116	0.250**	0.103	0.152	-0.175*
	(0.093)	(0.084)	(0.093)	(0.092)	(0.073)
Mother has Secondary Completed or More =1	0.174+	0.308**	0.139	0.216*	-0.091
	(0.094)	(0.097)	(0.105)	(0.103)	(0.093)
Mother Deceased or No Longer Living with Child =1	-0.575**	-0.548**	-0.534*	-0.554*	-0.285+
	(0.195)	(0.191)	(0.205)	(0.224)	(0.171)
Father Deceased or No Longer Living with Child =1	-0.062	-0.020	-0.076	0.008	-0.052
	(0.051)	(0.059)	(0.063)	(0.056)	(0.062)
First Child =1	0.043	0.091+	0.146**	-0.068	-0.001
	(0.051)	(0.050)	(0.054)	(0.051)	(0.060)
Child Biomedical Factors					
Premature (gestational age < 37 weeks) =1	-0.410+	-0.347	-0.326	-0.151	0.173
	(0.212)	(0.210)	(0.230)	(0.184)	(0.171)
Premature * Age	0.010	0.008	0.006	0.005	-0.009
	(0.007)	(0.007)	(0.008)	(0.007)	(0.006)
Birth Weight (/1000)	0.241**	0.064	0.056	0.174**	-0.044
	(0.058)	(0.056)	(0.060)	(0.057)	(0.062)
Stunted (z-height-for-age < -2 SD) =1	-0.031	-0.111+	-0.067	-0.179**	0.022
	(0.064)	(0.063)	(0.068)	(0.064)	(0.074)
Quality of the Home Environment Factors					
Books and Newspapers (FCI Score)	-0.008	0.004	0.025	-0.009	-0.007
	(0.016)	(0.016)	(0.016)	(0.017)	(0.013)
Play Materials (FCI Score)	0.045**	0.025+	0.018	0.033*	0.013
	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)
Play Activities (FCI Score)	0.039*	0.072**	0.048**	0.016	0.069**
	(0.016)	(0.017)	(0.017)	(0.019)	(0.017)
Institutional Care Factors					
Public Preschool =1	0.203**	0.223**	0.089	0.245**	-0.088
	(0.077)	(0.079)	(0.071)	(0.079)	(0.081)
Private Preschool =1	0.291**	0.278**	0.198+	0.196	-0.062
	(0.103)	(0.101)	(0.105)	(0.120)	(0.113)
Hogar Comunitario =1	0.273**	0.191*	0.075	0.160+	-0.171*
	(0.081)	(0.084)	(0.102)	(0.089)	(0.086)
Cubic Spline in Wealth * Age	Yes	Yes	Yes	Yes	Yes
Tester Dummies	Yes	Yes	Yes	Yes	Yes
p-value F-test: (Parental Factors =0)	0.03	0.00	0.01	0.06	0.08
p-value F-test: (Biomedial Factors =0)	0.00	0.01	0.03	0.00	0.55
p-value F-test: (Home Environment Factors =0)	0.00	0.00	0.00	0.03	0.00
p-value F-test: (Institutional Care Factors =0)	0.00	0.00	0.23	0.01	0.23
R-Sq Adjusted	0.18	0.18	0.15	0.12	0.10

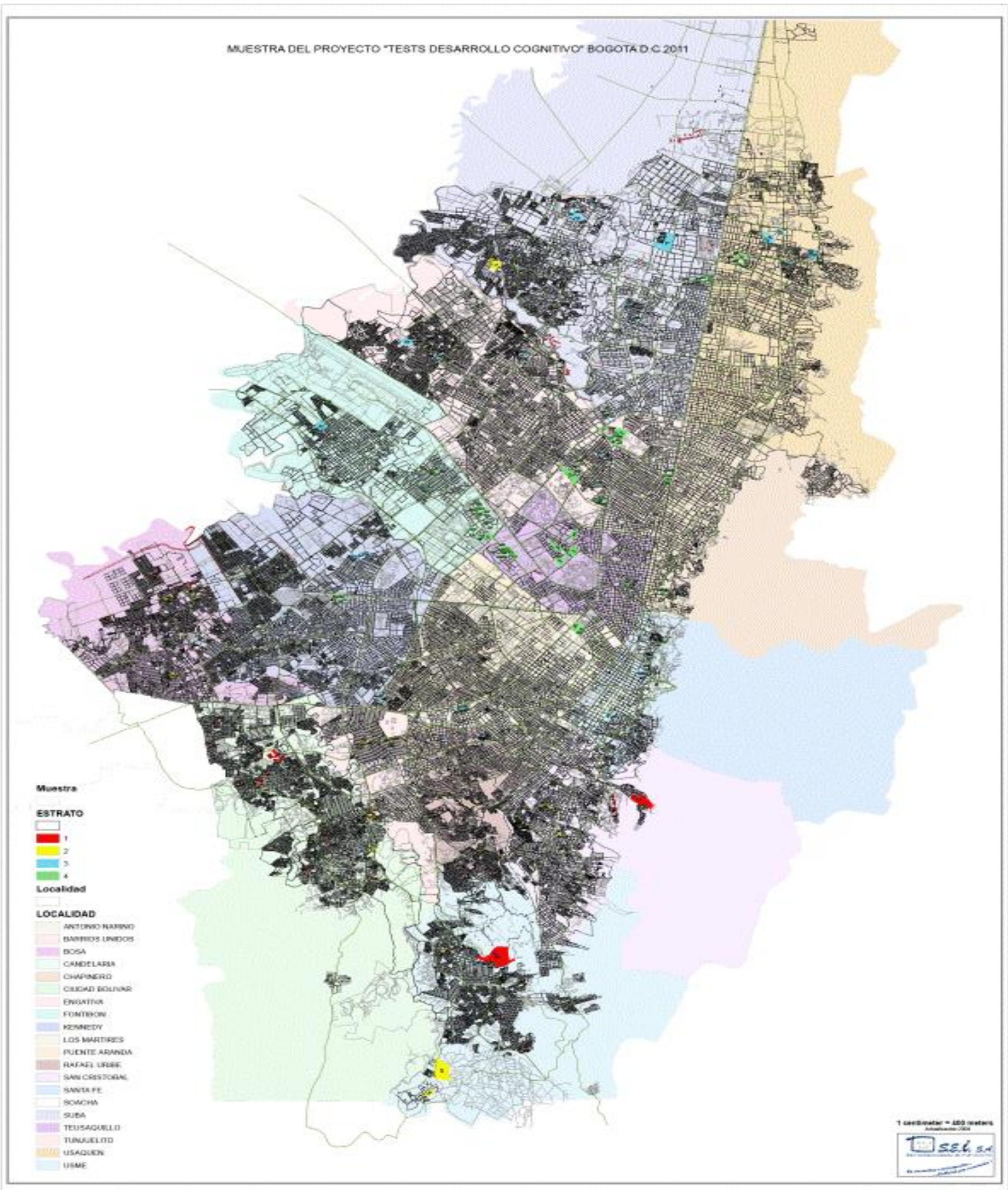
Notes: +significant at 10%, *significant at 5%, **significant at 1%. N= 1330. Regression coefficients from the regressions that generate the figures in Model 5 in Tables 8 and 9. For each developmental outcome, these regressions fit a cubic spline in wealth with two nodes interacted with age linearly, controlling for child's sex, tester dummies, parental factors, child biomedical factors, and variables related to the quality of the home environment and to the institutional care setting. Missing values for birth weight have been imputed with the predicted value from a regression of birth weight on sex, gestational age and height-for-age. We have accounted for the replacement with a dummy. SE are clustered at the neighbourhood level (primary sampling unit or sección).

Figure A1: Spatial Distribution of *Estratos* in the City of Bogota



Source: http://institutodeestudiosurbanos.info/endatos/0200/02-030-vivienda/02.03.01.htm

Figure A2: Spatial Distribution of the Study Sample



Source: Field Team at SEI S.A.

Figure A3: Kernel Densities Internally Standardized Scores (in blue) and Normal Density (in red)

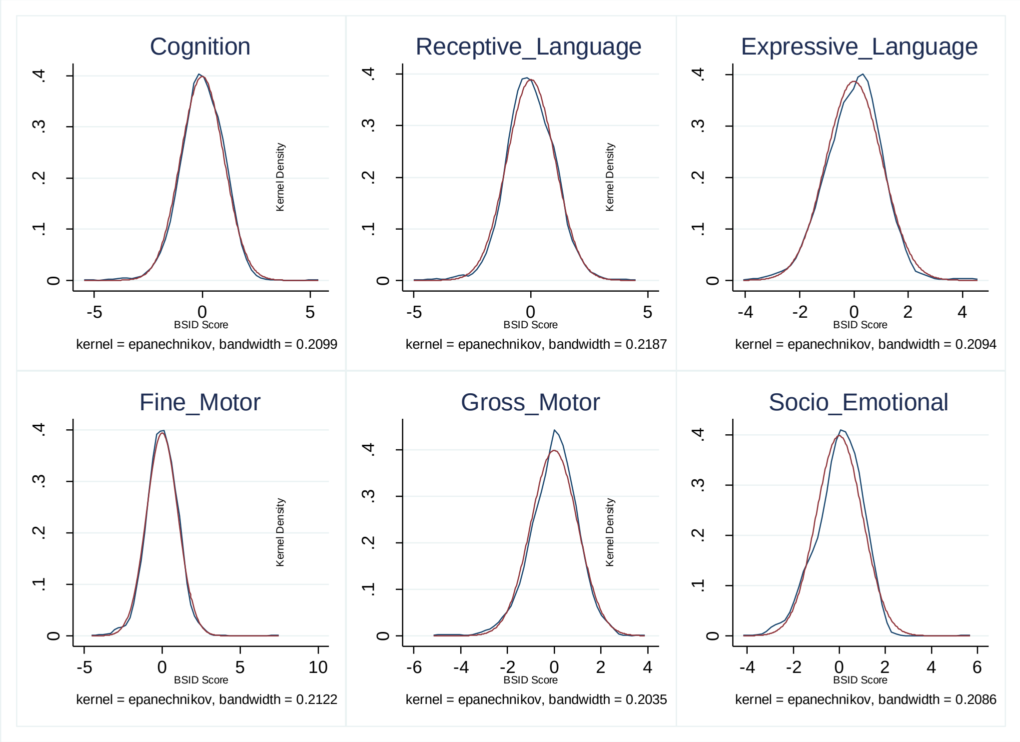


Figure A4: Bayley-III Internally Standardized Scores over Age -- Non-Parametric Regressions