Title

Labor Market Returns to Early Childhood Stimulation: a 20-year Followup to an Experimental Intervention in Jamaica

Authors

Gertler, Paul
Heckman, James
Pinto, Rodrigo
et al.

Publication Date

2013-04-16

Series Name: WPS
Paper No.: 027
Issue Date: 16 April 2013

Labor Market Returns to Early Childhood Stimulation: a 20-year Followup to an Experimental Intervention in Jamaica

Paul Gertler, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeerch, Susan Walker, Susan Chang-Lopez, Sally Grantham-McGregor

Working Paper Series

Center for Effective Global Action

University of California

The CEGA Working Paper Series showcases ongoing and completed research by faculty affiliates of the Center. CEGA Working Papers employ rigorous evaluation techniques to measure the impact of large-scale social and economic development programs, and are intended to encourage discussion and feedback from the global development community.

Recommended Citation:

Gertler, Paul; Heckman, James; Pinto, Rodrigo; Zanolini, Arianna; Vermeerch, Christel; Walker, Susan; Chang-Lopez, Susan; Grantham-McGregor. (2013). Labor Market Returns to Early Childhood Stimulation: a 20-year Followup to an Experimental Intervention in Jamaica . CEGA Working Paper Series No. WPS-0267 Center for Effective Global Action. University of Califoria, Berkeley.

Labor Market Returns to Early Childhood Stimulation: a 20-year Followup to an Experimental Intervention in Jamaica

Paul Gertler[a]
James Heckman[b,c]
Rodrigo Pinto[b]
Arianna Zanolini[b]
Christel Vermeerch[e]
Susan Walker[d]
Susan Chang-Lopez[d]
Sally Grantham-McGregor[f]

March 20, 2013 1:15pm

Acknowledgements: The authors gratefully acknowledge research support from the World Bank Strategic Impact Evaluation Fund, the American Bar Foundation, The Pritzker Children's Initiative, NICHD R37HD065072, R01HD54702, the Human Capital and Economic Opportunity Global Working Group - an initiative of the Becker Friedman Institute for Research in Economics funded by the Institute for New Economic Thinking (INET), and an anonymous funder. We acknowledge the support of a European Research Council grant hosted by University College Dublin, DEVHEALTH 269874. We have benefitted from comments of participants in seminars at the University of Chicago, UC Berkeley, MIT, the 2011 LACEA Meetings in Santiago Chile and the 2013 AEA Meetings. The authors have not received any compensation for the research nor do they have any financial stake in the analyses reported here.

Author Affiliations: [a]University of California Berkeley, [b]University of Chicago, [c]American Bar Foundation, [d]The University of The West Indies, [e]The World Bank, [f]University of London

Corresponding Author Contact: Gertler@haas.berkeley.edu

Abstract

We find large effects on the earnings of participants from a randomized intervention that gave psychosocial stimulation to stunted Jamaican toddlers living in poverty. The intervention consisted of one-hour weekly visits from community Jamaican health workers over a 2-year period that taught parenting skills and encouraged mothers to interact and play with their children in ways that would develop their children's cognitive and personality skills. We re-interviewed the study participants 20 years after the intervention. Stimulation increased the average earnings of participants by 42 percent. Treatment group earnings caught up to the earnings of a matched non-stunted comparison group. These findings show that psychosocial stimulation early in childhood in disadvantaged settings can have substantial effects on labor market outcomes and reduce later life inequality.

Key Words: early childhood development, stunting, randomized trial

JEL Codes: 015, I20, I10, I25

Contents

1 Introduction

Early childhood, when brain plasticity and neurogenesis are very high, is an important period for cognitive and psychosocial skill development.[1] Investments during this period create the foundations for the evolution of the cognitive and psychosocial skills that are key determinants of lifetime earnings.[2] Young children who experience negative shocks such as economic downturns, extreme weather, and infectious diseases suffer long lasting consequences in terms of their educational and labor market outcomes.[3] The seeds of inequality are planted in early life with remediation being less effective and more expensive later in life.[4]

Today more than 200 million children under the age of 5 living in developing countries are at risk of not reaching their full developmental potential. The vast majority of these children live in extreme poverty.[5] These children start disadvantaged, receive lower levels of parental investments, and are likely to continue to fall further behind without help than are children from more affluent environments.[6] Based on a growing body of evidence demonstrating positive impacts, early childhood development (ECD) interventions aimed at skill development are being promoted as cost-effective remediation policies to help these children.[7] While these ECD interventions are estimated to have substantially higher rates of return than investments in the human capital of the disadvantaged later in life,[8] there is little rigorous evidence on the long-term effects of ECD on earnings and inequality for developing countries.

[1] See Huttenlocher (1979, 2002) and Thompson and Nelson (2001).

[2] See e.g. Knudsen et al. (2006), Borghans et al. (2008), and Almlund et al. (2011a).

[3] See van den Berg et al. (2006), Almond et al. (2007), Bleakley (2007), Maccini and Yang (2009) and Almond and Currie (2011).

[4] See Carneiro and Heckman (2003), Cunha et al. (2006), Heckman (2008) and Cunha et al. (2010).

[5] See Grantham-McGregor et al. (2007) and Walker et al. (2007).

[6] See Paxson and Schady (2007), Fernald et al. (2011), Fernald et al. (2012) and Engle et al. (2011).

[7] See e.g. Engle et al. (2007), Heckman (2008) and Engle et al. (2011).

[8] See Heckman (2000, 2008), Cunha et al. (2006), Almond and Currie (2011).

This paper reports estimates of the labor market returns to an intervention that gave psychosocial stimulation and nutritional supplementation to growth-retarded toddlers living in poverty in Jamaica (Grantham-McGregor et al., 1991). Enrollment in the study was conditioned on stunting because it is an easily and accurately observed indicator of malnutrition that is strongly associated with poor cognitive development (Walker et al., 2007). The randomized treatment group assigned to stimulation received weekly visits for a period of two years from community health workers who actively encouraged mothers to interact and play with their children in ways designed to develop cognitive and psychosocial skills. Unlike the effects of many other early childhood interventions that fade out over time,[9] the Jamaican stimulation intervention proved to have large impacts on cognitive development twenty years later (Walker et al., 2011). In contrast, the nutritional intervention had no long-term impact on any outcome.

We use labor market information collected 20 years after the intervention when the participants were 22 years old. We show that stimulation increased average earnings by 42%. The magnitude of the estimated impact on earnings is put into perspective when compared to a non-stunted comparison group identified at baseline. In fact, the earnings of the treated stunted group completely caught up with the earnings of the matched non-stunted comparison group. These results provide evidence that stimulation interventions very early in life can compensate for developmental delays and thereby reduce inequality later in life.

We also examine pathways through which the intervention likely affected earnings. First, we find that the intervention increased maternal investment in children during the intervention period. Second, there are large effects on key determinants of earnings including schooling, cognitive development, and psychosocial development. Finally, we show that the treatment group was more likely to migrate to the U.S. or U.K., and thereby gained access to higher quality schools and better labor markets.

[9]See Cunha et al. (2006), Almond and Currie (2011) and Engle et al. (2011) for reviews.

To our knowledge, our study is the first experimental evaluation of the impact of an ECD stimulation intervention on long-term economic outcomes and inequality in a developing country.[10] This study contributes to a small literature on labor market returns to ECD programs including Perry Preschool, the Chicago Parent Child program, Abecedarian and Head Start, all of which are located in the U.S.[11] We find that the Jamaica stimulation program had substantially larger effects on earnings than any of the U.S. programs.

2 The Jamaican Study

2.1 The Intervention and Experimental Design

In 1986-1987, the Jamaican Study enrolled 129 stunted children age 9-24 months that lived in poor disadvantaged neighborhoods of Kingston, Jamaica (Walker et al., 1990). Stunting was defined as having a standardized height for age z-score less than -2. The children were stratified by age (above and below 16 months) and sex. Within each stratum, children were sequentially assigned to one of four groups using a randomly generated seed to begin the assignment. The four groups were (1) psychosocial stimulation (N=32), (2) nutritional supplementation (N=32), (3) both psychosocial stimulation and nutritional supplementation (N=32), and (4) a control group that received neither intervention (N=33). All children were given access to free health care regardless of the group to which they were assigned.

The stimulation intervention (comprising groups 1 and 3) consisted of two years of weekly one-hour play sessions at home with trained community health aides.[12] The curriculum for the cognitive stimulation was based on Piagetian concepts (Powell and Grantham-McGregor,

[10] While ours is the first to study labor market returns to ECD psychosocial stimulation in a developing county, there are labor market follow-ups to nutritional interventions. See, for example, Hoddinott et al. (2008), Maluccio et al. (2009).

[11] See Heckman et al. (2010a), Heckman et al. (2010b), Reynolds et al. (2004), Reynolds et al. (2007), Reynolds et al. (2011), Campbell et al. (2002), Campbell et al. (2012), Campbell, Conti, Heckman, Moon, and Pinto (2012); Aughinbaugh (2001), and Garces et al. (2002).

[12] The aides received 8 weeks of training in nutrition and primary health care and another 8 weeks of training in child development, teaching techniques and toy making.

1989). Mothers were encouraged to converse with their children, to label things and actions in their environments and to play educational games with their children (Grantham-McGregor et al., 1987). Particular emphasis was placed on language development, the use of praise, and on improving the self-esteem of both the child and of the mother. At age 24 months, the curriculum was enriched to include concepts such as size, shape, position, quantity, color, etc based on the curriculum in Palmer (1971).

The focus of the weekly play sessions was on improving the quality of the interaction between mother and child. Mothers were encouraged to continue practicing the activities and games learned during the visits on a continuing basis beyond the home visitation time. At every visit, homemade toys were brought to the home and left for the mother and child to use until the next visit when they were replaced with new ones. The intervention was innovative both for its focus on activities to promote cognitive and language development and for its emphasis on direct mother-child interactions.

The nutritional intervention (comprising groups 2 and 3) was aimed at compensating for the nutritional deficiencies that may have caused stunting. The nutritional supplements were provided weekly for a two-year period. The supplements consisted of one kilogram of formula containing 66% of daily-recommended energy (calories), 100% of daily-recommended protein, and 100% of daily-recommended micronutrients (Walker et al., 1992). In addition, in an attempt to minimize sharing of the formula with other family members, the family also received 0.9 kilograms of cornmeal and skimmed milk powder. Despite this, sharing was common and uptake of the supplement decreased significantly during the intervention (Walker et al., 1991).

Of the 129 study participants, two of the participants dropped out before completion of the two-year program. The remaining 127 participants were surveyed at baseline, resurveyed immediately following the the end of the two-year intervention, and again at ages 7, 11, and 18. Our analysis is based on a re-interview of the sample in 2007-08 when the participants were approximately 22 years old, some 20 years after the original intervention.

2.2 External Comparison Group

For comparison purposes, the study also enrolled a sample of non-stunted children from the same neighborhoods, where non-stunted was defined as having a height for age z-score greater than -1 standard deviations. At baseline, every fourth stunted child in the study was matched with one non-stunted child who lived nearby and was the same age (plus or minus 3 months) and sex. At age 7, this sample of 32 was supplemented with another 52 children who had been identified in the initial survey as being non-stunted and fulfilled all other inclusion criteria.

While the non-stunted group was better off than the stunted group in terms of their personal development and their socioeconomic status, the non-stunted children were still living in the same economically and socially disadvantaged Kingston neighborhoods. Members of the non-stunted comparison group did not receive any of the interventions, but did receive the same free health care as those in the stunted experimental group. From age 7 onwards, this group was surveyed at the same time as the participants in the experiment.

This sample is used to investigate the extent to which the early childhood stimulation intervention helped to compensate for initial disadvantage by comparing the stunted treatment group with the non-stunted external comparison group. We define complete catch-up as no difference between the treated stunted group and the non-stunted comparison group.

In order to better understand the external validity of the catch-up analysis we compare the non-stunted group to the general population using data from two surveys that are representative of urban Jamaica: (1) the 1992 Jamaican Survey of Living Conditions (JSLC) that was collected when the children were 7 years old and when most of the non-stunted sample was first surveyed, and (2) the 2008 Jamaica Labor Force (JLF) survey that was collected in the same year as the last follow-up. Unfortunately the labor supply and earnings questions in the JLF and in our survey were asked in different ways, and there was a 50% non-response rate in the JLF to the earnings questions among those who were employed. Only the education variables are directly comparable.

Comparing childhood conditions in 1992 we find that the non-stunted comparison group grew up in more disadvantaged settings than the general population living in the urban Jamaica.[13] The non-stunted sample was less likely to live in houses with piped water, their mothers were less likely to have completed grade 9 at school, and they were less likely to have the father present in the house. Despite this, by age 22, the non-stunted group attained comparable levels of human capital as those of the same age and living in the Kingston Area interviewed in the Labor Force Survey. The two samples are equally likely to still be in school and achieved the same level of educational attainment in terms of years of schooling and passing national comprehensive matriculation exams.[14]

2.3 Previous Studies

The stimulation and the combined stimulation-nutrition arms of the Jamaica Study proved to have a large long-term impact on cognitive development. At age 22 the order of magnitude of the impacts of stimulation were large at 0.6 standard deviations on a WAIS test (Walker et al., 2011). While the treatment groups' cognitive scores improved relative to those of the control group and caught up with the non-stunted sample in performance IQ , they did not completely catch up in all cognitive function domains (Walker et al., 2005, 2000). Moreover, both stimulation arms had positive impacts on psychosocial skills, schooling attainment and crime reduction Walker et al. (2011). However, there was no long-term impact on anthropometric measures (Walker et al., 1996).

While the stimulation arms had strong and lasting effects, the nutrition-only arm had no long-term effect on any outcome (Walker et al., 2005, 2000).[15] Hence, we combine the

[13] See Table A.1 Panel A in the Appendix.

[14] See Table A.1 Panel B in the Appendix.

[15] This is in contrast to the Guatemala Study in which nutritional supplementation did affect both long-term health status and earnings (Hoddinott et al., 2008 and Maluccio et al., 2009). This may be due to the fact that the Guatemala study started supplementing children earlier, in utero and right at birth, while when the Jamaica program started children were already undernourished. Since there is no study showing sustained benefits from supplementation in children who were malnourished before beginning supplementation, supplementation in Jamaica may have begun too late to have an impact. Other possible reasons for the difference include the fact that the supplement was offered for less time in Jamaica, the supplement

two psychosocial stimulation arms into a single treatment group (N=64) and combine the nutritional supplementation only group with the pure control group into a single control group (N=65). Henceforward we use the term stimulation effects of stunted participants to designate the analysis that compares groups 1 and 3 against groups 2 and 4.

3 The New Survey

We resurveyed both stunted (experimental) and non-stunted (comparison) study populations in 2007 and 2008 some 20 years after the original intervention when the participants were approximately 22 years old.[16] We attempted to find all of the study participants regardless of current location and followed migrants to the the US, Canada, the UK and the Caribbean. When we could not find a participant in Jamaica, we contacted relatives for further information to find the participants.

3.1 The Experimental Sample

We were able to find and interview 105 out of the original 127 (83%) stunted participants who completed the program. For this sample, Table 1 reports the baseline means for the treatment and control groups, the difference in the means of the two groups, and p-values for two sided permutation tests of equality of means. We observe significant differences in 3 out of 19 variables. Mothers of children in the treatment group were more likely to be employed and have completed less schooling than mothers of children in the control group, and children in treatment group had lower weight for height than children in the control group. These imbalances are already present in the full baseline sample of 127, which suggests that they

was more intensively shared with other family members in Jamaica, the formula provided in the Jamaica intervention had fewer micronutrients, and the supplement was a smaller share of the total food budget in Jamaica (Hoddinott et al., 2008; Walker et al., 1992, 1990.)

[16]The survey received ethical clearance from the IRB of the University of the West Indies in Kingston Jamaica.

were the result of sampling variation in the original randomization rather than differential sample attrition.[17]

The attrition rate from the experiment is 17%. Of 22 participants that dropped out of the sample, 10 were not found, 9 died, and 3 of those who were found refused to be interviewed.[18] In addition, treatment status is not a significant predictor of the overall probability of attrition or for any of the reasons for attrition. And, with just 4 exceptions out of 57, the means of individual variables are not significantly different between the group that dropped out and the group that stayed in the sample, even when we stratify by treatment and control.[19] Hence, in terms of measured variables, there appears to be no selective attrition and the remaining sample is representative of the original sample.

3.2 Non-Stunted Comparison Sample

We found and interviewed 65 children out of the 84 children originally surveyed with an implied attrition rate of 23%, which is slightly higher than that for the experimental sample. In the baseline samples, 9 out of 19 characteristics are statistically significantly different between stunted and non-stunted. As expected, the non-stunted were less disadvantaged.[20] Non-stunted children have taller mothers with higher Picture Peabody Vocabulary Test (PPVT) scores and have higher birth weight, larger head circumference, and higher initial developmental scores.

Attrition from the non-stunted sample appears to be selective.[21] Mothers in the attrition group are older, better educated, and perform better on the PPVT than mothers who do not attrite. In addition, children in the attrition group lived in homes with more verbal

[17] See Table A.2.
[18] See Table A.3.
[19] See Table A.4.
[20] Table A.6 in the Appendix shows the differences in baseline and 7 years characteristics for the stunted and non-stunted samples. The top panel only includes the baseline variables, where the non-stunted group only consists of 32 people, while the bottom panel also includes variables at 7 years, when the additional 52 non-stunted children were added.
[21] See Table A.5.

stimulation and better housing infrastructure. This suggests that the remaining sample is not representative of the original sample.

4 Methods

We investigate two questions. First, what are the impacts of treatment on earnings and associated outcomes? We identify the impacts by comparing the randomized treatment and control groups from the stunted sample. Second, did stimulation enable the stunted treatment group to catch up with the non-stunted group? We identify catch-up by comparing the stunted treatment group with the non-stunted comparison group. This comparison examines whether the intervention effectively remediated the initial disadvantage of stunted children.

4.1 Treatment Effect Analysis

Our analysis uses random assignment to identify treatment effects. Perfectly implemented randomized trials allow us to assess causal effects and are often called the "gold standard" for causal inference. However, most randomized trials are compromised in some way. In our case, we need to address 3 issues: (1) our small sample size, (2) imbalance of a few potentially key baseline variables between treatment and controls, and (3) a large number of outcomes and associated treatment effects. In what follows we will first describe our framework for causal inference and then describe the approaches used to address these three issues.

4.1.1 Randomization

The standard program evaluation model describes the observed outcome Y_i of participant i by $Y_i = D_i Y_i(1) + (1 - D_i)Y_i(0)$, where D_i denotes treatment effect assignment ($D_i = 1$ if treated, $D_i = 0$ otherwise) and $(Y_i(0), Y_i(1))$ are potential outcomes for individual i. Our objective is to estimate the average treatment effect, $E(Y_i(1) - Y_i(0))$. However, we are

unable to calculate the average treatment effect from ordinary observational data as we only observe either $(Y_i(1)|D_i = 1)$ or $(Y_i(0)|D_i = 0)$ for participant i.

We can estimate the simple difference-in-means, $E(Y_i(1)|D_i = 1) - E(Y_i(0)|D_i = 0)$. In observational samples, however, this difference is usually not a consistent estimator of the average treatment effect due to participants self-selecting into treatment. Selection bias occurs when the resulting distributions of participant characteristics differ between the treatment and control groups, and these differences are correlated with outcomes. Here, any difference in outcomes reflects a combination of both the underlying difference of unobserved characteristics and the treatment effect. Conditioning on observed characteristics may not fully control for all relevant sources of differences between treatment and controls.

Perfectly implemented random assignment solves the selection bias problem by inducing independence between the distribution of counter-factual outcomes $(Y_i(0), Y_i(1))$ and treatment status D_i conditional on the variables used in the randomization protocol. In the Jamaica Study, the randomization protocol first stratified the sample by age and sex, denoted by X, and then randomly assigned children within strata to each treatment group.

4.1.2 Permutation Tests

Our aim is to test the null hypothesis of no treatment effect. The small sample size of the Jamaican Study suggests that classical statistical inference methods that rely on large sample asymptotic theory to define the distribution of test statistics may be misleading. We address this problem by using non-parametric permutation tests as implemented in Heckman et al. (2010b). Permutation tests are valid in small samples because they are distribution free and do not rely on assumptions about the parametric sampling distribution.

Permutation tests are based on exchangeability properties generated directly from the randomization protocol. Exchangeability exploits the invariance of the joint distribution of (Y, D) under the null hypothesis. Random assignment guarantees that the vector of treatment assignments D is exchangeable within blocks of participants that share the same

values of X. Any swap of treatment status among participants who belong to the same pre-randomization strata of gender and age are just as likely to occur as the realized vector. Hence, under exchangeability we can permute treatment status between individuals with the same pre-program variables X and the joint distribution of (Y, D) will remain the same under the null hypothesis of no treatment effects.

Under exchangeability we can generate the exact distribution of a conditional test statistic $T(Y, D|X)$. Specifically, we generate the conditional distribution of the statistic given by all of the values that $T(Y, D)$ takes as we fully permute the elements of the vector of treatment status D within strata formed by X. The distribution generated in this way does not depend on any distributional or asymptotic assumption. We use this generated distribution for inference to test the null hypothesis that there is no difference between the treated and the untreated population means. We report a p-value that is simply the proportion of the test statistic values that are bigger than the ones computed using the actual data.

4.1.3 Baseline Imbalance

While randomization guarantees that any baseline variable Z is independent of the vector of treatment status D conditional on variables X used in the randomization protocol, the realization of baseline variables can turn out to be imbalanced across treatment groups. In the case of the Jamaica Study, three potentially important characteristics were not balanced at baseline. In order to control for potential bias, we estimate treatment effects by linear regression controlling for these variables when relevant for explaining outcomes.

The conditioning variables may also be used to increase the power of the statistical inference. Heckman et al. (2010b) address this problem by assuming partial linearity as suggested in Freedman and Lane (1983). Essentially, this involves permutation using the residuals from a multivariate linear regression. However, we would like to avoid linearity as this assumption is unlikely to hold for categorical outcomes and imposes a specific functional form.

Instead we employ a fully non-parametric technique that avoids invoking linearity assumptions by including the variables not balanced at baseline in X. This expands the number of strata blocks for the permutation tests described above. This is straightforward for discrete variables, but requires us to discretize continuous variables.

Increasing the number of blocks, however, comes at a cost as it may reduce the number of valid block permutations. This happens because it reduces the number of participants that share the same values of the conditioning variables. One can end up with blocks in which there are only treatments or only controls, rendering the observations in those blocks lost to the analysis.

Our conditioning set always includes the variables used in the randomization protocol plus the baseline variables that are imbalanced when their impact is statistically significant. Child age and sex as well as maternal employment and maternal education were constructed as discrete indicators. Weight-for-height is the only variable that we had to discretize. We chose the highest possible number of divisions that maximize the minimum number of observations in a block. This led to dividing the sample in three categories, those with a z-score higher than -1, those less than -1 but greater than -2, and those less than -2 in the standardized weight for height distribution. We lost no observations for permutation by following this rule.

Our method of inference is fully non-parametric and does not require any linearity assumption. It is theoretically exact. While the Freedman and Lane (1983) procedure is approximate, it often generates reasonably accurate inferences (Anderson and Robinson, 2001). Both the Freedman and Lane procedure and our procedure have drawbacks. The first imposes linearity, while the one we use requires us to discretize continuous variables when conditioning. While the results of our hypotheses do not change from what is obtained using the Freedman Lane procedure, our approach produces more precise estimates.

4.1.4 Accounting for Multiple Outcomes

The presence of multiple outcomes leads to the danger of arbitrarily selecting "statistically significant" outcomes where high values of test statistics arise by chance. Testing each hypothesis one at a time with a fixed significance increases the probability of a type-I error exponentially as the number of outcomes tested grows. We correct for this source of bias in inference by performing multiple hypothesis testing based on the Family-Wise Error Rate (FWER), which is the probability of rejecting at least one true null Hypothesis. We use the Stepdown algorithm proposed in Romano and Wolf (2005), which generates inference exhibiting strong FWER control. Associated with each outcome is a single null hypothesis of no treatment effect. We implement the Step-Down procedure for conceptually similar blocks of outcomes.

4.2 Catch Up Analysis

Our catch up analysis compares the non-stunted comparison group with the stunted treatment group. Despite being non-randomized, this analysis will employ inference using permutation tests. The intuition is that, under the null hypothesis, being non-stunted has no advantage with respect to the treatment stunted group. Exchanging stunted status within blocks should not change the distribution of outcomes. While inference in the treatment effect analysis tests whether the causal effect of the intervention is statistically significant compared to the control group, inference for the catch up group tests if the distribution of outcomes is statistically different between treatment and comparison groups. While the exchangeability criteria for inference between treatment and control groups comes from randomization, the exchangeability criteria for inference between non-stunted and treatment groups comes from the assumption of equality of outcome distributions under the null hypothesis.

One difference between the treatment and catch up analyses is attrition. As previously noted, attrition in the stunted sample is not a problem. However, attrition in the non-stunted

group appears to be selective. When attrition is not random, the observed sample may differ from the initial sample that was representative of the non-stunted population from poor urban areas. Hence, the catch up analysis would be a biased estimate of catch up to the non-stunted population.

We correct for attrition by using predicted probabilities of attrition to re-weight observed data. The predictions come from a logit model of attrition as a function of the baseline characteristics whose means are significantly difference between attrited and non-attrited. This procedure is termed Inverse Propensity Weighting (IPW).[22] This method gives more weight to those observations in the sample with a low propensity score for attrition correcting for the censoring effect of non-random attrition. We re-weight the data using variables measured at the onset of the intervention to correct for the potential bias of non-random attrition.

5 Migration

We begin by reporting the results of the impact of stimulation on migration to the U.S. or U.K. (Table 2). As discussed in greater detail below, migration has important implications for the earnings analysis. Migration is, itself, also an interesting outcome. The stimulation treatment may have improved skills enough so that beneficiaries or their families were encouraged to move overseas to take advantage of better education and labor market opportunities. Hence, migration might be an important pathway through which the intervention could have improved human capital and earnings outcomes.

We obtained migration status for the full baseline sample of both stunted and non-stunted children by filling missing values with information from relatives of study participants who dropped out of the sample. For the full stunted sample, 23 participants migrated and the treatment group was 10 percentage points (83%) more likely to migrate than the control

[22] See Robins et al. (1994).

group (p-value .08). Migration may have been a pathway to better education and earnings opportunities.

There is evidence of selective attrition of the migrants. We were able to locate and interview 14 out of the 23 (60%) migrants, a substantially lower share than the share of non-migrants that we were able to find and interview. Of the 14 migrants who were found and interviewed, 11 were in the treatment group and 3 were in the control group. This means that we found a much larger share of the treatment migrants than of the control migrants. This is apparent from comparing Row 1 of Table 2, which reports the results for the full baseline sample to Row 2 of Table 2, which reports the impact of treatment on migration using only the sample found in the follow-up. The third column of Table 2 indicates that the average migration rate for the observed control group is 6% compared to 12% for the full sample. The third column represents the conditional difference in average migration between the treatment and the control group. The treatment group is 15% more likely to be a migrant than the control group in the sample found at follow-up (p-value 0.02), but it is only 10% more likely to be a migrant in the full baseline sample (p-value 0.08).

This finding causes concern as it could lead to an overestimate of the impact of treatment on earnings. Migrants to the US and UK earn substantially more than those who remained in Jamaica. Due to the differential follow-up of migrants, control migrants are under-represented in the sample of 22 year-olds. Hence, the mean earnings of treatments might be higher than the mean earnings of controls even when the treatment effect is zero. We will address this concern with an additional set of analyses that (1) impute the earnings of the lost migrants, and (2) checks robustness by dropping migrants completely from the analysis. The later analysis produces a low bound estimate of the treatment effect as it does not allow migration to be a pathway to improved education and earnings.

6 Earnings Results

6.1 Measurement

We examine the impact of treatment on earnings histories as represented by earnings in the first job, in the last job and in the current job, as well as average monthly earnings over the lifetime. The current job is equal to the last job if the person is currently employed. We include the last job in order to reduce concerns over censoring as all but two of the study participants have had some labor market experience. Average monthly lifetime earnings are calculated as the ratio of total lifetime earnings divided by the number of months worked. All variables are expressed in terms of monthly earnings and are deflated to 2005 dollars using the CPI and then transformed into logs. Migrants' earnings are first deflated to 2005 using the local CPI, then converted to Jamaican Dollars using PPP adjusted exchange rates.[23]

One issue is that there is a significant portion of the sample that is both working and in school full time. Working, full-time students are likely to have lower earnings than non-students with the same education, and there are significantly more full time students working in the treatment group than in the control group. Hence, we likely underestimate the long-run earnings of those who are still in school. As a result, observed average earnings likely understate the long run earnings of the treatment more than the control group. This would imply that we are underestimating the long-run effects of treatment on earnings.

In order to assess the extent to which including working full time students in the sample underestimates the effect of treatment on long run earnings, we additionally analyze the impact of the program on samples restricted to workers in full time jobs and further restricted to workers in non-temporary permanent jobs. Restricting the sample to full time workers partially controls for this source of selection as many of the participants had part-time jobs while primarily attending school. We define full time as working at least 20 days per month.

[23]The PPP deflators are from the University of Pennsylvania, Center for International Comparison of Production, Income and Prices. As robustness checks, we also estimated the models by converting monetary amounts using the 2005 currency exchange rates and using the World Bank PPP exchange rates. Results in both cases are close to the estimates reported in the paper and are available upon request.

The sample of workers in non-temporary permanent jobs further omits students working in summer jobs that may have been full time. Non-temporary is defined as having a full time job working for 8 months a year or more.

6.2 Earnings Densities

We begin by examining the impact of the intervention on densities of different measures of log earnings. The panels of Figure 1 present the kernel density estimates of different earnings measures for the treatment and control group.[24] We display them separately for the first job, last job, current job, and average lifetime earnings. We also present them separately for all workers, full time workers, and non-temporary workers. The figures also report Komogorov-Smirnov test statistics for the null hypothesis that there is no difference in the treatment and control distributions.

The figures show that the densities of log earnings for the treatment group are shifted everywhere to the right of the control group densities for all comparisons with the single exception of the density of earnings on the first job for all workers. The Kolmogorov-Smirnov tests reveal that distributions of log earnings for the treatment group are significantly different than the distributions for the control group for almost all cases. The differences are greater when we restrict the sample to full time workers and even greater when we restrict the sample further to non-temporary workers.

6.3 Point Estimates

The estimated impacts on log earnings for the observed sample are reported in Table 3, in Panel I. The table reports the treatment effect, the conditional p-value for the hypothesis of no treatment effect taken in isolation, and the p-value obtained from the Step-Down

[24]The kernels of Figure 1 display the distribution of log earnings for the overall sample. We evaluate Epanechnikov kernels using a bandwidth that minimizes mean integrated squared error for Gaussian data.

procedure. In doing the Step-Down procedure, we group together the outcomes in each block of rows in each panel separately.

The results show that the treatment group had significantly higher earnings over the entire tenure in labor market and for all job types including part-time, full time and permanent jobs, as well as for log of average earnings. Average monthly lifetime earnings are 49% higher for all jobs and 60% higher for the treatment than for the control group when considering only non-temporary full time jobs.[25,26] Differences in earnings in the first, last and current jobs show similar magnitudes. However, the impact is substantially larger for full-time and even larger for full-time permanent (non-temporary) jobs.

6.4 Attrition of Migrants

We address potential bias from differential attrition of migrants by imputing earnings for the missing observations. Imputing the missing observations re-weights the data so that the treatment and control groups of migrants are no longer under- or over-represented in the sample. In order to minimize the amount of data imputed, we impute missing earnings only for migrants who were lost. As a result, we impute earnings for only 9 observations. We replace missing earnings values with predicted log earnings from an OLS regression on treatment, gender and migration status.

The results, reported in Panel II of Table 3, show that the impact on earnings remains large and statistically significant for the sample with imputed earnings. Not surprisingly, however, the point estimates are slightly lower. In this case, the estimated impact on the average monthly lifetime earnings for all workers is 42%, and for non-temporary workers is 49%. Again, we find similar effects of the adjustment on the magnitude of impact in the first job, last job and current job. The estimated impact increases when we restrict the sample

[25]We convert the estimated treatment effects on log earnings from Table 3 into percent change with the following transformation $exp(\beta) - 1$, where β denotes the treatment effect estimate.

[26]The kernels of Figure 1 suggest that the results are not driven by outliers. We also examine the influence of outliers by excluding values that are more extreme than the 5th and the 95th percentile (see Appendix Table A.9). Trimming leads to slightly smaller but still statistically significant point estimates.

to full time and non-temporary workers.

As a robustness check, we re-estimate the models excluding the migrants (Table 3, Panel III). Completely excluding all of the migrants is a very conservative approach because it rules out migration as a mechanism for obtaining higher earnings and it also significantly reduces sample size. Excluding migrants provides a lower-bound estimate of the impact on earnings. When we exclude migrants, we find that estimated effect sizes fall slightly, but still remain highly significant especially for full-time and non-temporary workers. The estimates excluding migrants show 38% higher earnings for the treatment group for all jobs and 45% for full time non-temporary jobs.

6.5 Employment and Labor Force Participation

Censoring is a concern with our estimates of the impact of treatment on current earnings for all workers because we only observe the earnings of those employed who are in the labor force. However, treatment does not appear to affect employment or labor force participation (Table 4), implying negligible bias from censoring in our results for current earnings.

6.6 Catch Up in Earnings

The results reported thus far indicate a substantial and significant impact of treatment on earnings. One important question, however, is whether the stimulation intervention was strong enough for the earnings of the treatment group to catch up to a population that was not stunted in childhood. This question is at the heart of the remediation issue: can early childhood intervention remediate initial disadvantage? We answer this question by comparing the earnings of the stunted population with the earnings of the non-stunted comparison group.

Overall, we find that the treatment group caught up with the comparison group on all measures of earnings, while the control group remained behind. Table 5 compares the non-stunted comparison group with the stunted treatment group using IPW weights to correct

for the higher attrition among the non-stunted group. Table 5 presents results in the same fashion as Table 3: Panel I examines the observed sample, Panel II uses imputed values for missing data, and Panel III focus on data for non-migrants only. The p-values represent one-sided tests for the hypothesis that the differences between the two groups are null versus the alternative hypothesis of the non-stunted group has higher earnings. The conditional differences in log-earnings between the non-stunted group and the stunted treatment group are never statistically significant and average around zero.

The panels of Figure 2 displays kernel density estimates of the densities of earnings for the non-stunted comparison group and the stunted treatment group. The figures generally show little separation between the earnings densities for the two groups, as confirmed the Kolmogorov-Smirnov tests. These results are consistent with the findings reported in Table 5.

In contrast, the stunted control group remains behind. Table 6 presents the mean differences between the non-stunted comparison group and the stunted control group. The table shows that the non-stunted comparison group consistently earns more than the stunted control group, with most differences in mean earnings being statistically significant.

7 Pathways to Earnings

7.1 Parental Investment

The stimulation intervention was designed to improve the maternal-child interaction, i.e. the quality of parenting. We begin by examining the extent to which treatment resulted in more maternal investment in stimulation at home during the experimental period when the children are very young. Although we cannot attribute a causal link between the increase in the measures of the quality of the home environment and the outcomes, these results suggest a possible mechanism.

We analyze the effects of treatment on a modified version of the Caldwell index of stimulation of the home, the infant toddler HOME inventory (Caldwell, 1967; Caldwell and Bradley,

1984). The HOME score captures the quality of parental interaction and investment in the children through the observation of the home environment.[27] It includes six main domains: emotional and verbal responsivity of the caregiver, avoidance of restriction and punishment, organization of the environment, provision of play materials, parental involvement with the child and opportunities for variety in daily stimulation.

The results show that the intervention did indeed increase the HOME score (see Table 7). At baseline there was no difference between treatment and control groups, but the HOME score of the stunted group was significantly lower than the HOME score of the non-stunted group. At the end of trial, however, the HOME score of the treated is significantly higher than that of the control group, and the HOME score of the treatment caught up to the HOME score of the non-stunted group.[28] In results available from the authors on request, we find that stimulation in the home was not different between the treated and control groups at 7 and 11 years old. These findings are consistent with the hypothesis that treatment improved the quality of maternal-child interaction in early life and this difference dissipated later. Hence, any impacts on schooling and earnings later in life are likely due to these investments made early in life.

7.2 Education

A key determinant of labor market success is education. Schooling in Jamaica comprises primary school, grades 1-6; junior secondary, grades 7-9; and senior secondary grades 10-13. At the end of grade 11, students take exams called the CXC exams which are similar to the British O Levels. Most students leave school after grade 11. At the end of grade 13 students take advanced level exams (CAPE) for college entry.

Overall the results show that treatment is associated with substantially more education (Table 8, Panel A). The treatment group has completed significantly more schooling than the

[27] Previous studies have found HOME to be highly correlated with cognitive, social and motor skills. For example, see Bradley (1993), Bradley et al. (1989), Grantham-McGregor et al. (1997).

[28] Moreover, there is no difference in the impact of HOME for boys versus girls.

control group for all indicators of educational attainment. They are three times more likely to have had some college education than the control group. Members of the treatment group have passed more CXC and CAPE exams than the control group (Table 8, Panel B). The treatment group also has about 0.61 year more of schooling than the control group. However, this is clearly a lower bound estimate as the impact of treatment on school participation is strongly positive. Persons in the treatment group are twice as likely to be in school and almost three times more likely to be in full time school (Table 8, Panel A). Since the earnings of those in school will likely increase as they fully enter that labor force, we likely underestimate the impact of treatment on earnings.

7.3 Cognitive and Psychosocial Skills

Cognitive and psychosocial skills are important determinants of labor market outcomes.[29] We examine the impact of treatment on psychosocial skills at age 18, a critical age for labor market decisions. The survey at age 18 collected multiple psychometric scales of cognitive and psychosocial skills. The cognitive scales included the WRAT math, WRAT reading comprehension, Picture Peabody Verbal Test, Verbal Analogies, Raven matrices, and WAIS full-scale IQ tests.[30] For psychosocial skills, available scales include the Conners' scale for oppositional behavior, the Conners' scale for inattention, and the Conner's scale for hyperactivity, as well as a self-esteem scale, an anxiety scale and a depression scale.

We use factor analysis to aggregate the scales above, extracting three factors: one for cognitive skills, and two for psychosocial skills, which represent externalizing behavior and internalizing behavior. All three types of skills have proven to have independent effects on earnings (Almlund et al., 2011a) and have been used as outcomes for the evaluation of early childhood policies (Heckman, Pinto, and Savelyev, 2013). Externalizing behavior was measured using the Conners' scales described above, while Internalizing behavior was measured

[29]See Heckman, Pinto, and Savelyev, 2013; Heckman, Stixrud, and Urzúa, 2006; Almlund, Duckworth, Heckman, and Kautz, 2011b, and Heckman and Kautz, 2012.

[30]See Walker et al. (2005) for results of impact on the individual scales using large sample inference methods

with self-esteem, anxiety and depression scales. Exploratory factor analysis confirmed that there were three factors. We then used confirmatory factor analysis to recover the factors using the Bartlett's method to extract the factor scores.[31] Measurements were dedicated to a factor, so that each measure only had positive loadings on one factor. However, factors are allowed to be correlated. The factors are measured in standard deviations and all are recoded so that a higher level of the transformed measure is more desirable.

The results reported in Table 8, Panel C, present the impact of stimulation treatment on aggregated factors of cognitive and psychosocial skills. The coefficient on the factors can be interpreted as the impact of treatment measured on standard deviations of the scale. Overall we find strong and statistically significant effects of treatment on all measures of cognition, as well as on internalizing behavior. Plots of the densities of these factors classified by treatment and control status (Figure 3) show a substantial impact of treatment and suggest that the treatment was particularly effective for those in the upper part of the distribution. Overall the results are consistent with strong positive effects of treatment on earnings and are in line with the recent literature on the importance of both cognitive and psychosocial skills for earnings (see Borghans et al., 2008 and Almlund et al., 2011a).

7.4 Catch Up in Education and Skills

We return to the question of whether the impact of stimulation helped the stunted group catch up with the non-stunted comparison group. The catch-up analysis shows that the treated group did indeed catch up with the non-stunted group for educational outcomes (Panel I of Table 9). The results confirm the findings from Walker et al. (2005) in that the treatment group did not completely catch up to the non-stunted group in cognitive skills. However, the treatment group completely caught up in terms of education and both

[31]See Appendix B for details about our factor analysis. Exploratory factor analysis confirmed that a three factor model was the most suitable to explain the measurements with each measure loading clearly on one and only one factor. See results in Table B.1. Kaiser and Scree tests also selected a three factor model. This finding was confirmed by confirmatory factor analysis fit statistics (Chi square test, RMSEA, CFI, AIC and BIC) with better performance for a three-factor model compared to a simpler two-factor model.

internalizing and externalizing behavior. Figure 4 shows that the treatment and comparison group densities overlap for the internalizing and externalizing factor, but not for the cognitive factor. The comparison group density is shifted to the right of the treatment group density for the cognitive factor. The results reported in Panel II of Table 9 show that the control group did not catch up with the comparison group.

8 Gender Differences

The literature on early childhood interventions shows that ECD treatment effects can differ substantially by gender (e.g. Heckman et al., 2010a, 2013). In this section we investigate the gender-specific effects of the Jamaican intervention. We interpret the results of these analyses with caution as the study was not originally designed nor powered for this purpose.

Table 10 reports the impact of treatment on log earnings separately for males and females.[32] We find statistically significant effects of stimulation on earnings for both males and females. While the point estimates are in general somewhat larger for females than for males, tests for equality cannot reject the hypothesis that the impact on earnings is equal for males and females.[33]

Not only do we find significant effects on earnings for both stunted males and females, but we also find that both males and females in the stunted treatment group catch up to the earnings of the non–stunted comparison group (Table 11, Panel I). The point estimates of the differences are generally close to zero and are not significantly different from zero. However, the earnings of the stunted control group are also not significantly different from those of the non–stunted comparison group. In this case, the point estimates are positive, indicating that the stunted control group earns consistently less than the non–stunted comparison group,

[32]We only display the results for the sample with imputed earnings for lost migrants. However, the results do not differ substantively for either the observed sample or the non-migrants only sample. These results are available upon request.

[33]See Table A.7 in Appendix A. This table reports estimates for female treatment effects and the difference of treatment effects for males versus females. The table also presents inference on gender difference of treatment effects.

but these differences are only statistically significant for earnings in full time last job and current job for females, and for earnings in first jobs for males (Table 11, Panel II).

We also compare gender differences in the determinants of earnings for treated and control groups (Table 12). All of the point estimates are generally positive. However, a couple of important differences emerge. For males (Table 12, Panel II) there are statistically significant treatment effects on cognitive ability and on the probability of being expelled from school, while for females (Table 12, Panel I) there are statistically significant treatment effects on passing exams and on reduction on externalizing and internalizing behavior. However, the hypothesis of equality of the treatment effects for males and females cannot be rejected for these outcomes in particular and in general for 10 out of 12 of these outcomes.[34]

The female stunted treatment group caught up with the non–stunted comparison in all of the educational and skills outcomes. However, males did not catch up completely in exams (Table 13, Panel I). While the female stunted control group did not catch up to the female comparison group, the male control group appears to have caught up to the male comparison group in terms of educational outcomes and psychosocial skills (Table 13, Panel II).

9 Conclusions

This is the first study to experimentally evaluate the long-term impact of early childhood stimulation on economic outcomes in a low income country. Twenty years after the intervention was conducted, we find that the average earnings of the stimulation group are approximately 42% higher than those of the control group. These findings show that simple psychosocial stimulation in very early childhood in disadvantaged settings can have a substantial effect on labor market outcomes. The magnitude of the estimated treatment effects can be put into perspective when the outcomes for the treated are compared to those for a non-stunted comparison group. The stunted children who received the stimulation intervention caught up to the earnings of a non-stunted comparison group. These results imply

[34]See Table A.7.

that stimulation interventions very early in life can compensate for developmental delays and thereby reduce inequality later in life. The estimated impacts found for Jamaica are substantially larger than the impacts reported for the US–based interventions. Early Childhood Development may be an especially effective strategy for improving long-term outcomes of disadvantaged children in developing countries.

References

Almlund, M., A. Duckworth, J. J. Heckman, and T. Kautz (2011a). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education*, Volume 4, pp. 1–181. Amsterdam: Elsevier.

Almlund, M., A. Duckworth, J. J. Heckman, and T. Kautz (2011b, February). Personality psychology and economics. *IZA Discussion Paper* (No. 5500). `http://ftp.iza.org/dp5500.pdf`.

Almond, D. and J. Currie (2011). Human capital development before age five. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4B, Chapter 15, pp. 1315–1486. North Holland: Elsevier.

Almond, D., L. Edlund, H. Li, and J. Zhang (2007, September). Long-term effects of the 1959-1961 China famine: Mainland China and Hong Kong. Working Paper 13384, National Bureau of Economic Research.

Anderson, M. J. and J. Robinson (2001, March). Permutation tests for linear models. *The Australian and New Zealand Journal of Statistics 43*(1), 75–88.

Aughinbaugh, A. (2001). Does Head Start yield long-term benefits? *Journal of Human Resources 36*(4), 641–665.

Bleakley, H. (2007, February). Disease and development: Evidence from hookworm eradication in the American South. *Quarterly Journal of Economics 122*(1), 73–117.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008, Feburary). The economics and psychology of personality traits. *IZA Discussion Paper* (3333). `http://ftp.iza.org/dp3333.pdf`.

Bradley, R. H. (1993). Children's home environments, health, behavior, and intervention efforts: a review using the HOME inventory as a marker measure. *Genetic Social and General Psychology Monographs 119*(4), 437–490.

Bradley, R. H., B. M. Caldwell, S. L. Rock, C. T. Ramey, K. E. Barnard, C. Gray, M. A. Hammond, S. Mitchell, A. W. Gottfried, L. Siegel, and D. Johnson (1989). Home environment and cognitive development in the first 3 years of life: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology 25*(2), 217–235.

Caldwell, B. M. (1967). Descriptive evaluations of child development and of developmental settings. *Pediatrics 40*(1), 46–54.

Caldwell, B. M. and R. H. Bradley (1984). *HOME observation for measurement of the environment*. Little Rock, AR: University of Arkansas at Little Rock.

Campbell, F., G. Conti, J. Heckman, S. Moon, and R. Pinto (2012). The long-term health effects of early childhood interventions. Under review, *Economic Journal*.

Campbell, F. A., E. P. Pungello, M. Burchinal, K. Kainz, Y. Pan, B. H. Wasik, O. A. Barbarin, J. J. Sparling, and C. T. Ramey (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology 48*(4), 1033–1043.

Campbell, F. A., C. T. Ramey, E. Pungello, J. Sparling, and S. Miller-Johnson (2002). Early childhood education: Young adult outcomes from the abecedarian project. *Applied Developmental Science 6*(1), 42–57.

Carneiro, P. and J. J. Heckman (2003). Human capital policy. In J. J. Heckman, A. B. Krueger, and B. M. Friedman (Eds.), *Inequality in America: What Role for Human Capital Policies?*, pp. 77–239. Cambridge, MA: MIT Press.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Engle, P. L., M. M. Black, J. R. Behrman, M. Cabral de Mello, P. J. Gertler, L. Kapiriri, R. Martorell, M. Eming Young, and The International Child Development Steering Group (2007, January). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet 369*(9557), 229–242.

Engle, P. L., L. C. H. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M. Cabral de Mello, M. Hidrobo, N. Ulkuer, I. Ertem, and S. Iltus (2011, October). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet 378*(9799), 1339–1353.

Fernald, L., P. Kariger, M. Hidrobo, and P. Gertler (2012, October). Socio-economic gradients in child development in very young children: Evidence from india, indonesia, peru and senegal. *Proceedings of the National Academy of Sciences (Suplement 2) 109*, 17273–17280.

Fernald, L. C., 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.

Freedman, D. and D. Lane (1983, October). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics 1*(4), 292–298.

Garces, E., D. Thomas, and J. Currie (2002, September). Longer-term effects of Head Start. *American Economic Review 92*(4), 999–1012.

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

Grantham-McGregor, S., W. Schofield, and C. Powell (1987). Development of severely malnourished children who received psychosocial stimulation: Six-year follow-up. *Pediatrics 79*(2), 247–254.

Grantham-McGregor, S., S. Walker, S. Chang, and C. Powell (1997). Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children. *American Journal of Clinical Nutrition 66*(2), 247–253.

Grantham-McGregor, S. M., C. A. Powell, S. P. Walker, and J. H. Himes (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican study. *The Lancet 338*(8758), 1–5.

Heckman, J. J. (2000, March). Policies to foster human capital. *Research in Economics 54*(1), 3–56.

Heckman, J. J. (2008, July). Schools, skills and synapses. *Economic Inquiry 46*(3), 289–324.

Heckman, J. J. and T. Kautz (2012, August). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464. Adam Smith Lecture.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010a, August). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010b, February). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics 94*(1-2), 114–128.

Heckman, J. J., R. Pinto, and P. A. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. Unpublished manuscript, University of Chicago, Department of Economics (first draft, 2008). Forthcoming, *American Economic Review*.

Heckman, J. J., J. Stixrud, and S. Urzúa (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Hoddinott, J., J. A. Maluccio, J. R. Behrman, R. Flores, and R. Martorell (2008). Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults. *The Lancet 371*(9610), 411–416.

Huttenlocher, P. (1979). Synaptic density in human frontal cortexdevelopmental changes and effects of aging. *Brain research 163*(2).

Huttenlocher, P. R. (2002). *Neural plasticity: The effects of environment on the development of the cerebral cortex*. Cambridge, MA: Harvard University Press.

Knudsen, E. I., J. J. Heckman, J. Cameron, and J. P. Shonkoff (2006, July). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences 103*(27), 10155–10162.

Maccini, S. L. and D. Yang (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American Economic Review 99*(3), 1006–1026.

Maluccio, J. A., J. Hoddinott, J. R. Behrman, R. Martorell, A. R. Quisumbing, and A. D. Stein (2009). The impact of improving nutrition during early childhood on education among Guatemalan adults. *Economic Journal 119*(537), 734–763.

Palmer, F. H. (1971). *Concept training curriculum for children ages two to five*. Stony Brook, NY: State University of New York at Stony Brook.

Paxson, C. and N. Schady (2007, Summer). Cognitive development among young children in Ecuador: The roles of wealth, health, and parenting. *Journal of Human Resources 42*(1), 49–84.

Powell, C. and S. Grantham-McGregor (1989). Home visiting of varying frequency and child development. *Pediatrics 84*(1), 157–164.

Reynolds, A. J., S.-R. Ou, and J. W. Topitzes (2004, September–October). Paths of effects of early childhood interventions on educational attainment and deliquency: A confirmatory analysis of the Chicago Parent-Child Centers. *Child Development 75*(5), 1299–1328.

Reynolds, A. J., J. A. Temple, S.-R. Ou, I. A. Arteaga, and B. A. B. White (2011, July). School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups. *Science 333*(6040), 360–364.

Reynolds, A. J., J. A. Temple, S.-R. Ou, D. L. Robertson, J. P. Mersky, J. W. Topitzes, and M. D. Niles (2007, August). Effects of a school-based, early childhood intervention on adult health and well-being: A 19-year follow-up of low-income families. *Archives of Pediatrics and Adolescent Medicine 161*(8), 730–739.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association 89*(427), 846–866.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Thompson, R. A. and C. A. Nelson (2001, January). Developmental science and the media: Early brain development. *American Psychologist 56*(1), 5–15.

van den Berg, G. J., M. Lindeboom, and F. Portrait (2006, March). Economic conditions early in life and individual mortality. *American Economic Review 96*(1), 290–302.

Walker, S., S. Grantham-McGregor, C. Powell, J. Himes, and D. Simeon (1992). Morbidity and the growth of stunted and nonstunted children, and the effect of supplementation. *American Journal of Clinical Nutrition 56*(3), 504–510.

Walker, S., C. Powell, and S. Grantham-McGregor (1990). Dietary intakes and activity levels of stunted and non-stunted children in Kingston, Jamaica. Part 1. Dietary intakes. *European Journal of Clinical Nutrition 44*(7), 527–534.

Walker, S. P., S. M. Chang, C. A. Powell, and S. M. 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, M. Vera-Hernández, and S. Grantham-McGregor (2011). Early childhood stimulation benefits adult competence and reduces violent behavior. *Pediatrics 127*(5), 849–857.

Walker, S. P., S. M. Grantham-McGregor, J. H. Himes, C. A. Powell, and S. M. Chang (1996). Early childhood supplementation does not benefit the long-term growth of stunted children in Jamaica. *Journal of Nutrition 126*(12), 3017–3024.

Walker, S. P., S. M. Grantham-McGregor, C. A. Powell, and S. M. Chang (2000, July). Effects of growth restriction in early childhood on growth, IQ, and cognition at age 11 to 12 years and the benefits of nutritional supplementation and psychosocial stimulation. *Journal of Pediatrics 137*(1), 36–41.

Walker, S. P., C. A. Powell, S. M. Grantham-McGregor, J. H. Himes, and S. M. Chang (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican study. *American Journal of Clinical Nutrition 54*(4), 642–648.

Walker, S. P., T. D. Wachs, J. M. Gardner, B. Lozoff, G. A. Wasserman, E. Pollitt, J. A. Carter, and The International Child Development Steering Group (2007, January). Child development: Risk factors for adverse outcomes in developing countries. *The Lancet 369*(9556), 145–157.

Table 1: Baseline (1986) Descriptive Statistics for Stunted Experimental Sample

	Sample Size	**Control Mean**	**Treatment Mean**	**Difference in Means**	**Single p-value**
A. Parental Characteristics					
Mother present	105	0.96	0.94	-0.02	1.00
Mother/guardian's age (years)	105	24.4	25.8	1.41	0.28
Mother /guardian employed	105	0.15	0.32	0.17	**0.05**
Mother/guardian school >= 9th grade	105	0.21	0.05	-0.16	**0.02**
Mother/guardian's PPVT	105	84.9	86.8	1.91	0.64
Mothers/guardian's height (cm)	103	159.3	159.4	0.06	0.96
Father present	105	0.46	0.45	-0.01	1.00
HOME score on enrolment	105	17.1	16.02	-1.08	0.22
Housing index	105	7.56	7.17	-0.39	0.20
B. Child Characteristics					
Age (years)	105	1.55	1.55	0.00	1.00
Male	105	0.56	0.53	-0.03	0.85
Child's birth order	105	2.98	3.38	0.40	0.38
Birth Weight $<$ 2500 grams	104	0.19	0.25	0.06	0.58
Head Circumference (cm)	105	46.2	45.9	-0.27	0.37
Daily Calories Consumed	105	1006	912.9	-93.11	0.31
Daily Protein Consumed (grams)	105	27.0	26.96	-0.04	1.00
Griffith Developmental Quotient	105	97.1	99.3	2.21	0.21
Height for Age z-Score	105	-2.87	-3.00	-0.13	0.28
Weight for Height z-Score	105	-0.87	-1.18	-0.31	**0.02**

Notes: This table reports and compares arithmetic means of variables of interest for the stunted treatment and control groups at baseline (1986) for the sample found in 2008. The p-values reported in the last column are for two-sided block permutation tests of the null hypotheses that the difference in means between treatment and control groups are zero. The permutation blocks are child's age and sex. Variable definitions include: PPVT denotes the raw score from Peabody Picture Vocabulary Test (Dunn and Dunn, 1981), HOME denotes the raw score from the HOME environment test (Caldwell, 1967), and Griffith Development Quotient reports the raw score for this test (Griffiths, 1954; Griffiths 1970).

Table 2: Impact of Stimulation Treatment on Migration and Catch-up in Migration.

A. Treatment Effect	**Sample Size**	**Control Group Mean**	**Treatment Effect**	**Single p-value**
Full baseline sample	127	0.12	0.10	**0.08**
Sample found at follow-up	105	0.06	0.15	**0.02**
B. Treatment Group Catch-Up	**Sample Size**	**Treatment Group Mean**	**Difference Comparison vs. Treated**	**Single p-value**
Full baseline sample	141	0.19	0.01	0.47
Sample found at follow-up	115	0.16	-0.06	0.83

Notes: Panel A reports the estimated impact of treatment on the probability of migration out of Jamaica to another country. The first row presents the results for the full sample enrolled in the experiment at baseline in 1986. The second row reports the results for the sample found at follow-up in 2008. The treatment effects are estimated by linear regression and are interpreted as the differences in the migration rates of the stunted treatment and stunted control groups conditional on baseline values of child age, gender, weight-for-height z-score, maternal employment, and maternal education. Our p-values are for one-sided block permutation tests of the null hypothesis of no treatment effect (Single p-value). Permutation blocks are based on the conditioning variables used in the treatment effect regressions. Panel B reports the difference in the probability of migration of the weighted non-stunted comparison group and the stunted treatment group. Again, the first row reports the results for the full sample and the second row for the sample found at follow-up in 2008. The comparison group observations are weighted using Inverse Probability Weights (IPW) to correct for attrition. Our p-values are for one-sided block permutation tests of the null hypothesis of complete catch-up (Single p-value). Permutation blocks are based on gender.

Table 3: Impact of Stimulation Treatment on Log Earnings

	I. Observed Sample			II. Imputed Missing Values			III. Non-Migrants Only		
	Treatment Effect	**Single p-value**	**Stepdown p-value**	**Treatment Effect**	**Single p-value**	**Stepdown p-value**	**Treatment Effect**	**Single p-value**	**Stepdown p-value**
A. First Job									
All	0.27	0.11	0.11	0.21	0.16	0.16	0.25	**0.05**	**0.05**
Full Time	0.35	**0.04**	**0.06**	0.27	**0.07**	**0.10**	0.28	**0.04**	**0.06**
Non-Temporary	0.53	**0.01**	**0.03**	0.45	**0.02**	**0.04**	0.46	**0.01**	**0.03**
B. Last Job									
All	0.27	**0.06**	**0.06**	0.23	**0.09**	**0.09**	0.15	0.19	0.19
Full Time	0.40	**0.00**	**0.01**	0.36	**0.01**	**0.01**	0.29	**0.02**	**0.03**
Non-Temporary	0.50	**0.00**	**0.00**	0.45	**0.00**	**0.00**	0.40	**0.01**	**0.02**
C. Current Job									
All	0.27	**0.09**	**0.09**	0.21	0.13	0.13	0.12	0.26	0.26
Full Time	0.43	**0.01**	**0.02**	0.37	**0.01**	**0.03**	0.35	**0.02**	**0.04**
Non-Temporary	0.44	**0.01**	**0.02**	0.34	**0.02**	**0.04**	0.40	**0.01**	**0.03**
D. Average Earnings									
All	0.40	**0.01**	**0.01**	0.35	**0.00**	**0.01**	0.32	**0.01**	**0.02**
Full Time	0.34	**0.01**	**0.01**	0.28	**0.01**	**0.01**	0.22	**0.03**	**0.03**
Non-Temporary	0.47	**0.00**	**0.01**	0.40	**0.01**	**0.01**	0.37	**0.01**	**0.02**

Notes: This table reports the estimated impact of treatment on log monthly earnings for 3 samples as indicated by the following column panels: (I) the observed sample, (II) the observed sample with imputations for the earnings of missing migrants, and (III) the observed sample without migrants. In each sample, treatment effects are reported for the following jobs as indicated by the row blocks: (A) First Job, (B) Last Job, (C) Current Job, and (D) Average Lifetime Earning over all jobs. Within each type of job, results are reported for the following types of workers as indicated by the rows: All workers, Full Time Workers, and full time Non-Temporary workers. The treatment effects are estimated by linear regression and are interpreted as the differences in the means of Log Earnings between the stunted treatment and stunted control groups conditional on baseline values of child age, gender, weight-for-height z-score, maternal employment, and maternal education. Our p-values are for one-sided block permutation tests of the null hypothesis of no treatment effect (Single p-value) and multiple hypotheses (Stepdown p-value) of no treatment. Permutation blocks are based on the conditioning variables used in the treatment effect regressions.

Table 4: Impact of Stimulation Treatment on Employment and Labor Force Participation

	Sample Size	**Control Mean**	**Treatment Effect**	**Single p-value**	**Stepdown p-value**
Current Employment Status					
Employed	105	0.65	0.12	**0.08**	0.16
Employed Full Time	105	0.58	0.03	0.31	0.31
Employed in Non-Temporary Job	105	0.34	0.07	0.18	0.26
Looking For Work	99	0.27	-0.09	0.17	0.34

Notes: This table reports the estimated impact of treatment on measures of employment status including currently employed, employed in a full time job, employed in a non-temporary job and looking for work. The treatment effects are estimated by linear regression and are interpreted as the differences in the means of employment outcomes between the stunted treatment and stunted control groups conditional on baseline values of child age, gender, weight-for-height z-score, maternal employment, and maternal education. Our p-values are for one-sided block permutation tests of the null hypothesis of no treatment effect (Single p-value) and multiple hypotheses (Stepdown p-value) of no treatment. Permutation blocks are based on the conditioning variables used in the treatment effect regressions.

Table 5: Treatment Group Catchup in Log Earnings

	I. Observed Sample			II. Imputed Missing Values			III. Non-Migrants Only		
	Difference Comp. vs. Treated	Single p-value	Stepdown p-value	Difference Comp. vs. Treated	Single p-value	Stepdown p-value	Difference Comp. vs. Treated	Single p-value	Stepdown p-value
A. First Job									
All	0.11	0.16	0.24	0.15	**0.09**	0.14	0.11	0.17	0.26
Full Time	0.16	0.11	0.19	0.19	**0.07**	0.12	0.18	**0.07**	0.14
Non-Temporary	-0.12	0.78	0.78	-0.11	0.79	0.79	-0.02	0.56	0.56
B. Last Job									
All	0.12	0.26	0.26	0.14	0.21	0.34	0.23	0.12	0.21
Full Time	-0.05	0.62	0.70	0.00	0.49	0.57	-0.01	0.50	0.59
Non-Temporary	-0.23	0.93	0.93	-0.17	0.88	0.88	-0.14	0.82	0.82
C. Current Job									
All	-0.04	0.59	0.74	0.05	0.40	0.54	0.08	0.36	0.51
Full Time	-0.25	0.88	0.92	-0.15	0.78	0.84	-0.20	0.82	0.88
Non-Temporary	-0.35	0.94	0.94	-0.26	0.91	0.91	-0.32	0.92	0.92
D. Average Earnings									
All	-0.04	0.64	0.74	0.00	0.52	0.63	0.03	0.43	0.55
Full Time	-0.07	0.69	0.74	-0.03	0.60	0.65	0.00	0.49	0.55
Non-Temporary	-0.21	0.91	0.91	-0.16	0.87	0.87	-0.11	0.77	0.77

Notes: This table compares the earnings of the stunted treatment group to the earnings of the non-stunted comparison group in terms of log monthly earnings for 3 samples as indicated by the following column panels: (I) the observed sample, (II) the observed sample with imputations for the earnings of missing migrants, and (III) the observed sample without migrants. In each sample, catch-up is reported for the following jobs as indicated by the row blocks: (A) First Job, (B) Last Job, (C) Current Job, and (D) Average Lifetime Earning over all jobs. Within each type of job category, results are reported for the following types of workers as indicated by the rows: All workers, Full Time Workers, and full time Non-Temporary workers. Catch-up is estimated as the difference in the means of Log Earnings between the weighted non-stunted comparison group and the stunted treatment group. The comparison group observations are weighted using Inverse Probability Weights (IPW) to correct for attrition. Our p-values are for one-sided block permutation tests of the null hypothesis of complete catch-up (Single p-value) and multiple hypotheses (Stepdown p-value) of complete catch-up. Permutation blocks are based on gender.

Table 6: Control Group Catchup in Log Earnings

	I. Observed Sample			II. Imputed Missing Values			III. Non-Migrants Only		
	Difference Comp. vs. Control	Single p-value	Stepdown p-value	Difference Comp. vs. Control	Single p-value	Stepdown p-value	Difference Comp. vs. Control	Single p-value	Stepdown p-value
A. First Job									
All	0.23	**0.04**	**0.04**	0.24	**0.02**	**0.04**	0.25	**0.03**	**0.03**
Full Time	0.31	**0.01**	**0.02**	0.31	**0.01**	**0.01**	0.32	**0.01**	**0.02**
Non-Temporary	0.26	**0.04**	**0.06**	0.23	**0.05**	**0.05**	0.28	**0.02**	**0.04**
Last Job									
All	0.29	**0.03**	**0.07**	0.28	**0.03**	**0.05**	0.26	**0.05**	0.10
Full Time	0.26	**0.03**	**0.05**	0.28	**0.02**	**0.04**	0.22	**0.06**	**0.09**
Non-Temporary	0.16	0.19	0.19	0.18	0.14	0.14	0.12	0.25	0.25
B. Current Job									
All	0.18	0.18	0.30	0.21	0.12	0.21	0.12	0.27	0.42
Full Time	0.14	0.21	0.28	0.17	0.16	0.22	0.06	0.36	0.47
Non-Temporary	0.02	0.50	0.50	0.03	0.47	0.47	-0.07	0.65	0.65
D. Average Earnings									
All	0.25	**0.03**	**0.05**	0.27	**0.02**	**0.03**	0.23	**0.04**	**0.07**
Full Time	0.17	**0.08**	0.11	0.18	**0.08**	**0.09**	0.15	0.11	0.14
Non-Temporary	0.16	0.16	0.16	0.17	0.13	0.13	0.14	0.18	0.18

Notes: This table compares the earnings of the stunted control group catch up to the earnings of the non-stunted comparison group in terms of log monthly earnings for 3 samples as indicated by the following column panels: (I) the observed sample, (II) the observed sample with imputations for the earnings of missing migrants, and (III) the observed sample without migrants. In each sample, catch-up is reported for the following jobs as indicated by the row blocks: (A) First Job, (B) Last Job, (C) Current Job, and (D) Average Lifetime Earning over all jobs. Within each type of job category, results are reported for the following types of workers as indicated by the rows: All workers, Full Time Workers, and full time Non-Temporary workers. Catch-up is estimated as the difference in the means of Log Earnings between the weighted non-stunted comparison group and the stunted treatment group. The comparison group observations are weighted using Inverse Probability Weights (IPW) to correct for attrition. Our p-values are for one-sided block permutation tests of the null hypothesis of complete catch-up (Single p-value) and multiple hypotheses (Stepdown p-value) of complete catch-up. Permutation blocks are based on gender.

Table 7: Impact of Treatment on HOME Scores: Treatment vs. Control (A) and Comparison Group vs. Treatment (B)

A. Treatment Effect	**Sample Size**	**Control Group Mean**	**Treatment Effect**	**Single p-value**
HOME at enrollment	127	16.64	-0.51	0.51
HOME at end of trial	127	15.98	2.53	**0.01**
B. Treatment Group Catch-Up	**Sample Size**	**Treatment Group Mean**	**Difference Comparison vs. Treated**	**Single p-value**
HOME at enrollment	94	16.14	1.28	0.13
HOME at end of trial	94	18.53	0.91	0.39

Notes: Panel A reports the estimated impact of treatment on HOME Scores compared to controls. The first row reports the results at baseline (1986) before the intervention began and the second row reports the results at end of the two-year intervention period (1988). The treatment effects are estimated by linear regression and are interpreted as the differences in the means of the HOME Scores of the stunted treatment and stunted control groups conditional on baseline values of child age, gender, weight-for-height z-score, maternal employment, and maternal education. Reported p-values are for one-sided block permutation tests of the null hypothesis of no treatment effect (Single p-value). Permutation blocks are based on the conditioning variables used in the treatment effect regressions. Panel B reports the difference in means of the HOME Scores of the weighted non-stunted comparison group compared to the stunted treatment group. Again, the first row reports the results at baseline and the second row at the end of the intervention period. The comparison group observations are weighted using Inverse Probability Weights (IPW) to correct for attrition. Our p-values are for one-sided block permutation tests of the null hypothesis of complete catch-up (Single p-value). Permutation blocks are based on gender.

Table 8: Impact of Treatment on Education and Skills

	Sample Size	**Control Mean**	**Treatment Effect**	**Single p-value**	**Stepdown p-value**
A. Schooling					
School years completed	105	10.96	0.61	**0.07**	0.16
Any vocational training	105	0.56	0.12	0.16	0.16
Any college	104	0.04	0.11	**0.07**	0.14
In school	97	0.15	0.17	**0.01**	**0.04**
In school full time	97	0.07	0.18	**0.01**	**0.01**
B. Exams					
Passed at least one CXC exam	94	0.22	0.15	**0.09**	0.15
Passed 4 or more CXC exams	94	0.1	0.16	0.12	0.12
Passed at least one CAPE	94	0.00	0.09	**0.03**	**0.08**
C. Skills					
Cognitive factor	102	-0.46	0.59	**0.00**	**0.01**
Externalizing Behavior factor	102	-0.23	0.22	0.17	0.30
Internalizing Behavior factor	102	-0.32	0.39	**0.02**	**0.05**
Ever expelled from school	105	0.17	-0.12	**0.02**	**0.02**

Notes: This table reports the estimated impact of treatment on educational and skill outcomes. Treatment effects are reported for the following sets of outcomes indicated by the row blocks: (A) Schooling, (B) Exam, (C) Skills. The treatment effects are estimated by linear regression and are interpreted as the differences in the means of the outcomes between the stunted treatment and stunted control groups conditional on baseline values of child age, gender, weight-for-height z-score, maternal employment, and maternal education. Our p-values are for one-sided block permutation tests of the null hypothesis of no treatment effect (Single p-value) and multiple hypotheses (Stepdown p-value) of no treatment. Permutation blocks are based on the conditioning variables used in the treatment effect regressions.

Table 9: Treatment Group and Control Group Catch-up in Education and Skills

	I. Non-Stunted Comparison Group vs. Treatment					II. Non-Stunted Comparison Group vs. Control				
	Sample Size	Treatment Mean	Difference in Means	Single p-value	Stepdown p-value	Sample Size	Control Mean	Difference in Means	Single p-value	Stepdown p-value
A. Schooling										
Total years of education	115	11.50	-0.12	0.60	0.84	108	11.08	0.47	**0.05**	0.16
Any vocational training	115	0.66	-0.09	0.81	0.93	116	0.52	0.03	0.37	0.37
Any college	110	0.13	0.01	0.38	0.73	113	0.08	0.11	**0.04**	0.14
In school	113	0.25	-0.03	0.58	0.86	113	0.20	0.10	**0.08**	0.19
In school full time	113	0.18	-0.08	0.84	0.84	113	0.09	0.06	0.14	0.25
B. Exams										
Passed at least one CXC exam	106	0.30	0.18	**0.02**	0.04	110	0.27	0.29	**0.00**	**0.00**
Passed 4 or more CXC exams	106	0.21	0.13	**0.06**	0.10	110	0.12	0.25	**0.00**	**0.00**
Passed at least one CAPE	106	0.08	0.01	0.35	0.35	110	0.04	0.10	**0.01**	**0.01**
C. Skills										
Cognitive factor	112	-0.07	0.38	**0.01**	**0.04**	118	-0.32	0.81	**0.00**	**0.00**
Externalizing Behavior factor	112	0.01	0.06	0.27	0.43	118	-0.18	0.37	**0.02**	**0.04**
Internalizing Behavior factor	112	0.07	0.06	0.38	0.41	118	-0.41	0.48	**0.00**	**0.01**
Ever expelled from school	115	-0.02	-0.09	0.98	0.98	116	-0.10	0.06	0.15	0.15

Notes: This table reports the results of the analyses of treatment and control groups catch-up to the comparison group in educational and skill outcomes. Column panel I reports the results of the analyses of stunted treatment group catch-up to the non-stunted comparison group while Column panel II reports the control group catch-up. Catch-up is reported for the following sets of outcomes indicated by the row blocks: (A) Schooling, (B) Exam, (C) Skills. Catch-up is estimated as the difference in the means of the outcomes between the weighted non-stunted comparison group and the designated stunted group. The comparison group observations are weighted using Inverse Probability Weights (IPW) to correct for attrition. Our p-values are for one-sided block permutation tests of the null hypothesis of complete catch-up (Single p-value) and multiple hypotheses (Stepdown p-value) of complete catch-up. Permutation blocks are based on gender.

Table 10: Impact of Stimulation Treatment on Log Earnings by Gender

	I. Females			II. Males		
	Treatment Effect	**Single p-value**	**Stepdown p-value**	**Treatment Effect**	**Single p-value**	**Stepdown p-value**
A. First Job						
All	0.22	0.30	0.30	0.11	0.11	0.16
Full Time	0.34	**0.09**	0.12	0.11	0.13	0.13
Non-Temp	0.45	**0.04**	**0.09**	0.37	**0.05**	**0.09**
B. Last Job						
All	0.18	0.18	0.18	0.29	0.15	0.15
Full Time	0.51	**0.01**	**0.01**	0.26	0.12	0.17
Non-Temp	0.53	**0.01**	**0.02**	0.40	**0.03**	**0.07**
C. Current Job						
All	0.28	0.13	0.13	0.15	0.29	0.29
Full Time	0.66	**0.00**	**0.01**	0.16	0.26	0.36
Non-Temp	0.56	**0.02**	**0.04**	0.24	0.15	0.28
D. Average Earnings						
All	0.30	0.19	0.19	0.44	0.01	0.02
Full Time	0.39	**0.07**	0.11	0.24	**0.09**	**0.09**
Non-Temp	0.49	**0.04**	**0.09**	0.37	**0.05**	**0.06**

Notes: This table reports the estimated impact of treatment on log monthly earnings using the sample with imputations for the earnings of missing migrants. Column panel I reports the results for Females while Column panel II reports the results for Males. In each sample, treatment effects are reported for the following jobs as indicated by the row blocks: (A) First Job, (B) Last Job, (C) Current Job, and (D) Average Lifetime Earning over all jobs. Within each type of job, results are reported for the following types of workers as indicated by the rows: All workers, Full Time Workers, and full time Non-Temporary workers. The treatment effects are estimated by separate linear regressions for each gender. The treatment effects are interpreted as the differences in the means of Log Earnings between the stunted treatment and stunted control groups conditional on baseline values of child age, weight-for-height z-score, maternal employment, and maternal education. Our p-values are for one-sided block permutation tests of the null hypothesis of no treatment effect (Single p-value) and multiple hypotheses (Stepdown p-value) of no treatment. Permutation blocks are based on the conditioning variables used in the treatment effect regressions.

Table 11: Treatment and Control Group Catch-up in Log Earnings by Gender

	I. Non-Stunted Comparison Group vs. Treatment						II. Non-Stunted Comparison Group vs. Control					
	Females			Males			Females			Males		
	Difference in Means	Single p-value	Stepdown p-value	Difference in Means	Single p-value	Stepdown p-value	Difference in Means	Single p-value	Stepdown p-value	Difference in Means	Single p-value	Stepdown p-value
A. First Job												
All	-0.12	0.79	0.90	0.37	**0.01**	**0.02**	0.00	0.49	0.49	0.46	**0.01**	**0.01**
Full Time	-0.14	0.82	0.90	0.46	**0.01**	**0.01**	0.03	0.41	0.57	0.56	**0.00**	**0.00**
Non-Temp	-0.35	0.95	0.95	0.04	0.42	0.42	0.02	0.46	0.61	0.39	**0.02**	**0.02**
B. Last Job												
All	0.08	0.35	0.51	0.19	0.24	0.38	0.25	**0.09**	0.14	0.31	0.08	0.14
Full Time	0.02	0.46	0.56	-0.01	0.52	0.60	0.40	**0.02**	**0.03**	0.18	0.19	0.19
Non-Temp	-0.27	0.90	0.90	-0.10	0.68	0.68	0.13	0.28	0.28	0.22	0.15	0.21
C. Current Job												
All	-0.09	0.63	0.77	0.16	0.26	0.37	0.18	0.23	0.31	0.24	0.18	0.29
Full Time	-0.30	0.90	0.90	-0.06	0.60	0.69	0.39	**0.06**	0.10	**0.00**	0.50	0.61
Non-Temp	-0.30	0.88	0.92	-0.23	0.79	0.79	0.19	0.26	0.26	-0.08	0.61	0.61
D. Average Earnings												
All	0.03	0.43	0.55	-0.02	0.52	0.60	0.23	**0.09**	0.15	0.30	**0.05**	**0.07**
Full Time	-0.09	0.69	0.75	0.02	0.46	0.54	0.15	0.18	0.22	0.21	0.13	0.13
Non-Temp	-0.33	0.93	0.93	-0.06	0.61	0.60	0.06	0.39	0.39	0.26	0.11	0.12

Notes: This table reports the results of Treatment and Control Groups Catch-up in log earnings by Gender. Column panel I reports the results for stunted treatment group catch-up to the non-stunted comparison group for Females and Males. Column panel II reports the results for stunted control group catch-up also by gender. In each sample, catch-up is reported for the following jobs as indicated by the row blocks: (A) First Job, (B) Last Job, (C) Current Job, and (D) Average Lifetime Earning over all jobs. Within each type of job, results are reported for the following types of workers as indicated by the rows: All workers, Full Time Workers, and full time Non-Temporary workers. Catch-up is estimated as the difference in the means of Log Earnings between the weighted non-stunted comparison group and the stunted group for each gender. The comparison group observations are weighted using Inverse Probability Weights (IPW) to correct for attrition. Our p-values are for one-sided block permutation tests of the null hypothesis of complete catch-up (Single p-value) and multiple hypotheses (Stepdown p-value) of complete catch-up. Permutation blocks are based on gender.

Table 12: Impact of Stimulation Treatment on Education and Skills by Gender

	I. Females			II. Males		
	Treatment Effect	**Single p-value**	**Stepdown p-value**	**Treatment Effect**	**Single p-value**	**Step Down p-value**
A. Schooling						
Total years of education	0.77	**0.03**	**0.08**	0.41	0.38	0.53
Any training	0.07	0.34	0.51	0.15	0.10	0.23
Any college	0.14	**0.04**	0.15	0.08	0.40	0.40
Currently in school	0.06	0.36	0.36	0.25	**0.01**	**0.05**
Currently in school full time	0.13	0.13	0.28	0.14	**0.05**	0.16
B. Exams						
Passed at least 1 CXC exams	0.28	**0.03**	**0.06**	-0.01	0.56	0.56
Passed 4 or more CXC exams	0.2	**0.05**	**0.05**	0.09	0.39	0.53
Passed at least 1 CAPE exams	0.15	**0.02**	**0.06**	0.04	0.30	0.56
C. Skills						
Cognitive factor	0.36	0.13	0.32	0.57	**0.01**	**0.04**
Externalizing Behavior factor	0.58	**0.05**	0.10	-0.07	0.6	0.71
Internalizing Behavior factor	0.76	**0.01**	**0.02**	0.08	0.37	0.53
Ever expelled from school	-0.09	**0.09**	**0.09**	-0.15	**0.04**	**0.04**

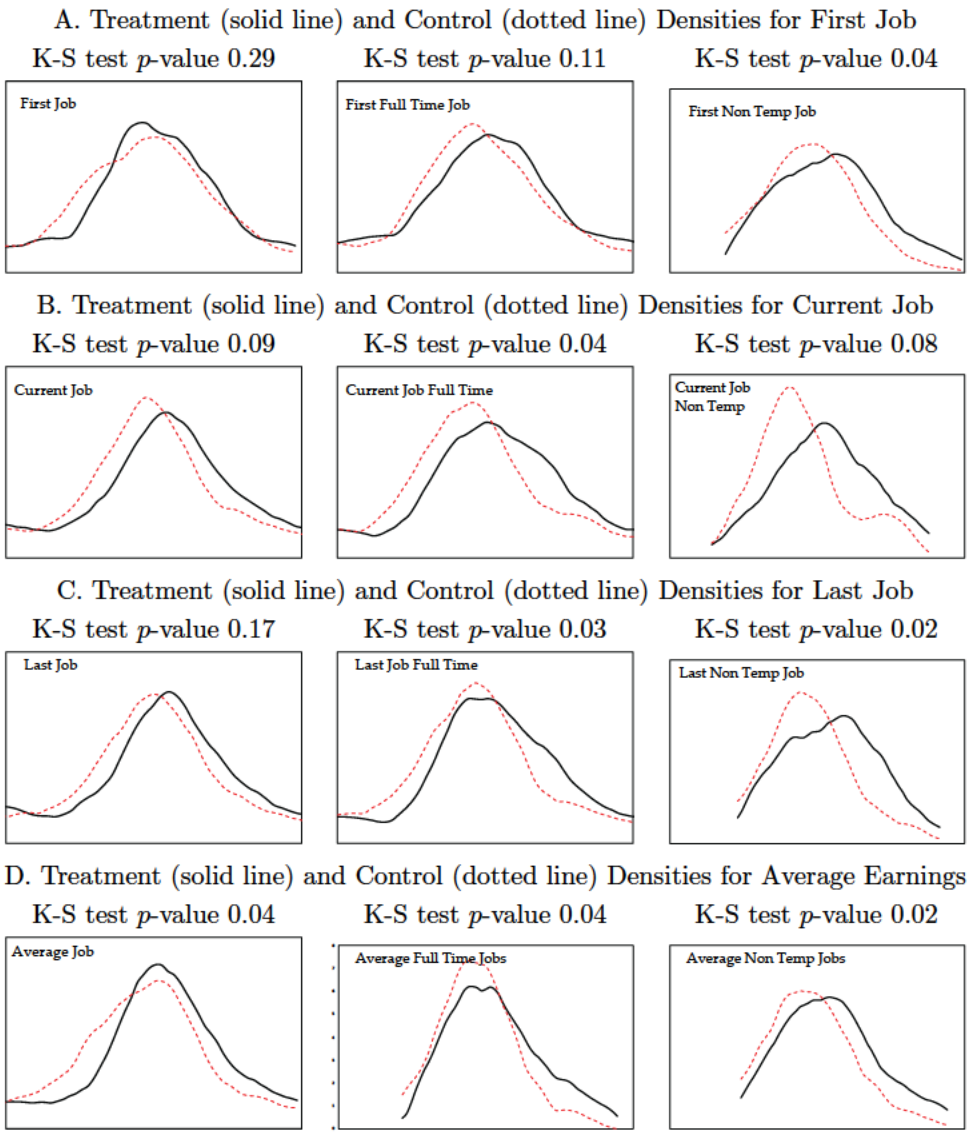
Notes: This table reports the estimated impact of treatment on educational and skill outcomes by gender. Column panel I reports the results for Females while Column panel II reports the results for Males. The treatment effects are reported for the following sets of outcomes indicated by the row blocks: (A) Schooling, (B) Exam, (C) Skills. The treatment effects are estimated by separate linear regressions for each gender and are interpreted as the differences in the means of the outcomes between the stunted treatment and stunted control groups conditional on baseline values of child age, weight-for-height z-score, maternal employment, and maternal education. Our p-values are for one-sided block permutation tests of the null hypothesis of no treatment effect (Single p-value) and multiple hypotheses (Stepdown p-value) of no treatment. Permutation blocks are based on the conditioning variables used in the treatment effect regressions.

Table 13: Treatment and Control Group Catch-up in Education and Skills by Gender

	I. Non-Stunted Comparison Group vs. Treatment						II. Non-Stunted Comparison Group vs. Control					
	Females			Males			Females			Males		
	Diff. Means	Single p-value	Stepdown p-value	Diff. Means	Single p-value	Stepdown p-value	Diff. Means	Single p-value	Stepdown p-value	Diff. Means	Single p-value	Stepdown p-value
A. Schooling												
Total years of education	0.01	0.52	0.74	-0.22	0.73	0.95	1.10	**0.00**	**0.02**	-0.14	0.64	0.64
Any training	-0.13	0.83	0.83	-0.05	0.63	0.96	0.06	0.33	0.33	0.01	0.47	0.74
Any college	0.05	0.33	0.63	-0.03	0.65	0.94	0.25	**0.01**	**0.02**	-0.02	0.63	0.77
Currently in school	0.02	0.42	0.68	-0.07	0.76	0.85	0.10	0.21	0.31	0.09	0.11	0.37
Currently in school full time	-0.04	0.62	0.79	-0.11	0.91	0.91	0.10	0.18	0.35	0.02	0.29	0.65
Passed $\geq$ 1 CXC exams	0.09	0.25	0.45	0.26	**0.02**	**0.04**	0.42	**0.00**	**0.00**	0.16	0.11	0.11
B. Exams												
Passed $\geq$ 4 CXC exams	0.05	0.35	0.49	0.20	**0.04**	**0.07**	0.32	**0.00**	**0.01**	0.17	**0.07**	0.17
Passed $\geq$ 1 CAPE exams	0.03	0.40	0.40	0.00	0.44	0.44	0.17	**0.03**	**0.03**	0.03	0.10	0.18
Cognitive factor	0.34	0.11	0.24	0.42	**0.04**	0.13	0.88	**0.00**	**0.00**	0.75	**0.00**	**0.01**
C. Skills												
Externalizing Behavior factor	0.08	0.37	0.52	0.04	0.44	0.61	0.82	**0.00**	**0.00**	-0.02	0.52	0.72
Internalizing Behavior factor	0.04	0.43	0.61	0.08	0.42	0.5	0.90	**0.00**	**0.00**	0.10	0.33	0.40
Ever expelled from school	-0.03	0.43	0.43	-0.14	0.97	0.97	-0.09	**0.09**	**0.09**	-0.03	0.39	0.39

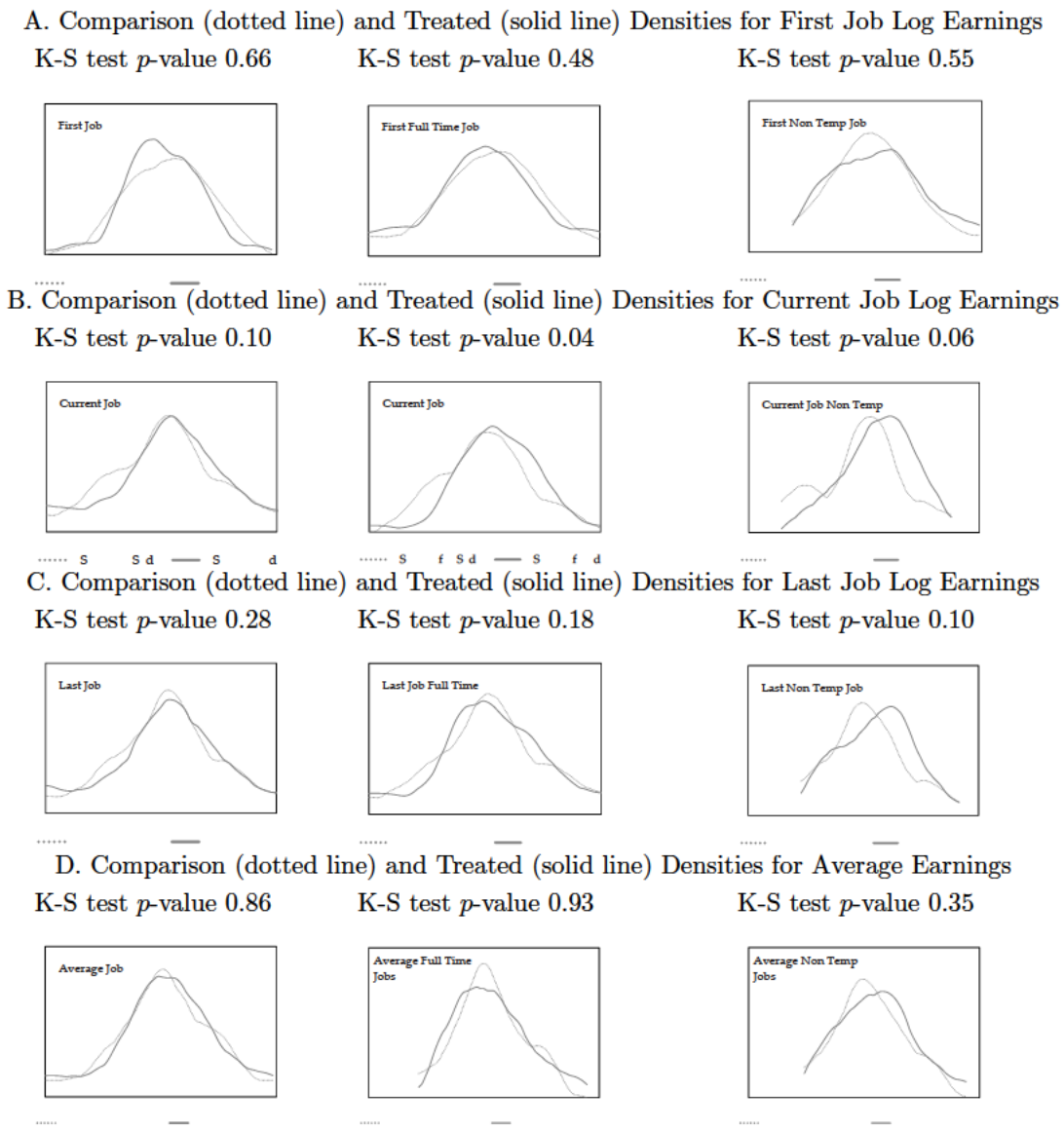
Notes: This table reports the results of Treatment and Control Group Catch-up in educational and skills outcomes by Gender. Column panel I reports the results for stunted treatment group catch-up to the non-stunted comparison group for Females and Males. Column panel II reports the results for stunted control group catch-up also by gender. In each sample, catch-up is reported for the following outcomes as indicated by the row blocks: (A) Schooling, (B) Exam, (C) Skills. Catch-up is estimated as the difference in the means of the outcomes between the weighted non-stunted comparison group and the stunted group for each gender. The comparison group observations are weighted using Inverse Probability Weights (IPW) to correct for attrition. Our p-values are for one-sided block permutation tests of the null hypothesis of complete catch-up (Single p-value) and multiple hypotheses (Stepdown p-value) of complete catch-up. Permutation blocks are based on gender.

Figure 1: Impact of Stimulation Treatment on the Densities of Log Earnings



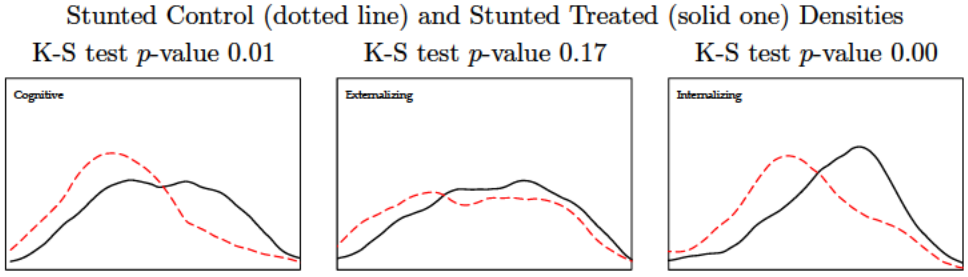
Notes: These figures present the log earnings densities for the Treatment and Control Groups. The Control density is the dotted line and the Treatment density the solid one. Separate densities are presented for earnings in the first, last and current jobs as well as average lifetime earnings by all workers, full-time workers, and full-time non-temporary workers. The densities are estimated using Epanechnikov kernels. The treatment densities were estimated with an optimal bandwidth defined as the width that would minimize the mean integrated squared error under the assumption that the data are Gaussian. For comparability purposes, the same bandwidth used was used for the corresponding Control Group. The p-values are for Kolmogorov-Smirnov tests of the equality of treatment and control densities.

Figure 2: Catch-up of Treatment Group Earnings to Comparison Group Earnings



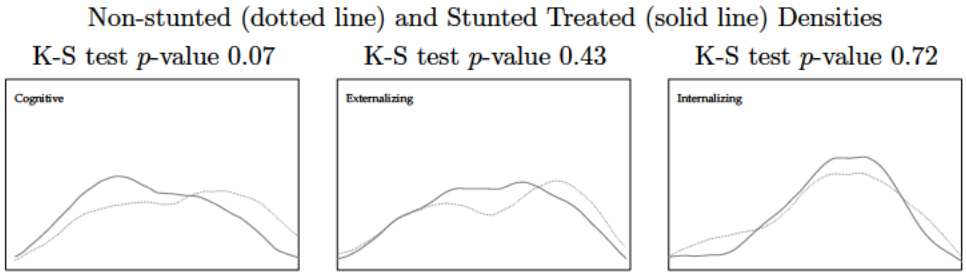
Notes: These figures present the log earnings densities for the non-stunted Comparison and stunted Treatment Groups. The comparison group density is the dotted line and the Treatment group density the solid one. Separate densities are presented for earnings in the first, last and current jobs as well as average lifetime earnings by all workers, full-time workers, and full-time non-temporary workers. The densities are estimated using Epanechnikov kernels. The treatment densities were estimated with an optimal bandwidth defined as the width that would minimize the mean integrated squared error under the assumption that the data are Gaussian. For comparability purposes, the same bandwidth used was used for the corresponding Control Group. The p-values are for Kolmogorov-Smirnov tests of the equality of comparison and treatment densities.

Figure 3: Impact of Stimulation Treatment on Skills



Notes: These figures present the cognitive, internalizing and externalizing factor densities for the Treatment and Control Groups. The Control density is the dotted line and the Treatment density the solid one. The densities are estimated using Epanechnikov kernels. The treatment densities were estimated with an optimal bandwidth defined as the width that would minimize the mean integrated squared error under the assumption that the data are Gaussian. For comparability purposes, the same bandwidth used was used for the corresponding Control Group. The p-values are for Kolmogorov-Smirnov tests of the equality of treatment and control densities.

Figure 4: Catch-up of Treatment Group Skills to Comparison Group Earnings



Notes: These figures present the cognitive, internalizing and externalizing factor densities for the non-stunted Comparison and Treatment Groups. The density for the comparison group is the dotted line and the density for the treated is the solid one. The densities are estimated using Epanechnikov kernels. The treatment densities were estimated with an optimal bandwidth defined as the width that would minimize the mean integrated squared error under the assumption that the data are Gaussian. For comparability purposes, the same bandwidth used was used for the corresponding Control Group. The p-values are for Kolmogorov-Smirnov tests of the equality of treatment and control densities.

A Appendix: Supplemental Tables (For Reviewers Only)

Table A.1: External Validity of Non-stunted Comparison Group

A. Comparison with JSLC 1992	**JSLC Mean**	**Comparison Group Mean**	**Difference Means**	**Single p-value**
Mother completed 9th grade	0.57	0.83	0.26	**0.00**
Father present in the house	0.61	0.73	0.12	**0.09**
Poor sanitation	0.16	0.13	-0.03	0.36
Piped water in the house	0.66	1.12	0.46	**0.00**
B. Comparison with JLFS 2008	**JLFS Mean**	**Comparison Group Mean**	**Diff. in Means**	**Single p-value**
Studying full time	0.09	0.06	-0.03	0.46
Highest Grade Completed	10.83	10.87	0.04	0.76
Passed at least one CXC exam	0.44	0.36	-0.08	0.22
Passed 4 or more CXC exams	0.28	0.32	0.04	0.33
Passed at least one CAPE	0.13	0.20	0.07	**0.02**

Notes: This table compares the Non-stunted Comparison Group with the Jamaican Survey of Living Conditions (JSLC) in Panel A and with the Jamaican Labor Force Survey 2008 (JLFS) in Panel B. The JSLC sample is restricted to households with children between the ages of 9 and 24 months from the Kingston Metropolitan Area. The JLFS sample includes individuals of ages 22 and 23 years old living in Kingston Metropolitan Area. The p-values reported in the last column are for two-sided permutation tests of the null hypotheses that the difference in means between the two samples is zero.

Table A.2: Baseline (1986) Descriptive Statistics for Stunted Sample Enrolled in the Study

	Sample Size	**Control Mean**	**Treatment Mean**	**Difference in Means**	**Single p-value**
A. Parental Characteristics					
Mother present	127	0.97	0.96	-0.03	0.48
Mother/guardian's age (years)	127	23.9	25.4	1.53	0.19
Mother/guardian employed	127	0.15	0.29	0.14	**0.09**
Mothers/guardian education	127	0.20	0.06	-0.14	**0.02**
Mother/guardian's education, any training	127	0.20	0.27	0.07	0.42
Mother/guardian's PPVT	127	84.5	86.1	1.65	0.63
Mothers/guardian's height (cm)	125	159.5	159.1	-0.36	0.73
Father presence	127	7.55	7.21	-0.34	0.28
HOME score on enrolment	127	-0.21	0.07	0.14	0.65
Housing index	127	7.55	7.21	-0.34	0.28
B. Child Characteristics					
Child age (years)	127	1.53	1.54	0.01	0.9
Male	127	0.59	0.55	-0.04	0.69
Child's birth order	127	2.89	3.32	0.43	0.28
Birth Weight $<$ 2500 grams	126	0.19	0.22	0.03	0.77
Head Circumference (cm)	127	46.1	45.9	-0.21	0.47
Daily Calories Consumed	127	970.4	939.0	-31.4	0.71
Daily Proteins Consumed (grams)	127	25.6	27.6	2.04	0.55
Griffith Developmental Quotient	127	97.2	98.9	1.7	0.3
Stunting	127	-2.91	-3.01	-0.1	0.33
Wasting	127	-0.94	-1.17	-0.23	**0.05**

Notes: This table presents baseline means of variables of interest for the stunted treatment and control groups at baseline (1986) for the sample enrolled at baseline. The p-values reported in the last column are for two-sided block permutation tests of the null hypotheses that the difference in means between treatment and control groups are zero. Variable definitions include: PPVT denotes the raw score from Peabody Picture Vocabulary Test (Dunn and Dunn, 1981), HOME denotes the raw score from the HOME environment test (Caldwell, 1967), and Griffith Development Quotient reports the raw score for this test (Griffiths, 1954; Griffiths 1970).

Table A.3: Attriton From the Stunted Sample

	Total Sample Size	**Control Sample Size**	**Difference in Attrition Rates**	**Single p-value**
A. Attrition				
Baseline Sample Size	127	65	–	–
Sample lost in 2008 resurvey	22	13	-0.05	0.4
B. Reason lost				
Could not locate	10	6	-0.03	0.57
Located but refused interview	3	1	0.02	0.55
Died	9	6	-0.04	0.37

Notes: This table presents baseline sample sizes and attrition in panel A. The difference in attrition rates is the difference between the treatment and control groups. The p-values reported in the last 2 columns are for two-sided block permutation tests of the null hypotheses that the difference in treatment and control means are zero. The permutation blocks are child's age and sex.

Table A.4: p-Values for Tests of Attrition Bias in the Stunted Sample

	Full Sample	**Treatment Group**	**Control Group**
A. Parental Characteristics			
Mother present	1.00	0.68	1.00
Mother/guardian's age (years)	**0.02**	0.12	0.18
Mother has $>$ 9 years of Schooling	0.31	1.00	0.30
Mothers education	1.00	0.72	1.00
Mother has any job training	0.56	0.20	0.69
Mother/guardian's PPVT	0.47	0.74	0.57
Mother/guardian's height (cm)	0.88	0.60	0.37
Father present	0.80	0.75	1.00
HOME score on enrolment	0.30	0.12	0.71
Housing index	0.68	0.95	0.53
B. Child Characteristics			
Child age (years)	0.18	0.28	0.47
Male	0.26	0.56	0.50
Child's birth order	0.26	0.41	0.56
Birth Weight $<$ 2500 grams	0.12	1.00	**0.09**
Head Circumference (cm)	0.21	0.25	0.52
Daily Calories Consumed	0.73	0.05	0.20
Daily Protein Consumed (grams)	0.47	0.02	0.49
Griffith Developmental Quotient	0.59	0.87	0.34
Stunting	0.30	0.24	0.80
Wasting	0.27	0.12	0.81

Notes: This table presents p-values for two-sided permutation tests of the null hypotheses that the difference in baseline means of the sample found in the 2008 and the sample not found in 2008 are equal. The first column reports that results for the full sample and the next two columns report the results separately for the treatment and control samples.

Table A.5: Attrition From the Non-Stunted Comparison Group

	Non-Attrited Group Mean	**Attrited Group Mean**	**Difference in Means**	**Single p-value**
Maternal age	32.38	37.45	5.07	**0.05**
Mother Present	0.86	0.66	-0.20	0.13
Maternal employment	0.66	0.56	-0.10	0.47
Maternal education	0.36	0.17	-0.19	**0.10**
Maternal PPVT Score	94.78	84.35	-10.43	**0.09**
Home stimulation: books +paper	0.46	0.20	-0.26	0.30
Home stimulation: games and trips	0.03	-0.01	-0.04	0.89
Home stimulation: verbal stimulation	0.12	-0.30	-0.42	**0.05**
Home stimulation: writing material	0.09	-0.06	-0.15	0.44
Housing score	8.83	9.56	0.73	**0.09**
Child misses school because of money	0.33	0.28	-0.05	0.77
Weight for Age z-Score	0.19	0.16	-0.03	0.88
Height for Age z-Score	0.81	0.90	0.09	0.76
Stanford Binet	82.23	80.74	-1.49	0.48
Ravens	13.86	12.84	-1.02	0.24

Notes: This table presents baseline descriptive statistics for the sample of non-stunted comparison group member found (Non-Attrited) in the 2008 survey and the group lost (Attrited) in the 2008 survey. The p-values reported in the last 2 column are for two-sided permutation tests of the null hypotheses that the difference in non-attrited and attrited group means are zero.

Table A.6: Baseline (1986) Descriptive Statistics for the Non-Stunted versus Stunted Samples

	Non-stunted Group Mean	**Stunted Group Mean**	**Difference in Means**	**Single p-value**
A. Parental Characteristics				
Mother present	1.0	1.0	0	0.3
Mother/guardian's age (years)	25.1	25.0	-0.1	1
Mother /guardian employed	0.2	0.2	0.0	0.9
Mother/guardian school 9th grade	0.1	0.1	0.0	0.7
Mother/guardians PPVT	85.8	98.2	12.4	**0**
Mothers/guardians height (cm)	159.3	163.8	4.5	**0**
Father present	0.5	0.6	0.1	0.5
HOME score on enrolment	16.6	17.9	1.3	0.2
Housing index	7.4	8.7	1.3	**0**
B. Child Characteristics				
Age (years)	1.6	1.6	0.0	1
Male	0.5	0.4	-0.1	0.5
Child's birth order	3.2	2.1	-1	**0.1**
Birth Weight $<$ 2500 grams	0.2	0.0	-0.2	**0**
Head Circumference (cm)	46.1	47.8	1.7	**0**
Daily Calories Consumed	959.8	909.0	-50.8	0.6
Daily Protein Consumed (grams)	27.0	25.9	-1.1	0.8
Griffith Developmental Quotient	98.2	106.6	8.4	**0**
Height for Age z-Score	-2.9	0.1	3.0	**0**
Weight for Height z-Score	-1	0.1	1.1	**0**
C. Variables at 7 years				
Mother present	0.9	0.9	0.0	0.9
Mother/guardian's age (years)	32.8	32.4	-0.4	0.7
Mother /guardian employed	0.4	0.6	0.2	**0**
Mother/guardian school 9th grade	0.3	0.4	0.1	0.5
Mother/guardians PPVT	86.6	94.7	8.1	**0**
Father present	0.3	0.5	0.2	**0**
Housing index	8.4	8.9	0.5	**0.1**
Missed school due to lack of money	0.5	0.3	-0.2	**0**

Notes: This table presents baseline descriptive statistics for the Non-stunted Comparison Group vs. Full Stunted Sample (Treatment and Control Groups). The p-values reported in the last 2 column are for two-sided permutation tests of the null hypotheses that the difference in non-stunted and stunted group means are zero.

Table A.7: Testing for Gender Differences in the Impact of Treatment and Catch Up in Log Earnings

	I. Treatment vs. Control			II. Comparison vs. Treatment		
	Treatment Effect	**Differential Effect for Women**	**Single p-Value**	**Treatment Effect**	**Differential Effect for Women**	**Single p-Value**
A. First Job						
All	0.23	-0.15	0.58	-0.14	0.47	**0.07**
Full time	0.30	-0.20	0.50	-0.15	0.59	**0.03**
Non temp	0.46	-0.08	0.81	-0.35	0.38	0.24
B. Last Job						
All	0.25	0.03	0.94	0.06	0.11	0.77
Full time	0.54	-0.28	0.33	-0.02	-0.06	0.85
Non temp	0.52	-0.13	0.69	-0.31	0.13	0.71
C. Current Job						
All	0.37	-0.23	0.59	-0.18	0.24	0.55
Full time	0.81	-0.65	**0.08**	-0.38	0.56	0.60
Non temp	0.72	-0.46	0.29	-0.39	0.07	0.89
D. Average Earnings						
All	0.27	0.14	0.63	0.00	-0.08	0.76
Full time	0.38	-0.14	0.60	-0.11	0.08	0.75
Non temp	0.50	-0.14	0.67	-0.36	0.25	0.45

Notes: This table reports the results of tests for gender differences in the treatment effect and catch-up of the treatment group in log earnings. The columns in Panel I present the results for the treatment effects. The columns in Panel II present the analysis of treatment group catch-up to the comparison group. The treatment effects and treatment effects interacted with a female dummy variable to estimate the Differential Effect for Women are estimated by linear regression. The treatment effect is interpreted as the difference in the means of log earnings for males between the stunted treatment and stunted control groups conditional on baseline values of child age, gender, weight-for-height z-score, maternal employment, and maternal education. The treatment effect interacted with the female dummy is interpreted as the difference between the treatment effect for females versus males conditional on the same variables. Our p-values are for two-sided block permutation tests of the null hypothesis of no differential effect for women (Single p-value). Permutation blocks are based on the conditioning variables used in the treatment effect regressions. Catch-up for males is estimated as the difference in the means of the outcomes between the weighted non-stunted comparison group and the stunted group. The comparison group observations are weighted using Inverse Probability Weights (IPW) to correct for attrition. The interaction of the difference in means and females is interpreted as the difference in catch-up for females versus males.

Table A.8: Testing for Gender Differences in the Impact of Treatment on Education and Skills

	Treatment Effect	**Differential Treatment Effect for Women**	**Single p-Value**
A.Schooling			
Total years of education	0.95	-0.90	0.11
Any vocational training	0.16	-0.07	0.73
Any college	0.16	-0.12	0.27
In school	0.07	0.12	0.40
In school full time	0.12	0.03	0.80
B.Exams			
Passed at least one CXC exam	0.31	-0.37	**0.04**
Passed 4 or more CXC exams	0.25	-0.24	0.12
Passed at least one CAPE	0.14	-0.10	0.21
C.Skills			
Cognitive factor	0.49	-0.11	0.76
Externalizing Behavior factor	0.60	-0.59	0.13
Internalizing Behavior factor	0.88	-0.71	**0.06**
Ever expelled from school	0.09	0.08	0.51

Notes: This table reports the results of tests for gender differences in the treatment effect on schooling and skills. The treatment effects and the Differential Treatment Effect for Women are estimated by linear regression. The treatment effect is interpreted as the differences in the means of the outcome variables for males between the stunted treatment and stunted control groups conditional on baseline values of child age, gender, weight-for-height z-score, maternal employment, and maternal education. The differential treatment effect is interpreted as the difference between the treatment effect for females versus males conditional on the same variables. Our p-values are for two-sided block permutation tests of the null hypothesis that the interactions of the treatment effect and female are zero (Single p-value). Permutation blocks are based on the conditioning variables used in the treatment effect regressions.

Table A.9: Outliers Robustness Analysis of Stimulation Treatment on Log Earnings

	I. Original Sample		II. Trimmed Sample	
	Treatment Effect	**Single p-value**	**Treatment Effect**	**Single p-value**
A. First Job				
First Job	0.27	0.11	0.24	**0.03**
First Full Time	0.35	**0.04**	0.24	**0.03**
First Non Temp	0.53	**0.01**	0.23	**0.07**
B. Last Job				
All	0.27	**0.06**	0.18	**0.10**
Full Time	0.40	**0.00**	0.31	**0.00**
Non-Temp	0.50	**0.00**	0.33	**0.01**
C. Current Job				
All	0.27	**0.09**	0.26	**0.06**
Full Time	0.43	**0.10**	0.44	**0.00**
Non-Temp	0.44	**0.10**	0.40	**0.02**
D. Average Earnings				
All	0.40	**0.01**	0.28	**0.01**
Full Time	0.34	**0.01**	0.13	0.11
Non-Temp	0.47	**0.00**	0.18	**0.10**

Notes: This table reports the estimated impact of treatment on log monthly earnings using the original sample. The results from Table 3 are presented in column panel I and the results for the sample trimmed of the lowest and highest 5% values is presented in panel II. In each sample, treatment effects are reported for the following jobs as indicated by the rows blocks: (A) First Job, (B) Last Job, (C) Current Job, and (D) Average Lifetime Earning over all jobs. Within each type of job, results are reported for the following types of workers as indicated by the rows: All workers, Full Time Workers, and full time Non-Temporary workers. The treatment effects are estimated by linear regression and are interpreted as the differences in the means of log earnings between the stunted treatment and stunted control groups conditional on baseline values of child age, gender, weight-for-height z-score, maternal employment, and maternal education. Our p-values are for one-sided block permutation tests of the null hypothesis of no treatment effect (Single p-value). Permutation blocks are based on the conditioning variables used in the treatment effect regressions.

B Appendix: Construction of the Factors for Skills (For Reviewers Only)

Cognitive, internalizing and externalizing factors are estimated as factor scores from a confirmatory factor analysis. We first perform Exploratory Factor Analysis to confirm the model predicted by the theory.

The confirmatory factor analysis assumes dedicated measures. Preliminary exploratory factor analysis is used to identify such measures. Exploratory Factor Analysis reveals that antisocial behavior and digit span measures do not load on any factor, and are also poorly correlated with any other measurement. They were excluded from the final measurement system. Another measurement, ADHD Conners scale, was removed from the system because theory suggests that ADHD loads on both internalizing and externalizing behavior, hence violating our assumption of dedicated measurements.

Table B.1 reports the factor loadings from our Exploratory Factor Analysis. It shows that each measurement loads clearly on only one factor. Exploratory factor analysis also suggests that a three factor model is the most suitable for explaining the measurements according to the Kaiser test, the Scree test and the Velicer test, with each measure loading clearly on one and only one factor. The confirmatory factor analysis confirms that a 3 factor structure (Chi square test, RMSEA, CFI, AIC and BIC) is preferable to a simple two factor model. Bartlett scores were used to create factor scores.

Measurements for cognitive factors included WRAT math, WRAT reading, reading comprehension, PPVT, verbal analogies, Raven matrices, WAIS full scale IQ, all measured at 18 years old. Measurements for externalizing behavior include the Conners' scale for oppositional behavior, the Conners' scale for inattention, and the Conners' scale for hyperactivity. Measures for internalizing behavior include a self esteem scale, an anxiety scale, and a depression scale. We allow for factors to be correlated. In fact, cognitive factor and internalizing factor have a correlation of -0.25, externalizing and internalizing of 0.21 and cognitive and

externalizing of -0.25.

Table B.1: Exploratory Factor Analysis on Cognitive and Psychosocial Measurements for Full Stunted Sample (Treatment and Control Groups)

	Cognitive	**Internalizing**	**Externalizing**	**Uniqueness**
WRAT math	**0.84**	-0.04	0.04	0.29
WRAT reading	**0.95**	0.07	-0.05	0.11
Reading comprehension	**0.90**	0.05	0.03	0.22
PPVT	**0.88**	0.04	-0.04	0.22
Verbal analogies	**0.87**	-0.03	0.04	0.24
Ravens matrices	**0.74**	-0.04	0.00	0.43
Wais fullscale IQ	**0.91**	-0.03	-0.02	0.14
Conners oppostional behaviror	-0.02	-0.07	**0.84**	0.30
Conners cog. problems/inattention	0.02	0.00	**0.68**	0.55
Conners hyperactivity	-0.02	0.08	**0.67**	0.51
Self-Esteem	-0.13	**0.54**	0.01	0.65
Anxiety	0.01	**0.78**	0.04	0.37
Depression	0.07	**0.95**	-0.05	0.16

Notes: This table show the factor loadings estimations of 13 measures used to identify three latent factors associated with cognition, internalizing and externalising skills. Our factor model estimation uses standard procedures of Exploratory Factor Analysis. We adopt the Oblique promax factor rotation and our estimations are based on the full sample that includes Stunted Treatment Group, Stunted Control Group and the Non-stunted Comparison Group. Our data totals 166 participants.