

Looking Beyond Outcome Evaluations of Conditional Cash Transfers: The Differential Impact of Peru's *Juntos* on the Educational Achievement of Child Beneficiaries

Kristen McCollum

This thesis submitted in part fulfilment of the requirements for the degree of MSc in Impact Evaluation for International Development, University of East Anglia.

The data used come from Young Lives, a longitudinal study of childhood poverty that is tracking the lives of 12,000 children in Ethiopia, India (in the states of Andhra Pradesh and Telangana), Peru and Vietnam over a 15-year period. www.younglives.org.uk

Young Lives is funded by UK aid from the Department for International Development (DFID) and co-funded by Irish Aid from 2014 to 2016.

The views expressed here are those of the author. They are not necessarily those of the Young Lives project, the University of Oxford, DFID or other funders.

**Looking Beyond Outcome Evaluations of
Conditional Cash Transfers:
The Differential Impact of Peru's *Juntos* on the
Educational Achievement of Child Beneficiaries**

100072677

A dissertation submitted to the School of International
Development of the University of East Anglia in Part-fulfillment of
the requirements for the Degree of Master of Science

August 2015

Word Count: 8087

Table of Contents

ABSTRACT	1
INTRODUCTION	1
CCTs AND LATIN AMERICA	2
<i>Impact of CCTs on Education: What We Know So Far</i>	3
JUNTOS, PERU'S CONDITIONAL CASH TRANSFER PROGRAM	4
<i>The Impact of Juntos on Education</i>	5
<i>Examining Differential Effects: Maternal Education and Child Outcomes</i>	7
DATA	9
METHODS	10
<i>Theory of Change</i>	10
<i>Establishing Treatment and Control Groups</i>	12
<i>Propensity Score Weighting</i>	12
<i>Assumptions</i>	14
RESULTS	14
<i>Probability Model and Inverse Probability of Treatment Weights</i>	14
<i>Regression Analysis</i>	18
DISCUSSION	23
<i>Juntos and Ability Tracking</i>	23
<i>Reliability of Results</i>	24
<i>Implications for Juntos</i>	25
CONCLUSION	26
REFERENCES	28

Abstract

Despite the popularity of conditional cash transfers (CCTs) across Latin America, and proponents' insistence that the programs alleviate poverty in the long term, there is a lack of evidence on the impact of such programs on students' educational achievement. This paper represents one of just a handful of studies on the impact of CCTs on educational achievement, and to the author's knowledge, it is the first evaluation of its kind for Peru's *Juntos*. Using quasi-experimental methods, this study analyzes the differential impacts *Juntos* has by maternal education. Results show that *Juntos* has a negative impact on students whose mothers completed primary school. These results contribute to the wider debate on ability tracking in education and the long-term effectiveness of CCTs across Latin America.

Introduction

Since their first introduction in 1995, conditional cash transfers (CCTs) have proved a popular intervention in Latin America and the Caribbean. In just one decade, an estimated 70 million Latin Americans, amounting to around 12 percent of the population, were beneficiaries of a CCT (Lomelí, 2008). CCTs also proved an important step in the field of evaluation, as they were frequently implemented alongside a systematic evaluation framework. Such forethought supported the subsequent surge of evaluations on the topic of their effectiveness.

Most CCTs in Latin America have an educational component to their conditions. However, few evaluations have traveled beyond evaluating short-term educational outcomes. Research has shown CCTs to be an effective way of increasing enrollment and attendance (Baird et al., 2014), but little is known about long-term educational results, such as changes in educational achievement as measured with test scores.

This paper marks the first attempt to address the long-term questions of the educational impact of *Juntos*. Do the child beneficiaries score better on tests compared to their non-beneficiary counterparts? Are children from families with low education "catching up" to their peers whose parents have higher education, thus helping to break the intergenerational transmission of

poverty? The answers to such questions are key to the current stage of evaluation of CCTs, and may help to justify (or challenge) the widespread use of such programs across Latin America.

CCTs and Latin America

After the seeming success of *Bolsa Escola* in Brazil and *Progres*a (later, *Oportunidades*) in Mexico, CCTs quickly found themselves to be a popular social policy across Latin America (Handa & Davis, 2006). Though CCTs differ depending on context, they always involve transfers of cash from the state to poor households in order to support their economic well-being. To receive these transfers, however, families must comply with expectations “deemed to be in the broader public interest.” Generally, children must attend school and families must participate in nutritional and healthcare programs (Lomelí, 2008:476; Handa & Davis, 2006). The immediate economic relief provided by the transfer, combined with the continual conditions regarding health and education, allow CCTs to be regarded as both a means of poverty-alleviation and of human capital development in a community (Handa & Davis, 2006).

It is this simultaneous short- and long-term approach that make CCTs at once attractive to policy makers and yet difficult to evaluate. Reimers, DeShano da Silva, and Trevino (2006) elaborate on the “black box” that exists in education programs in development: Policy makers often assert that with increased years of education and higher attendance, children will become better-educated and more capable adults. However, with education-focused policies (such as CCTs), the inputs and outputs are made very clear, while the connecting theory, which encompasses multiple influencing factors, is often made invisible. Because of this, CCTs are caught up in an ongoing debate over whether demand-side interventions (such as cash transfers) are sufficient for boosting a country’s educational outcomes. Some scholars argue that supply-side constraints are the true impeding factor, and that without well-trained teachers or quality curriculum, CCT programs are at best moot (Lomelí, 2008; Handa & Davis, 2006).

Impact of CCTs on Education: What We Know So Far

Now, two decades since the start of the first CCT in Latin America, the impact of CCTs on educational achievement¹ is still insufficiently demonstrated. Interestingly, most CCTs were set up with evaluation in mind, using a randomized implementation method to allow for an experimental evaluation. Thus, the first generation of CCTs was followed by a flood of randomized control trials (RCTs) evaluating the programs' targeting, delivery, and short-term outcomes (Rawlings & Rubio, 2005). As a result, CCTs were largely deemed effective at increasing school attendance and enrollment (Reimers et al., 2006) and improving child health and nutrition (Rawlings & Rubio, 2005).

However, the implications of randomly withholding benefits for research purposes often make randomized implementation politically risky and thus unsustainable. In the case of *Progresa* in Mexico, for example, the randomized implementation incited accusations that the government was intentionally withholding benefits from needy families for the purposes of experimentation, resulting in a much quicker roll-out than originally planned for the evaluation (Handa & Davis, 2006). When randomization can only be carried out for the short-term, the ability to measure longer-term effects is limited (Rawlings & Rubio, 2005). In the case of CCTs, when half of their charm is owed to the potential for human capital development and thus increased lifetime earnings, randomized control trials are not sufficient for accurately measuring the impact of the transfers (Handa & Davis, 2006).

It is likely this dependence on RCTs that accounts for the dearth of literature on the long-term effects of conditional cash transfers. Nonetheless, a few evaluations have studied the effects of CCTs on children's educational achievement. One such study in Ecuador found positive impact on the mathematics and language scores of children enrolled in *Bono de Desarrollo Humano* (BDH), one year and a half into participation. The differences, however, were not statistically significant (Ponce & Bedi, 2010). The authors conclude that expectations of CCTs to bolster learning outcomes are unfounded, and that supply-side interventions would be a better alternative.

¹ In this paper, I refer to educational *achievement* as distinct from educational *attainment*; the former being an indicator of cognitive development (often measured using test scores), and the latter being an indicator

Similarly, Behrman et al. (2000) find no significant difference when comparing the cognitive abilities of *Progresa* beneficiaries and their counterparts one year and a half into their participation. A follow up study, conducted after five years of participation, concluded with the same results (Behrman et al., 2005). Still today, one of the least studied aspects of CCTs is their effect on learning outcomes, with few evaluations able to document any significant results (Ponce & Bedi, 2010; Lomelí, 2008).

Indeed, in a systematic review conducted by Baird and colleagues (2014), the authors find just five studies of CCTs that report impact on test scores, only one of which was in Latin America² (Nicaragua's *Red de Proteccion Social*, which showed modest gains in educational achievement (Barham et al., 2014)), leading the authors to conclude that the effects of CCTs on educational achievement are "small at best." Clearly, there is a distressing lack of evidence on the educational impacts of CCTs, despite their global popularity and proponents' insistence that the programs will have positive long-term impacts.

What *can* we say about the educational impacts of CCTs? Their effects on enrollment and attendance appear promising, though as Handa & Davis (2006) point out, this proves little more than effective implementation, since both enrollment and attendance are conditions for participation. Furthermore, with poor school quality, these outcomes may not result in any change in the educational or economic status of children, thus subverting the principal objectives of CCTs. The next stage of evaluations must use alternative evaluation methods to examine medium- and long-term impacts of CCTs in order to justify their somewhat premature popularity.

***Juntos*, Peru's Conditional Cash Transfer Program**

Peru's own CCT was introduced in early 2005. Called *Juntos*, or "together" in Spanish, the program is targeted at families with children under 14 years old or families with a pregnant woman (Jones et al., 2008; Gajate-Garrido, 2014). Eligible families receive a flat-rate cash transfer of 100 soles

² These limited results were despite the fact that the literature search included studies in English, Spanish, and Portuguese. Search included both experimental and quasi-experimental designs (Baird et al., 2014).

(about US\$30, or 17% of the national minimum wage) in exchange for compliance with certain conditions: all school-aged children must enroll and maintain an 85 percent attendance rate, and younger children and pregnant women must attend regular health and nutrition checks (Gajate-Garrido, 2014; Perova & Vakis, 2009). Through these conditions aimed at incentivizing education, health, and nutrition, the *Juntos* program aims to “build capacities of future generations and break the intergenerational transmission of poverty” (Streuli, 2012:591). Families are allowed to participate in the program for up to eight years, as long as they continue to be eligible (Streuli, 2012).

Eligibility is determined in three stages: geographic, household, and community. First, communities are selected according to their level of extreme poverty, access to services, and history of political violence. Then households within participating communities are visited and deemed eligible using a social demographic questionnaire. Lastly, members of the community, local authorities, and representatives from government departments meet to discuss any inclusions or exclusions that need to be made, based on factors unidentified with the questionnaire. For example, those who have substantial wealth not identified by the survey, or those whose primary place of residence is outside of the community, are excluded. Any other impoverished families who are erroneously missed by surveyors are identified and included (Jones et al., 2008).

Within just three years, *Juntos* was implemented in 638 of the poorest districts in the country (Streuli, 2012). The program was placed under a centralized department while integrating representation from government departments of education, health, and social development. Additionally, community-level program facilitators help to monitor school attendance as well as attempt to ensure educational quality by organizing teacher training initiatives as a response to increased enrollment (Jones et al., 2008). Every three months, local coordinators conduct home visits and review school and clinic records in order to monitor compliance (Streuli, 2012).

The Impact of Juntos on Education

Juntos provides a unique case study for CCT evaluation for two reasons. First, *Juntos* did not integrate an evaluation design into its original

implementation. Thus, input on its effectiveness has been relatively quiet compared to other programs with a social experiment design. Second, Peru already had relatively high enrollment and attendance rates before *Juntos* began, with rates around 80 percent for both measures (Perova & Vakis, 2009). Thus, the largest educational impacts of *Juntos* will not likely be found in the area of enrollment or attendance, as those measurements are already high and have little space to improve. Researchers interested in the educational impact of *Juntos* must instead look beyond these outcome variables to more telling indicators such as educational achievement. If *Juntos* does indeed make an impact on education, it is most likely to be found by measuring test scores rather than attendance rates.

A few qualitative studies have led the way in exploring the impact of *Juntos*. In studies conducted in 2008 and 2012, parents, especially fathers, indicated they had taken more responsibility for their children's education as a result of the program, suggesting that *Juntos* has positive effects on the educational achievement of children through increased parental involvement (Jones et al., 2008; Streuli, 2012). Furthermore, respondents suggested that the necessity to monitor children's school attendance acted as an indirect incentive for teacher attendance. However, the increased attendance led to the overcrowding of classrooms, leading some to complain that school quality was not sufficient for demand (Jones et al., 2008).

Perova and Vakis (2009) are cited as being the first quantitative impact evaluation of *Juntos*. Using the same quasi-experimental method employed for this paper, the authors investigate the program's effects two years after its implementation. They find no statistically significant change in attendance rates, but do find a small impact on enrollment, suggesting limited educational impact overall. Furthermore, the authors discover that impact is concentrated at key points during a child's educational trajectory. *Juntos* appears effective when examining the decision to enroll a child in primary school, and the decision to transition from primary to secondary. Because enrollment in Peru is high at baseline, the impact of the program is only visible at crucial points when the continued enrollment of children represents a smaller opportunity cost that the transfer is able to offset. In other words, the transfer is just enough to convince families to finish

milestones already deemed as more valuable; namely entering, and later finishing, primary school.

A second study by the same authors, conducted 3 years later, found somewhat differing results by using an instrumental variable to mediate selection bias. This time, no significant impacts were found on enrollment. Beneficiaries who *were* enrolled were more likely to be attending school than enrolled non-beneficiaries – though the attendance rates of the latter were already impressively high at 86 percent (Perova & Vakis, 2012). Disappointingly, despite the authors' claim that the research measures both short- and long-term impacts, no attempt was made to observe impacts on cognitive development.

In sum, qualitative research suggests positive effects of *Juntos* on education, notwithstanding the supply-side adjustments that appear lacking. However, quantitative research cannot conclude that *Juntos* has overall positive effects on enrollment or attendance, though there is evidence of both positive and negative differential impacts. Furthermore, the available studies of *Juntos* are limited to only the above educational indicators, arguably providing mere outcome evaluations rather than evaluations of impact. The "ultimate test of success," as described by Reimers et al. (2006:5), has yet to be conducted on the *Juntos* program: that of examining the actual life chances of children.

The current study seeks to fill a long-overlooked gap in CCT research by evaluating the long-term impacts of such programs. This paper represents one of just a handful of studies on the impact of CCTs on educational achievement, and it is the first evaluation of its kind for Peru's CCT, *Juntos*.

Examining Differential Effects: Maternal Education and Child Outcomes

Given that one key aim of *Juntos* is to "break the intergenerational transmission of poverty," it serves to ask if the program is helping the education of children whose parents were, for various reasons, left out of educational opportunities. Indeed, a true testament to the success of *Juntos* would be evidence that it both increases educational achievement, and that this impact is greater for those whose parents had lower levels of education. For this reason, the current study will examine the differential effects of

Juntos on child beneficiaries whose mothers completed primary school, compared to those whose mothers did not.

There are two reasons for using maternal education as the measure of choice. The first is that *Juntos* directly targets families in which the mothers have lower levels of education (SISFOH, 2013). The second reason is made clear in the literature: the effect of maternal education on children's outcomes such as health and education is well documented. In Peru, maternal education is one of the most influential factors on the health of a child, particularly in non-coastal regions (that is, where *Juntos* is implemented) (Shin, 2007). One study by Urke and colleagues (2011) found that the children of mothers who had completed primary school had a stunting prevalence of 30 percent, compared to a 48 percent prevalence rate among children whose mothers had not completed primary. The study illustrates that the importance of maternal education in a child's life trajectory remains persistent, even after controlling for other key socio-economic indicators.

Of course, the effects on a child's health go hand in hand with his or her educational outlook. Children who are stunted due to poor health or nutrition have lower cognitive capacity and are inhibited in their abilities to succeed in school (Shin, 2007). However, maternal education affects a child's education through other means as well. In another study in Peru, researchers found that the effect of parental education on a child's test scores was realized through increased parental expectations. That is, the higher a mother's educational status, the more she expects her children to succeed in school as well, encouraging better educational outcomes for each child (Castro et al., 2002). Fernald et al. (2012) expand these results to even younger children, explaining that even at just 2 years old, Peruvian children with highly educated mothers perform significantly better on cognitive development tests, holding other relevant covariates constant.

One can see how these results may set up an increasing educational inequality in Peru. If those who are well educated have children who are more educated, and those who are not have children whose education suffers, then the gap between the two groups will continue to increase. This in turn may have serious consequences for income inequality in the long-term. However, a CCT program such as *Juntos* may have the opportunity of

bridging such a gap by positively impacting the educational achievement of beneficiaries with less-educated mothers enough to “catch up” to their peers.

Data

Data for the current research comes from the Peru section of Young Lives, an international study of poverty examining the life trajectories of children in four countries. The project follows two cohorts of children in Peru: a younger cohort of around 2000 children, and an older cohort of an additional 1000 children. Known locally as Niños del Milenio, the first round of Young Lives surveys in Peru took place in 2002, with the ages of the younger cohort ranging from 6 to 18 months, and the older cohort aged 7 to 8 years. The sampling strategy is a two stage randomized process, first randomly selecting 20 “sentinel sites,” and then taking a random sample of households within each site. The representativeness of the dataset is discussed later in this paper.

Each year, household surveys are conducted to gather information on household composition, consumption and expenditure, recent life history, child health and access to basic services, as well as parental background and education. The second round of surveys took place in 2006, just one year after the beginning of *Juntos*, and surveys now include a variety of questions on *Juntos* benefits and transfer use. This research makes use of the third round, conducted in 2009 when children were 7 and 8 years old. It is during this round when mathematics tests were first introduced to the survey for this cohort³. Data from round one is also used in order to provide a baseline that is not impacted by the cash transfer.

The Young Lives Peru Study is a collaborative effort between the University of Oxford, Save the Children Fund, and Grupo de Análisis para el Desarrollo (GRADE) and Instituto de Investigación Nutricional (IIN) in Peru. Data are analyzed using Stata version 14.0 (StataCorp., 2015).

³ Mathematics test scores were chosen as educational achievement indicators rather than scores for reading and vocabulary, as mathematics scores had the highest response rate. Because the families participating in the surveys were chosen randomly, high response rates help to ensure the results are still representative of the larger population.

Methods

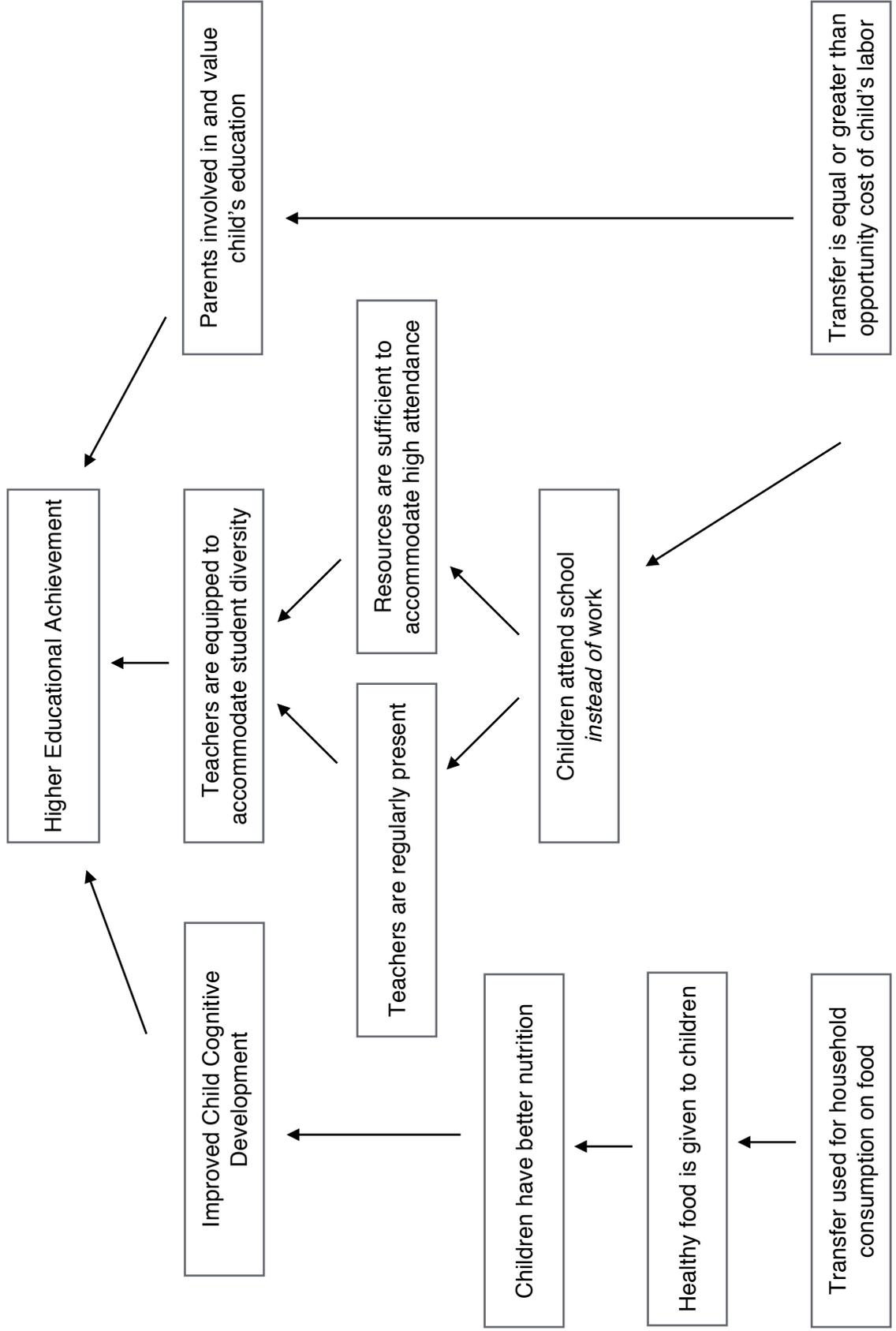
Theory of Change

Few, if any, CCT programs were created with clearly constructed theory of change models in mind. Even more interestingly, Reimers and colleagues (2006) study a collection of CCTs from around the world and note that none clearly mentioned improved learning as an aim, despite their insistence that short-term education incentives will result in long-term human capital accumulation. The current research, therefore, maps out an original theory of change based both on program literature and outcome evaluations conducted from various Latin American CCTs (see Figure 1). As Perova and Vakis (2009) point out, Peru had high attendance rates (around 80 percent) even before *Juntos* was initiated. Thus, one portion common to CCT theories of change — that children attend school instead of working, due to the transfer — is only one of multiple ways in which *Juntos* families could experience positive educational impacts. The theory of change will also focus on other mechanisms by which *Juntos* could improve test scores.

A few of these mechanisms are identified in Ponce and Bedi's (2010) evaluation of BDH in Ecuador. The cash transfer, when used for increased consumption on food (and coupled with informative nutritional check-ups) will result in better cognitive development for young children. Secondly, if the transfer is more than enough to compensate for loss of a child's labor, it will allow the child to give up the labor and find extra time and energy to achieve better grades.

These casual pathways, of course, assume that there is adequate training and resources in schools that serve beneficiary children. Particularly in schools with high student diversity, schools must be prepared with bilingual and culturally appropriate material. Poor educational management could instead result in no impact, or even a negative impact on educational achievement, as even a slight increase in attendance due to *Juntos* could attribute to overcrowding of schools. And, as *Juntos* is not directly managed by the ministry of Education, it is possible for delays in improvements when these challenges are identified (Jones et al., 2008; Reimers et al., 2006).

Figure 1. Theory of Change:
Juntos and Educational Achievement



Establishing Treatment and Control Groups

As a first step to the evaluation, the data is limited to households who live within regions who qualify for the transfer, but who have not yet received it due to the timing of the program's rollout, or to a fault of targeting. Other evaluations of *Juntos* have found this to be the most appropriate way of assessing impact (Perova & Vakis, 2009; Gajate-Garrido, 2014). This effectively eliminates richer households, but also families who might qualify for the program at a household-level, but who live within better-off neighborhoods. As both the propensity score (discussed below) and the regression include only household-level covariates, limiting the comparison group helps to control for these regional differences. Furthermore, establishing the control group as such allows us to make a meaningful interpretation of the Average Treatment Effect (ATE), which calculates the effect that the program would have on the entire population sampled. As the control is now limited to all those who should qualify for the program, the ATE is an appropriate and illuminating measure of impact. Secondly, as this research is focused on long-term educational impacts, children who have only recently begun participating in the study will be dropped from the sample. The study limits the treatment group to those children who are in their second year of the program or higher.

Propensity Score Weighting

Because *Juntos* was not implemented randomly, there exists inevitable selection bias between those in the treatment and control. Though it is difficult to identify and control for all factors that may have influenced a household's selection into the treatment group, propensity score weighting is one method that can help to correct for omitted-variable bias (Freedman & Berk, 2008).

Propensity score weighting helps to mimic randomized assignment by assigning weights to each observation that correspond to their probability of participation (Stuart, 2010; Arteaga et al., 2014). First, a probability model (either logit or probit) is used to predict each observation's selection to the treatment based on a set of covariates that influence participation. Then, using these probabilities, or propensity scores, weights are assigned to each observation such that higher weights are given to observations whose

treatment status is underrepresented in the sample. For example, an observation in the control group that has a high probability of treatment (say, a propensity score of 0.75) will be weighted more heavily, as this control observation is likely more similar to treatment observations than others with lower propensity scores. This method is preferable to traditional ordinary least squares regression in our case, because it minimizes the weights inherently given to variables that are 'outliers' or farther from the majority of the sample (Freedman & Berk, 2008). Propensity score weighting is very similar to its methodological cousin, propensity score matching (PSM), in this aspect (Caliendo & Kopeinig, 2005).

Though there are various types of weights that can be used with this method, the current study uses weights that measure the average treatment affect (ATE), as outlined in Stuart (2010) and Freedman and Berk (2008):

$$\text{Weight } w = T/p + (1-T)/(1-p)$$

Where T is a dummy variable such that a value of 1 indicates assignment to treatment, and p is the observation's propensity score. The above weighting scheme reflects an inverse probability of treatment weighting (IPTW) such that observations in the control group with high propensity scores are weighted more heavily, while observations in the treatment group with low propensity scores are weighted less heavily.

Recently, social sciences have used propensity score weighting to evaluate educational outcomes, particularly when interventions have differential impacts. Arteaga and colleagues (2014) test the marginal effects of preschool attendance among different at-risk subgroups, testing their educational achievements (as measured through reading and math scores) in adulthood. In a study of Brazil's *Bolsa Familia*, the largest CCT in the world, De Brauw and colleagues (2015) measure the program's impact on grade progression, repetition, and dropout rates. They use propensity score weighting while disaggregating by the age and sex of the beneficiaries. Indeed, one of the benefits of propensity score weighting compared to its popular counterpart, PSM, is that the method allows for interactions between treatment and covariates in order to tease out differential effects on subgroups (Stuart, 2010). Thus, this method is particularly appropriate for

the current research as it allows for mediation of the selection bias of the program and for the observation of differential impacts between different levels of maternal education.

Assumptions

Admittedly, this method comes with a few strong assumptions. Most notable is the “ignorable treatment” assumption, which asserts that after controlling for observable differences in the probability model, the potential outcomes are independent from treatment assignment (Stuart, 2010). The variables used in the probability model will simultaneously influence the treatment and the outcome of interest (Caliendo & Kopeinig, 2005). This of course implies that researchers are able to observe the important variables that influence the probability of selection into treatment (and that any unobserved bias is also corrected in the process). At the same time, the sample must exhibit overlap in the distribution of the propensity score between treatment and control, so that the groups are similar enough to make comparisons (Stuart, 2010).

Freedman and Berk remark, “the costs of misapplying the technique, in terms of bias and variance, can be serious” (2008: 392). They suggest that two questions can determine the suitability of the method for a particular data set: First, is there selection bias in the causal model? Second, does the data give enough information to estimate the propensity scores with accuracy? Given the non-random assignment of the program in question, there is certainly selection bias between treatment and control in our sample. However, as described earlier, the data set in use is sufficiently rich to provide good predictions of the propensity score while maintaining overlap between treatment and control.

Results

Probability Model and Inverse Probability of Treatment Weights

Variables from the probability function come from the selection criteria for *Juntos* eligibility (SISFOH, 2013) and are corroborated with studies of educational attainment in Peru (see Perova and Vakis, 2009; Jones et al.,

2008; Gajate-Garrido, 2014). In keeping with the ignorable treatment assumption, indicators of wealth and access to services come from round one of the survey (conducted in 2002), and as such, are not affected by treatment. See Table 1 the probit regression, and Figure 2 for a graph of the area of common support.

Upon first assigning weights, some control observations carry extreme weights – one as high as 103.54, meaning that this unique observation is representing over 103 observations. This is a result of a control observation being, in essence, *too* similar to treatment observations. That is, when the propensity score of a control observation is very close to one, its weight will be very large. For example, a control observation with a propensity score of .99 would yield:

$$w = 0/.99 + (1-0)/(1-.99) = 100$$

This carries consequences for our estimator: if the variance of the weights is high, the estimator may be imprecise, as just a handful of control observations will be disproportionately influential to the results (Arteaga et al., 2014; Stuart, 2010). To limit the influence of these high weighted observations, the weights were trimmed to the lowest level possible while still maintaining the balance between groups. After trimming, the highest weight in the control group was 65 (2 observations), with just 8 control observations assigned a weight over 20. Table 2 shows the balance of covariates between the treated and control groups after weighting. All but one of the covariates used in the probability model are well-balanced, though treated individuals have on average lower consumer durables index scores. As this difference is only slightly significant (at the $p < .10$ level), the analysis continues assuming covariates are balanced.

Table 1. Probit model used for propensity scores

VARIABLES	TREATMENT
INDIGENOUS (dummy)	1.532*** (0.119)
HOUSEHOLD SIZE (number of members)	0.0745*** (0.0248)
MATERNAL EDUCATION (dummy with 1 = mother completed primary)	-0.495*** (0.118)
SEX (dummy with 1 = Female)	-0.115 (0.0906)
CONSUMER DURABLES INDEX ⁴ , SQUARED (scale, 0 to 1)	-3.922*** (1.089)
ACCESS TO ELECTRICITY (dummy with 1 = has access)	-0.341*** (0.106)
ACCESS TO SANITATION (dummy with 1 = has access)	-0.372*** (0.095)
ACCESS TO COOKING FUEL (dummy with 1 = has access)	-0.314* (0.177)
Constant	-0.500 (0.215)
Observations	1,445
Pseudo-R2	.3666

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

⁴ Calculated mean composed of dummies indicating ownership of household items. Squared functional form allows for balance between treatment and control groups (when left un-squared, imbalance remains between groups).

Figure 2. Histogram showing area of common support

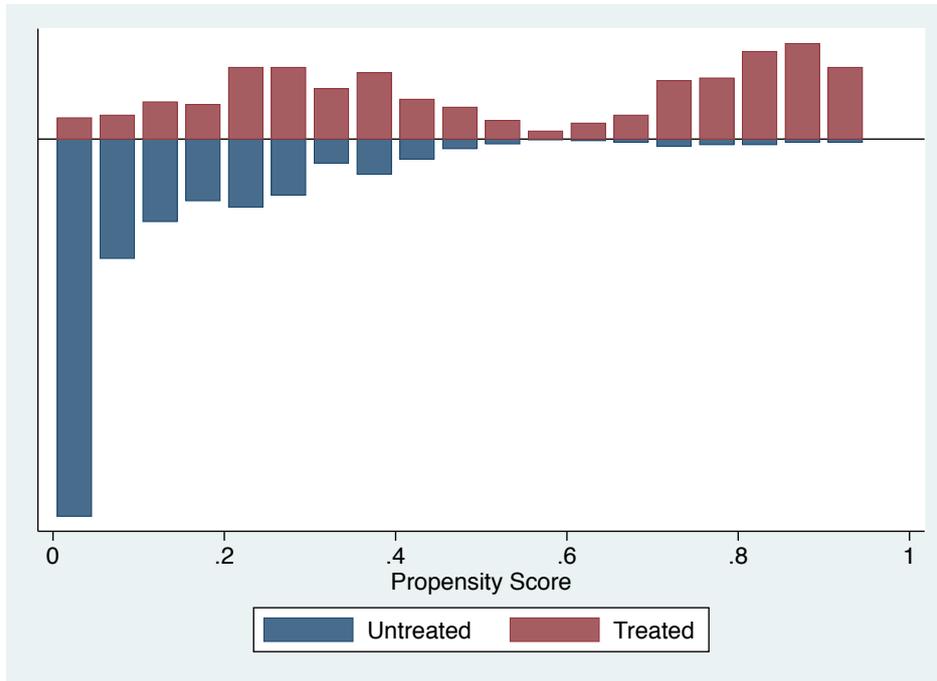


Table 2. Balance of Covariates in sample after weighting

VARIABLE	Mean in Treated	Mean in Untreated	P> t
INDIGENOUS	0.16	0.14	.578
HOUSEHOLD SIZE	5.75	5.48	.185
MATERNAL EDUCATION	0.35	0.43	.283
SEX	0.51	0.50	.945
CONSUMER DURABLES	0.18	0.23	.071
ACCESS TO ELECTRICITY	0.52	0.56	.483
ACCESS TO SANITATION	0.63	0.72	.123
ACCESS TO FUEL	0.25	0.30	.500

(weights trimmed to max = 65)

Regression Analysis

While the variables used to estimate the propensity score could theoretically also be included in the regression analysis, this is redundant as the means of these covariates should already be balanced between treatment and control groups upon weighting (Freedman and Berk, 2008). Thus, the current regression analysis includes only the variables of interest (that is, the treatment variable and the variable for maternal education). Additionally, the model includes regional dummy variables, as geographic region has been shown to moderate the effect of maternal education on educational outcomes (Shin, 2007). Table 3 gives the results of the regression.

Note that the regional dummy variables were not used in the propensity score estimation. This is due to the design of the program implementation (and thus the design of the evaluation) as described in a previous section, Establishing Treatment and Control Groups. Because *Juntos* was implemented by region, it follows that a region covariate would be a very strong predictor of program participation. However, there exist too few cases in which treated and untreated individuals share the same region. Estimating propensity scores using this covariate would result in these few individuals being weighed very heavily, without accurately representing the other individual- and household-level covariates.

Table 3. Regression using inverse probability of treatment weights

VARIABLES	Math Score	Math Score
TREATMENT	-0.190 (0.974)	0.133 (0.936)
MATERNAL EDUCATION	4.256*** (0.390)	3.241*** (0.414)
TREATMENT*MATERNAL EDUCATION	-4.577*** (1.562)	-3.877** (1.835)
REGION2		0.386 (0.897)
REGION3		-0.290 (0.958)
REGION4		1.554 (0.964)
REGION5		1.513 (0.930)
REGION6		1.396 (0.938)
REGION7		4.487*** (1.053)
REGION8		4.264*** (0.885)
REGION9		0.438 (0.993)
REGION10		3.974*** (0.902)
REGION11		0.135 (1.107)
REGION15		0.409 (1.173)
REGION16		1.250 (1.177)
REGION17		2.436 (1.968)
REGION18		-0.900 (1.112)
Constant	11.80*** (0.321)	10.74*** (0.811)
Observations	1,414	1,414
R-squared	0.102	0.163

Robust standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Two things are of note: first, the coefficient on the single treatment variable is small and not statistically significant, indicating that participation in *Juntos* alone likely does not affect a child’s mathematics scores. However, the coefficient for the interaction between maternal education and treatment is both large and statistically significant at the $p < .05$ level. This is the coefficient of interest when measuring the differential impact of *Juntos* between children with mothers that have low levels of education, and children whose mothers have higher levels of education. The results here are surprising: the negative coefficient on the interaction suggests that child beneficiaries whose mothers completed at least primary school are *negatively* impacted by enrollment in *Juntos*, in terms of their test scores.

The magnitude of this negative impact is not immediately obvious from our regression, however. Table 4 gives the marginal effects of treatment by maternal education, and provides a more intuitive interpretation of the results. Children enrolled in *Juntos* whose mothers did not complete primary school have math scores that do not differ significantly from their untreated counterparts. However, for children whose mothers did complete at least primary school, we can expect the test scores of *Juntos* beneficiaries to be 3.67 points lower on average than similar students in the control group. This difference is significant at the $p < .01$ level.

Table 4. Marginal effects and Predictive Margins for interaction of MATERNAL EDUCATION and TREATMENT

Marginal Effects				
MATERNAL EDUCATION		dy/dx TREATMENT	t	P>t
Mother did not complete primary	0	0.1328982	0.14	0.887
Mother completed primary	1	-3.743868	-2.70	.007
Predictive Margins				
MATERNAL EDUCATION		TREATED	UNTREATED	
Mother did not complete primary	0	12.06534	11.93245	
Mother completed primary	1	11.42994	15.17380	

In other words, rather than disadvantaged children catching up to their peers as one would hope, the opposite is happening: the test scores of treated children whose mothers completed primary school are *sinking* to the level of their peers with low-educated mothers. What could explain this effect?

As mentioned in the literature review, a qualitative evaluation of *Juntos* found that the overcrowding of classrooms might be a serious unintended consequence of *Juntos*, resulting in a lower quality of schools. Several key informants questioned the mandatory school attendance provision of the transfer, and whether it had any influence on a child's ability to learn, given the lack of teachers available to handle increased attendance (Jones et al., 2008). When classrooms are overcrowded, teachers must make decisions about how to allocate limited resources, and more specifically, to which students these resources are allocated. Such situations set the stage for programs to have differential effects on students. While the Young Lives data set does not include information on the number of children enrolled in each school, nor the number of teachers available, the household survey does include a question for the child's caregiver regarding their opinion of the quality of education the child receives.

The question is as follows: "How would you rate the quality of teaching at (child)'s school?" with possible responses of "all or most are very good teachers," "the majority are very good teachers," "only a few are good teachers," or "none or almost no teachers are good." The question was recoded into a dummy variable with the first two responses coded as 1 for good quality, and the last two responses coded as 0 indicating bad quality. Then, the responses were averaged across each region and used as a proxy for that region's school quality. Comparing between treatment and control groups, while maintaining the weights, shows a significant difference in the quality of schooling (see Table 5). Children in the control group are more likely to be in schools deemed as having high quality teaching, with a mean quality ranking of 0.57 compared to 0.48 in the treatment group.

Table 5. Means of reported regional school quality, by treatment

VARIABLE	Mean in Treated	Mean in Untreated	P> t
SCHOOL QUALITY	0.48	0.57	.001

Sample restricted to observations in regression. Adjusted for clustering at region level.

This would provide a good explanation, had our results shown a negative impact on *all* students. However, it does not fully explain the findings observed earlier: the *differential* effects between students. One clue of the mechanisms behind the adverse affects found in this study lies in the margins shown previously in Table 4. Interestingly, the difference in test scores between treated children with low maternal education and treated children with higher maternal education is almost nonexistent. That is, children with high maternal education are not only doing worse than their untreated counterparts, their scores are exactly low enough to be level with their classmates. This finding hints at the dynamics inside the classroom: with the increase in attendance and added challenges of overcrowding, teachers are likely teaching to the level of the low performers, and thus stunting the educational growth of the potential high performers.

These results make sense in the context of *Juntos*, where the primary attendance rates were high (around 80%) even before the introduction of the CCT. It also serves to remember how the control group was constructed. Since *Juntos* was implemented by region to all who qualify, the control group comes from regions in which none of the students receive the transfer (but who will receive *Juntos* when the gradual roll-out of the program arrives to their region). This means two things. First, that the vast majority of the control group is enrolled in school, and second, that they are *not* enrolled in schools with *Juntos* participants. The only portion of the control group who are suffering educationally are the 20 percent or so who remain un-enrolled. In the treatment group, however, a surge of enrollment of even 4 percent (as found by Perova & Vakis, 2009) is enough to create strain on a community's educational resources – particularly when the new students have much higher attendance rates (as found in Perova & Vakis, 2012). In this scenario, those that are suffering educationally are all those who share the classroom.

And when a teacher is teaching to the level of the low-performers, those who suffer the most are those who have the most potential.

Discussion

Juntos and Ability Tracking

The findings of this evaluation contribute to the contemporary debate on ability tracking, or separating students into different classes or schools based on their academic performance. Ability tracking has long been a hotly debated subject in education, though the conversation is largely concentrated in developed countries. Central to the debate is the paradox that tracking (argued by many as beneficial for all students) may reduce the impacts of the 'peer effect,' or the positive spillover effect by which lower performing students learn from their higher performing peers (Zimmer, 2003). Indeed, an evaluation conducted during Finland's conversion from a tracking system to a non-tracking system found that students whose parents had only a basic education experienced positive effects on their arithmetic test scores (Kerr et al., 2013). This suggests that non-tracking systems may actually be better for students from families with lower parental education.

Duflo et al. (2011) claim the first rigorous impact evaluation of school tracking in a developing country. The authors' randomized experiment in Kenya found that the average score of students in tracked schools was 0.14 standard deviations higher than the average score of students in non-tracked schools. However, the authors admit that these results may vary significantly depending on school context, saying "in a system where the incentive is to focus on the weakest students... tracking could have a very strong positive effect on high-achieving students, and a weak or even negative effect on weak students" (Duflo et al., 2011:1770). So while tracking students in Peru that demonstrate high potential (which in Peru are likely those who come from higher educated mothers) may yield better results for higher performing students, a non-tracked system may be beneficial for those who are at risk of falling behind (namely, those with lower educated mothers). This is certainly an area for which more evidence is needed, particularly in the Peruvian and wider Latin American context.

Reliability of Results

To the author's knowledge, no STATA programs are yet available to conduct sensitivity analyses when using propensity score weighting. Thus, reliability can be assessed only to the extent that the assumptions used for the method are logically met. These are discussed below.

First, under the ignorable treatment assumption, selection into treatment must be independent of outcome, conditional on observable variables. Reliability of our findings could thus be challenged if there is reason to believe unobserved variables influence treatment and outcome. However, because data available through Young Lives closely resembled the household-level socio-demographic questionnaire used to determine eligibility, this study was able to closely replicate the selection process in its probability model. With regard to regional-level characteristics, only those regions whose characteristics had already deemed them eligible for the program were used in the study, and the regression model included regional dummy variables to control for variation after weighting. Because of the careful construction of the control groups and the wealth of variables available in the data-set, the ignorable treatment assumption is reasonable. The second assumption – that there exists sufficient overlap between treatment and control groups – was demonstrated to be met in Figure 2.

Questions of external validity of the results can be answered by observing the surveying technique of the Young Lives project. First, it must be noted that the current analysis only evaluates the impact of *Juntos* on children ages 7 and 8 during the year 2009. Furthermore, an evaluation of the Young Lives sampling methods, conducted by Escobar and Flores (2008), reveals that the children surveyed for Young Lives are on average slightly better-off than children sampled in Peru's 2005 census. If the data set leaves out the most needy families, for which a transfer might make a larger difference, it is possible that the current study underestimates the impact of *Juntos*. Lastly, there were 143 observations out of 1,557 that were missing one or more of the variables needed to conduct this analysis. Ten of the 140 missing observations were from the treatment group (out of 335 treated children), all of which were only missing test scores. A comparison between the treated with test scores and those without reveals a statistically

significant difference in indigenous ethnicity and maternal education, shown in Table 6. Treated children whose data is used in the regression were more likely to have mothers who completed primary school and less likely to be indigenous, compared to treated children whose test scores were missing.

Table 6.

VARIABLE	Included in regression	Missing scores	P> t
INDIGENOUS	0.46	0.90	.000
HOUSEHOLD SIZE	6.08	6.30	.724
MATERNAL EDUCATION	0.10	0.00	.000
SEX	0.49	0.50	.964
HOUSING QUALITY	0.24	0.25	.595
ACCESS TO ELECTRICITY	0.27	0.30	.866
ACCESS TO SANITATION	0.52	0.60	.629
WEALTH INDEX	0.20	0.23	.494

Nonetheless, what the data lacks in representativeness, it makes up for in wealth of indicators and availability of a baseline. Particularly because *Juntos* administration does not collect data on non-participants, the Young Lives data set is currently the most appropriate instrument for analyzing program impact.

Implications for Juntos

While the results seem to cast doubt on the effectiveness of *Juntos*, they certainly do not support a removal of the program entirely. Though it does not appear that *Juntos* directly improve children’s test scores, the current research echoes past findings of significant relationship between a child’s educational achievement and his or her mother’s education level (holding all things constant, maternal education was significantly correlated with higher test scores– see Table 3 on page 20).

This in mind, the impact that increased enrollment has on the next generation of children can not be underestimated. Though *Juntos* may not be improving educational performance in the current child beneficiaries, previous research shows that it *has* increased enrollment, and in turn is educating a new generation of mothers and fathers. If maternal education continues to have a strong influences on children's educational outcomes in Peru, it may mean researchers must only wait to observe the next generation of children before the long-term educational impact of *Juntos* is apparent. This, of course, depends on Peru's ability to improve the educational quality of schools over time – providing adequate training and resources for the increased enrollment that the educational system will no doubt experience with the continued roll-out of the program. In this aspect, this paper's findings in part support critiques of CCTs which call for more supply-side interventions.

Results also hint that tracking may be an effective approach to harness the potential of *Juntos* to have positive impacts. Additional research is needed to understand if tracking is a viable option in the Peruvian context, and if it would help to mediate the differential impacts of *Juntos* without harming lower-performing students.

Conclusion

This research took advantage of the gradual roll-out of Peru's conditional cash transfer *Juntos* to assess the progress the program is making on its long-term educational goals, using quasi-experimental methods and a carefully constructed control group. Results are surprising: though the average treatment effect on the whole population appears negligible, participation in *Juntos* has differential impacts, depending on the education level of the child's mother. For children whose mothers completed primary school, enrollment in the CCT has a negative impact on test scores, contrary to what would be expected for that demographic.

The distribution of test scores between the subgroups and evidence collected in a household survey hint at the underlying cause for such an effect. Teachers are likely left without adequate training or resources to

handle a sudden increase in attendance, and instead cater lessons to the lower performing students, while stunting the educational growth of others. On the other hand, the vast majority students in the control group are enrolled in schools without *Juntos* beneficiaries, as the program has not yet been rolled out to their region, and thus do not suffer from overcrowded classrooms. Higher performing students are free to excel in the control environment, thus creating the drastic and statistically significant difference between themselves and their counterparts in treated communities.

However, previous studies emphasize the importance of maternal education on a child's educational achievement, and the current research echoes such findings. *Juntos* has been shown to increase enrollment, and will thus contribute to an increase in maternal education in the long-run. So while students aren't excelling in achievement, mere attainment of higher schooling levels may be enough to encourage long-term educational improvement and poverty reduction, assuming supply-side adjustments are made. To the author's knowledge, this is the first study to evaluate the impact of *Juntos* on children's educational achievement, and more evaluations are needed. Future research should focus on inconsistencies in school quality as a result of *Juntos*, and the outcomes of the children of today's young beneficiaries, in order to further assess the ability of *Juntos* to tackle the intergenerational cycle of poverty.

References

Arteaga, I., Humpage, S., Reynolds, A. and Temple, J., 2014. One year of preschool or two: Is it important for adult outcomes? *Economics of Education Review*, 40, pp. 221-237.

Baird, S., Ferreira, F. H. G., Özler, B. and Woolcock, M., 2014. Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness*, 6(1), pp.1-43.

Barham, T., Macours, K., Maluccio, J. A., Regalia, F., Aguilera, V. and Moncada, M. E., 2014. *Assessing long-term impacts of conditional cash transfers on children and young adults in rural Nicaragua. 3ie Impact Evaluation Report 17*. New Delhi: International Initiative for Impact Evaluation (3ie).

Behrman, J., Sengupta, P. and Todd, P., 2005. Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico. *Economic Development and Cultural Change*, 54(1), pp. 237-275.

Caliendo, M. and Kopeinig, S., 2005. Some Practical Guidance for the Implementation of Propensity Score Matching. *Forschungsinstitut zur Zukunft der Arbeit (IZA) Discussion Paper No. 1588*, May.

Castro, D., Lubker, B., Bryant, D. and Skinner, M., 2002. Oral language and reading abilities of first-grade Peruvian children: Associations with child and family factors. *International Journal of Behavioral Development*, 26(4), pp. 334-344.

De Brauw, A., Gilligan, D., Hoddinott, J. and Roy, S., 2015. The Impact of *Bolsa Familia* on Schooling. *World Development*, 70, pp. 303-316.

Duflo, E., Dupas, P. and Kremer, M., 2011. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5), pp. 1739-1774.

Escobal, J. and Flores, E., 2008. An Assessment of the Young Lives Sampling Approach in Peru. Young Lives Technical Note No. 3, March.

Fernald, L., Kariger, P., Hidrobo, M. and Gertler, P., 2012. Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal. *Proceedings of the National Academy of Sciences*, 109(2), pp. 17273-17280.

Freedman, D. and Berk, R., 2008. Weighting Regressions by Propensity Scores. *Evaluation Review*, 32(4), pp. 392-409.

Gajate-Garrido, G., 2014. Assessing the Differential Impact of "Juntos" Conditional Cash Transfer on Indigenous Peoples. In Population Association of America, Session 178 on Safety net programs and their impacts on human capital investments and learning outcomes. Boston, MA, 1-3 May 2014. Population Association of America: Boston.

Handa, S. and Davis, B., 2006. The Experience of Conditional Cash Transfers in Latin America and the Caribbean. *Development Policy Review*, 24(5), pp. 512-536.

Jones, N., Vargas, R. and Villar, E., 2008. Cash transfers to tackle childhood poverty and vulnerability: an analysis of Peru's *Juntos* programme. *Environment & Urbanization*, 20(1), pp. 255-273.

Kerr, P. S., Pekkarinen, T. and Uusitalo, R., 2013. School Tracking and Development of Cognitive Skills. *Journal of Labor Economics*, 31(3), pp. 577-602.

Lomelí, E. V., 2008. Conditional Cash Transfers as Social Policy in Latin America: An Assessment of their Contributions and Limitations. *Annual Review of Sociology*, 34, pp.475-499.

Perova, E. and Vakis, R., 2009. Welfare impacts of the “Juntos” Program in Peru: Evidence from a non-experimental evaluation. *The World Bank*, pp. 1-59.

Perova, E. and Vakis, R., 2012. 5 Years in *Juntos*: New Evidence on the Program’s Short and Long-Term Impacts. *Economía*, 35(69), pp.53-82.

Ponce, J. and Bedi, A. S., 2010. The impact of a cash transfer program on cognitive achievement: The *Bono de Desarrollo Humano* of Ecuador. *Economics of Education Review*, 29, pp.116-125.

Rawlings, L. B. and Rubio, G. M., 2005. Evaluating the Impact of Conditional Cash Transfer Programs. *The World Bank Research Observer*, 20(1), pp. 29-55.

Reimers, F., DeShano da Silva, C. and Trevino, E., 2006. Where is the “Education” in Conditional Cash Transfers in Education? UNESCO Institute for Statistics Working Paper No. 4, Montreal.

Shin, H., 2007. Child Health in Peru: Importance of Regional Variation and Community Effects on Children’s Height and Weight. *Journal of Health and Social Behavior*, 48, pp.418-433.

Sistema de Focalización de Hogares (SISFOH), 2013. Ficha Socioeconómica Única. [Online]. Available at <http://www.sisfoh.gob.pe/nosotros.shtml?x=1478> [accessed 1 April 2015]

StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Streuli, N., 2012. Child protection: a role for conditional cash transfer programmes? *Development in Practice*, 22(4), pp. 588-599.

Stuart, E., 2010. Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, 25(1), pp. 1-21.

Urke, H., Bull, T. and Mittelmark, M., 2011. Socioeconomic status and chronic child malnutrition: wealth and maternal education matter more in the Peruvian Andes than nationally. *Nutrition Research*, 31, pp. 741-747.

Zimmer, R. 2003. A new twist in the educational tracking debate. *Economics of Education Review*, 22, pp. 307-3015.

