No.

Impact Evaluation of Third Elementary Education Project in the Republic of the Philippines

December 2011

JAPAN INTERNATIONAL COOPERATION AGENCY

International Food Policy Research Institute

EVD
JR
11-67

Preface

Japan's ODA charter revised in 2003 shows Japan's commitment to ODA evaluation under the section "Enhancement of Evaluation", stating the importance of objective evaluation on the outcome of ODA projects.

Recently in the context of increasing concern on development outcomes, in order to implement more effective and efficient assistances, impact evaluation that precisely measures change that occurs as the result of an intervention or project implementation is beginning to receive more attention. JICA has been working on promoting the utilization of impact evaluation methods.

The volume shows the results of the impact evaluation of an ODA Loan project, "Third Elementary Education Project" in the Republic of the Philippines. This evaluation was conducted by the International Food Policy Research Institute. The result drawn from the evaluation will be shared with the JICA's stakeholders for the sake of improving the quality of ODA projects.

Lastly, deep appreciation is given to those who have cooperated and supported the creation of this volume of evaluation.

December 2011
Masato Watanabe
Vice President
Japan International Cooperation Agency (JICA)

Disclaimer

This volume of evaluation shows the result of impact evaluations made by external evaluators. The views and recommendations herein do not necessarily reflect the official views and opinions of JICA.

Minor amendments may be made when the volume is posted on JICA's website.

Impact Evaluation Report on Third Elementary Education Project in the Philippines

Futoshi Yamauchi
International Food Policy Research Institute
Washington D.C.

December 2011

Contents

Executive Summary

(a) Brief description of this project

The objective of this study is to assess short-term and long-term impacts of Third Elementary Education Projects in the Philippines. For this purpose, we have collaborated with the Philippine Department of Education and two local collaborators to collect various types of data including unique individual-level data.

The Third Elementary Education Project (TEEP) was implemented from 2000 to 2006 by the Philippine Department of Education in all public primary and elementary schools in the 23 provinces identified as the most socially depressed in the Social Reform Agenda.[1] The total project cost was US$221.16 ($91.07 million from JICA[2] and $82.84 million from World Bank, $47.25 from the Philippine government). The unique feature of TEEP is a combination of investments in school facility and education materials and school governance reform. Not only were school facilities and textbook supply improved, but the decision making process was also decentralized to the school and community levels. TEEP introduced a package of investments to schools in the selected 23 provinces. Specifically, the package of investments included (1) school building construction and renovation, (2) textbooks, (3) teacher training, (4) school-based management, and (5) other facility and equipment support.

In the short-term impact study, we use school-level database including national achievement test data to assess the TEEP impact on students' learning achievements. In the long-term impact study, we quantify the impact on students' subsequent schooling outcomes, migration behavior, and labor market earnings by conducting a unique tracking survey.

For the long-term impact analysis, we have conducted surveys simultaneously in 8 provinces: Ifugao, Neuva Vizcaya, Antique, Iloilo, Negros Oriental, Cebu, Leyte and Western Samar to capture schooling and work histories of former students. Ifugao, Antique, Negros Oriental and Leyte are TEEP divisions, and the others are non-TEEP divisions. Basic idea of identification is to compare TEEP and non-TEEP divisions in the same area. For example, Negros Oriental (treatment) and southern part of Cebu (comparison), both adjacent to each other sharing similar socio-economic situations, are compared for identifying TEEP impacts. In the tracking survey, we successfully collected information on schooling and work histories from 3451 students from 101 schools (Pre-TEEP and TEEP cohorts in both TEEP and non-TEEP divisions).

(b) Narrative explanation about main findings

In the short-term impact analysis, because the original phase-in plan of TEEP was not followed in practice, we cannot explore the pipeline design to identify the impact of TEEP on school

[1] The loan agreement was signed in 1997. Due to delays experienced in its preparation, actual implementations started in 2000/2001.
[2] At that time, it was JBIC, which merged to JICA at a later stage.

performance. Therefore, we formed a control group based on the schools in the non-TEEP provinces to estimate the counterfactual of the treatment group, which are the schools in the TEEP provinces. Double differences (DD) based on the cohort panel from grade 4 (SY 2002/03) and grade 6 (SY 2004/05) is used to eliminate cohort-specific fixed effects. Since the allocation of TEEP was purposive, the initial school conditions are likely to have different distributions in the treatment and control groups. If the initial conditions affect subsequent changes of the outcome variables, DD would give a biased estimate of the TEEP impacts.

We use two strategies to deal with the potential bias due to nonrandom program placement. First, we use the sample from Visayas only. TEEP divisions are relatively evenly distributed throughout Visayas compared with the other two macroregions. The TEEP and non-TEEP provinces are more comparable in Visayas. Second, we use propensity score (PS) matching to balance observable cohort characteristics and initial conditions between the treated and the control groups.

The short-term impact analysis demonstrated significant impacts of TEEP on students' learning achievements. Our estimates show an increase of 12 to 15 score point in national achievement tests (NAT) with a 6-year exposure to TEEP. Interestingly, the impact estimate is much smaller if we do not use matching methods. Component-wise impact analysis showed that (i) new school building and renovations, (ii) instructional training to teachers, and (ii) textbooks significantly contributed to an increase in NAT. In particular, additional textbooks distributed to grade-4 students helped students improve their performance up to grade 6.

In the long-term impact analysis, we conducted a tracking survey to capture data on the school and labor market performance of TEEP cohort (students enrolled in Grade 6 in SY2004/05 or SY2005/06) and non-TEEP cohort (students enrolled in Grade 6 in SY1999/2000) in both treatment and control schools. This data structure allows us to use double differences (DD) to identify the gender-specific impacts of TEEP. We looked at students' school performance (years of schooling completed, number of repetitions in high school, and college entry), migration, and return in labor market.

As discussed, the placement of TEEP is not random as TEEP tends to target poorer provinces. To deal with this problem, we estimate a selection function and weigh observations with the estimated propensity scores of schools. To further check robustness of our empirical results, we also run the regressions using trimmed sample which trimmed off the observations with extreme propensity scores. We also explicitly control potential factors that induce trends differentiated between TEEP and non-TEEP divisions.

We found that improved school quality enhanced female advantage (or decreased female disadvantage) in subsequent schooling investments, migration and labor market earnings. That is, females study more (relative to males), and tend to migrate and earn more if they receive high-quality educational investments at the early stage. Based on the estimated impacts on schooling and wages, we calculated an internal rate of return. Our computation shows a high rate of return to TEEP.

Our findings from the long-term impact analysis are also consistent with the observation that returns to schooling are significantly higher among females than males, especially after high school completion. Labor market imbalance, represented by significant gender gap in returns to schooling, seems to create a unique situation where the impact of improved school quality is biased in favor of females. This result is also supported by the behavior of parents, who were found to prioritize schooling investments in their daughters in the face of financial constraints. Our example demonstrates the importance of linking education and labor markets when predicting the impact of any intervention in schooling.

(c) Conclusion and recommendations

Our study clearly showed that early stage investments in education have relatively large payoffs at later stages. School system is probably the most important institutional innovation in modern history as it has a potential to sustainably transform many people over time. Our study recommends that JICA should invest in early stages even within the elementary education cycle. For example, good education at early grades (e.g., grades 1 and 2) critically determines achievements at later grades. Moreover, if such a cumulative effect exists in the process of human capital formation, it is reasonable to conjecture that early childhood and preschool investments affect schooling outcomes too.

Our results also showed that a large intervention at the elementary school stage (TEEP) improved schooling outcomes among females relative to males, and increased their migration and labor market outcomes among females. This finding proves positive long-term impacts (though they are biased in favor of females), resulting in a relatively large internal rate of return. Therefore, investing in the early stage of education system has longer term payoffs beyond elementary schools.

Public investments in elementary education likely have positive dynamics impacts on subsequent schooling and labor market outcomes. If so, social returns to an early-stage investment can be greater than what the current study seems to show. This argument justifies large public investments to improve school quality at the early stage of public education, because the cumulative benefits are gradually realized at later stages in the education system and labor markets.

On evaluation design, we strongly felt that it would have been best to build evaluation component into TEEP from its onset. By doing so, we could have well prepared necessary data collection including baseline and follow up surveys. In our study, we retrospectively collected data on pre-TEEP stages from households and students. However, it is best to collect data before the intervention. One benefit of institutionalizing evaluation for a large project such as TEEP is that we are able to trace changes concurrently occurring with TEEP intervention. However, as our objective was to understand long-term impacts, it was necessary for us to face tracking challenges in any case to capture their schooling and work histories.

We also observed that data management was not ideal at various levels. In many schools, we encountered the situation that principals and teachers did not systematically keep records. In many cases, principals bring out important data when they are transferred to other schools (therefore, no data remains). The situation is better at central schools (that is, district offices), but they can make improvements to systematically compile data including student lists. In TEEP division offices, we attempted to collect detailed data on implementation timings and quantities of TEEP investments since one available data source (DEDP) originally (as part of the division-level TEEP completion reports) is not well formatted and organized. Not surprisingly, the quality significantly varied across divisions. At large, Dep Ed and donor agencies are recommended to strictly monitor the quality of data especially on implementations since this monitoring effect is expected to seriously improve implementations and the quality of evaluation works later on.

Main Report

1. Introduction

This report summarizes methodologies and findings of the TEEP evaluation study. The project had two large components: short-term impact and long-term impact studies. In the short-term impact study, we use school-level database including national achievement test data to assess the TEEP impact on students' learning achievements. In the long-term impact study, we quantify the impact on students' subsequent schooling outcomes, migration behavior, and labor market earnings by conducting a unique tracking survey. For the latter, we chose 8 provinces from northern Luzon and Visayas regions: Ifugao, Neuva Vizcaya, Antique, Iloilo, Negros Oriental, Cebu, Leyte and Western Samar and tracked former students to capture their schooling and work histories.

As described in detail below, we find significant impacts of TEEP on students' test scores (as a measure of learning achievements). TEEP increased national achievement test score by 12 to 15 points if students were exposed for 6 years. The long-term impact analysis demonstrates significant impacts biased in favor of females. Years of schooling completed significantly increased among females relative to makes. TEEP significantly increased females' migration and labor market earnings. Though the positive impacts were directed toward females, our study also shows a high rate of return in TEEP since it covered a larger population who are expected to work in labor markets for years.

2. Program Background

The Third Elementary Education Project (TEEP) was implemented from 2000 to 2006 by the Philippine Department of Education in all public primary and elementary schools[3] in the 23 provinces[4] identified as the most socially depressed in the Social Reform Agenda.[5] The total project cost was US$221.16 ($91.07 million from JICA and $82.84 million from World Bank, $47.25 from the Philippine government). The unique feature of TEEP is a combination of investments in school facility and education materials and school governance reform. Not only were school facilities and textbook supply improved, but the decision making process was also decentralized

[3] Primary schools cover grades 1 to 4, while elementary schools cover grades 1 to 6.

[4] The program covered both primary (grades 1–4) and elementary (grades 1–6) schools. This paper analyzes the impacts on only elementary schools. However, converting primary schools to elementary schools by extending enrollment up to grade 6 was also an important part of the TEEP program. Students who complete primary schools are likely to attend elementary schools in grades 5 and 6, which changes the student body of those schools between grades 1–4 and grades 5 and 6.

[5] The Ramos administration, along with their medium-term development plan, called Philippines 2000, identified reforms as the key to bridging social gaps and alleviating poverty. The objective of enhancing development through social reforms led to the formulation of the blueprint for social development in the Philippines, the Social Reform Agenda (SRA), marked as the first instance of social reforms in the history of the Philippines (Ramos 1995). As a result of the initial success of the SRA, the Congress of the Philippines in 1998 passed Republic Act 8425, widely known as the Social Reform and Poverty Alleviation Act (Republic of the Philippines, Congress, 1998). The law institutionalized the poverty alleviation program and a host of grassroots development strategies.

to the school and community levels. TEEP introduced a package of investments to schools in the selected 23 provinces. Specifically, the package of investments included (1) school building construction and renovation, (2) textbooks, (3) teacher training, (4) school-based management, and (5) other facility and equipment support.

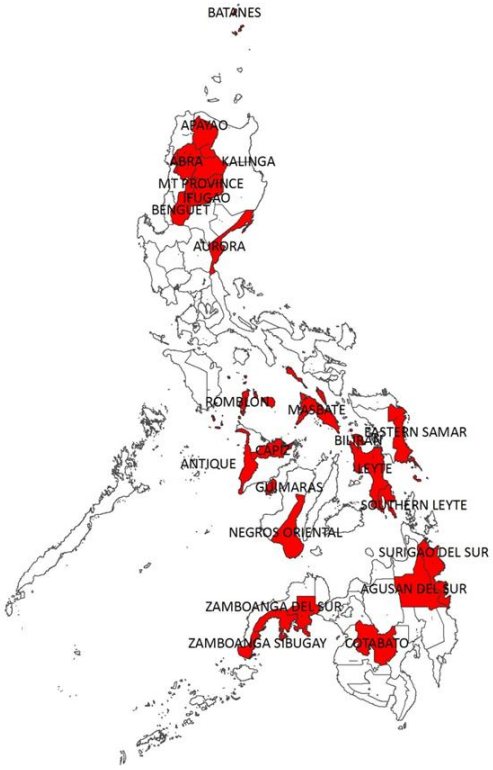
One unique feature of the program is a combination of hard and soft components. Through school-based management, schools are given an incentive to manage proactively and more independently of the government. Schools were partnered with communities and parents to decide key issues, such as improvement plans and school finance. Teachers were also trained systematically to improve teaching skills. Information management is being improved so that schools are responsible for systematically organizing information on enrollment, learning achievements, finance, and so forth, and reporting it to the division office. Schools are required to set improvement plans every year and compare them with actual achievement. This dynamic process is monitored by the division-level education department. School finance is also being decentralized to some extent to relax the school budget constraints because Philippine public schools are not allowed to charge school fees. TEEP schools are free to raise their own funds from communities, parents, and others, although resources are admittedly limited in many poor communities. These reforms in public schools are expected to improve education quality, which would then in turn increase returns to schooling in labor markets (see Yamauchi [2005] on returns to schooling).

The selection of TEEP provinces was purposive because it intended to cover the most depressed provinces identified in the Social Reform Agenda. TEEP allocation is rather different in the Philippines' three macro-regions. As shown in Figure 2.1, in the northern macro-region of Luzon, TEEP was concentrated in the Cordillera Administrative Region, a mountainous region in the center of northern Luzon. In the central macro-region of Visayas, TEEP divisions were relatively evenly distributed. In the southern Mindanao macro-region, TEEP divisions were clustered, although not as clustered as in northern Luzon.

TEEP was initially designed to follow a phase-in plan with three batches at the province level. However, the plan was altered in practice due to variations in preparedness across divisions. Because understanding the implementation process of TEEP is important in choosing the appropriate strategy to identify the TEEP impacts, we collected school-level data on program implementation time and investment amounts of different components. The data confirm that actual implementation did not follow the batch plan and suggest that the first and second batches were implemented almost simultaneously.[6]

[6] Khattri, Ling, and Jha (2010) used the lag between the first and second batches to identify the effect of school-based management on student test scores. Their analysis also includes TEEP investments such as new constructions as exogenous controlling variables. Their identification strategy is questionable, given that, in reality, the initial phase plan was changed due to variations in preparedness across divisions.

Figure 2.1—Map of TEEP and non-TEEP divisions in the Philippines (TEEP areas are in black)



3. Evaluation Design

3.1 Structure

This evaluation study has two components: short-term impact analysis using school panel data and long-term impact analysis using survey data. For the short-term analysis, we use Basic Education Information System (BEIS), National Achievement Test (NAT) score and TEEP implementation data. At this stage, the data on TEEP implementation has not been available except school construction/renovation information. Two consultants visited 23 TEEP divisions to collect detailed information on implementations of different TEEP components.

For the long-term impact analysis, we have conducted surveys simultaneously in 8 provinces (Figure 3.1): Ifugao, Neuva Vizcaya, Antique, Iloilo, Negros Oriental, Cebu, Leyte and Western Samar. Ifugao, Antique, Negros Oriental and Leyte are TEEP divisions, and the others are non-TEEP divisions. Basic idea of identification is to compare TEEP and non-TEEP divisions in the same area. For example, Negros oriental (treatment) and southern part of Cebu (comparison), both adjacent to each other, are compared for identifying TEEP impacts. Municipalities are selected from each division so that our sample areas in both treatment and comparison share similar socio-economic situations.

Figure 3.1—Sample provinces

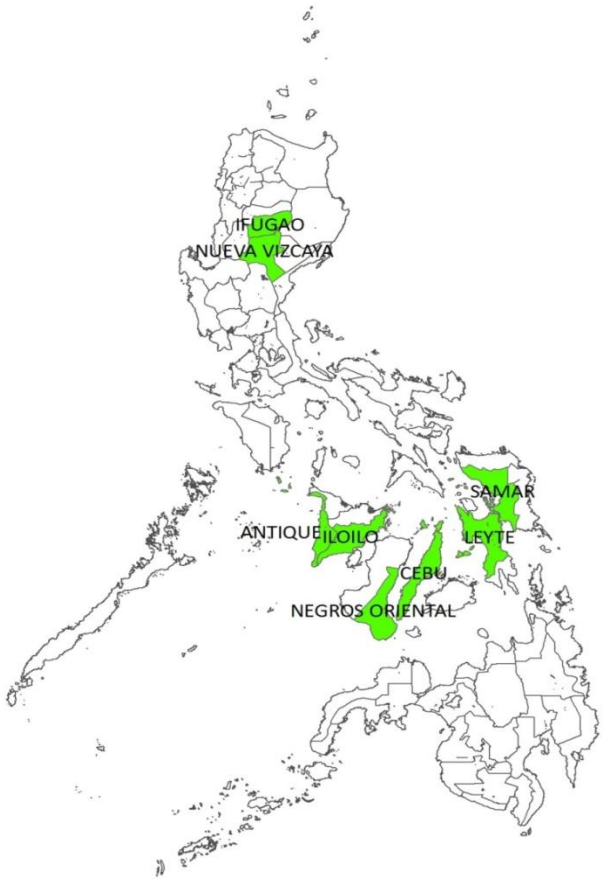


Table 3.1 Sample – Student composition

Division type Cohort	TEEP	non-TEEP
Pre-TEEP	Comparison	Comparison
TEEP	Treatment	Comparison

We also sample students from pre-TEEP and TEEP cohorts. In this way, we have four combinations of cohorts and division types: (i) TEEP division and TEEP cohort, (ii) TEEP division and pre-TEEP cohort, (iii) non-TEEP division and TEEP cohort, and (iv) non-TEEP division and pre-TEEP cohort. These combinations, controlling school-specific fixed unobservables (fixed effects), provide a basis of identifying TEEP impacts in our study (see Table 3.1).

In our study, we observed that the original batch plan was not accurately implemented. Largely due to delays in the initial stage, batches 1 and 2 were almost simultaneously introduced.

Preparedness in each division critically determined this process. This resulted in difficulty to use phase-in method to evaluate TEEP impacts.

It is also important to mention that our study could not assess SBM because SBM and other components were simultaneously introduced. Though SBM was introduced to non-TEEP divisions after 2006, we could not use TEEP as a comparison group for the post-2006 SBM treatment since TEEP divisions had already SBM (and many other TEEP investments). Therefore, we do not assess SBM in this current study. As a result, the estimates of component-wise TEEP impacts likely contain SBM contributions if SBM augments the impacts of those components.

On generalizability or representativeness of our findings largely based on Visayas in both short-term and long-term impact assessment studies, there are two key factors that we considered.

First, we found that TEEP was concentrated in CAR in northern Luzon and also clustered in 3 regions in Mindanao, but seemingly more randomly allocated in Visayas. Our preliminary work showed difference in income class distribution between TEEP and non-TEEP is small in Visayas. Therefore, comparison between TEEP and non-TEEP makes much sense there. In the long-term impact analysis, we included an experimental area from northern Luzon basically to validate the sample (we did not find any differences in behavior in that region, so it makes sense to pool all the divisions). Our judgment on TEEP/non-TEEP comparability is a critical factor.

Second, in both works, we use matching to increase comparability (now in a statistical sense). Short-term impact analysis – we used municipality-level income class. In long-term, we used school data as well as household-level asset (durables plus house and residential and commercial-purpose land) to construct school weights. For the above two reasons, we conclude that the results are highly generalizable. In other words, if we include Luzon and Mindanao, our estimates will be largely biased; this creates additional burdens to reduce bias.

3.2 Implementation of tracking survey

As summarized in the project proposal, we have worked with two teams: Fe Gascon's and the Office of Population Studies (OPS), University of San Carlos. Fe Gascon's team mainly consisted of members who have rich experiences in conducting rural household surveys (mainly based in the International Rice Research Institute, Los Banos). OPS team had expertise in conducting socio-economic surveys in Visayas region. One great achievement of the OPS team is an internationally famous Cebu Longitudinal Study. Fe Gascon's team conducts surveys in Northern Luzon and Western Visayas, whereas OPS conducts surveys in Central and Eastern Visayas.

3.2.1 Survey preparation

From March 2010, our teams discussed survey methodologies including sampling and instruments. Instruments were finalized in the beginning of June after several preliminary versions. The instruments have household and student questionnaires as a core part, supplemented by school,

barangay captain and Parents Teacher Community Association (PTCA) head questionnaires. School questionnaire was distributed to sample schools through the Department of Education (Dep Ed).

Pretests were implemented in different locations, by the two teams. Observations were summarized and used for revising the instruments. Combination of different expertise in this project truly contributed to improving the instruments, and formatting until the final stage.

3.2.2 Sampling

One of the largest challenges in this project was sampling. We have two stages of sampling: schools and former students. As explained earlier, we have chosen school districts (mostly the same as municipalities) from certain areas in treatment and comparison that share similar socio-economic conditions (for example, census 2000 municipality income classification, local language, etc.). We screened complete mono-grade schools with enrollment size greater than 120 from the 2002 master list. Random sampling was done to have 15 schools from TEEP division and 10 schools from non-TEEP division respectively.

Late April 2010 we requested Dep Ed Bureau of Elementary Education (BEE) to send letters to our sample divisions to have lists of students who were enrolled at grade 6 in 1999/2000, 2004/2005 and 2005/2006 school years. In case, some of our team members (including myself) directly visited division offices to follow up. Neuva Vizcaya, Negros Oriental, and Samar were the first group to submit the requested student lists. For Ifugao and Leyte, we needed an extended period of time to have the requested lists (Leyte until mid July 2010). National election in May also delayed this process. In total, we spent about 3 months to have student lists from all the schools.

In case that we found school districts are unsafe for our study (due to various reasons), and/or schools were found incomplete (up to grade 4) in 1999/2000, etc., we replaced them with some other schools.

Three provinces – Leyte, Ifugao and Antique – deserve special explanation.

Leyte – The division did not submit the requested student lists until July. Judith Borja and Lorna Perez visited the division office on July 13 to follow up our request. We also offered an alternative list of schools sampled from the districts who already submitted. Such a delay experienced in Layte affected our survey schedules in both the division and Samar.

Ifugao – There were two schools located in Tinoc district from which we could not receive student lists until late June. One of them was found incomplete (up to grade 4) in 1999/2000. My decision at that time was to drop them from our sample without replacement. When I visited the region around Aug 10, I changed the decision by bringing the district back to our sample since (i) the Ifugao supervisor informed that some enumerators (originally contacted) are ready to conduct survey in Tinoc, and (ii) I realized an urgent need to increase sample size from Ifugao for the

reason described in Section 4.4. In Tinoc, I have chosen Tinoc central and Wangwang schools after consulting with one school principal residing in Kiangan in Ifugao (at the municipality hall with their staff who coordinates our survey in the municipality/district). It took more than 4 hours one way to reach Tinoc from Kiangan all in the mountains. I, Marites, Cze (Ifugao supervisor) and one enumerator visited Tinoc directly to have the lists from Tinoc Central school. The Wangwang school head also brought the lists to us in Kiangan. Therefore we have 15 schools in Ifugao.

Antique – First, two schools in Valderrama - Lublub and Borocboroc were destroyed by a typhoon-related flash flood in 2002-2005. The impact is not small in our sample. As of Aug 30 when I visited the site, it was impossible to stop the operation. The other schools - Bunsod and Manlacho are fine in the sense that they did not have direct impacts of the flood. My decision on this issue is that we keep the sample of Lublub and Borocboroc, but need to add two more new schools in the sample. Second, initial oversampling from Sibalom was a serious issue too. As done already, the inclusion of Hamtic North and South districts (under Hamtic municipality) could be a solution. Sibalom is an income-class 3 (middle income) municipality, so Sibalom is not a good substitute for San Remegio which was dropped due to safety reasons. In any case, since oversampling from Sibalom unbalances our sample (and comparability to Iloilo), we decided to include 6 schools from Hamtic. This decision was made on September 2, and our team has received student lists from the district offices for sample. Therefore, we believe that the Antique sample is finally well balanced, being comparable to the Iloilo samnple.

The finalized list of schools follows.

Table 3.2 Sample schools

Ifugao (TEEP)

Name of Elementary School	District
Poitan ES	Banaue
Namulditan	Hingyon
Cababuyan	Hingyon
Hingyong ES	Hingyon
Ambasa ES	Lamut
Ilap ES	Lamut
Lawig ES	Lamut
Pindongan ES	Kiangan
Nagadacan ES	Kiangan
Mungayan ES	Kiangan
Kiangan CS	Kiangan
Duit ES	Kiangan
Bokiawan ES	Hungduan
Tinoc CS *	Tinoc
Wangwang ES *	Tinoc

*Two schools in Tinoc were added in August (we initially dropped two other schools from Tinoc)

Neuva Vizcaya (NON-TEEP)

Name of Elementary School	District
Barat ES	Bambang I
Bambang North CS	Bambang II
Sto. Domingo ES	Bambang I
Aliaga ES	Bambang I
Binalian ES	Kayapa
Quezon CS	Quezon
Solano North ES	Solano II
Solano East CS *	Solano I
Solano South ES *	Solano II

*Oversampled from two schools, as one inaccessible school was dropped from the sample

Antique (TEEP)*

Name of Elementary School	District
Manlacho ES	Valderrama
Borocboroc ES **	Valderrama
Lublub ES **	Valderrama
Bunsod ES	Valderrama
Vilafont ES	Sibalom North
Juan Vego ES	Sibalom North
Esperanza ES	Sibalom North
Lacaron ES	Sibalom North
Initan ES	Sibalom South
Nagdayao ES	Sibalom South
Catmon ES	Sibalom South
Hamtic CS***	Hamtic North
Buhay ES***	Hamtic North
Masanay ES***	Hamtic North
Linaban***	Hamtic South
Fabrica***	Hamtic South
Lanag***	Hamtic South

* Total sample is 17 schools, ** two schools were destroyed by floods, *** added

Iloilo (NON-TEEP)

Name of Elementary School	District
Tacuyong Norte ES	Leon I
Gines ES	Alimodian
Magsaysay ES	Maasin
Amurogtong ES	Igbaras
Cabacanan ES	Alimodian
Sibucauab ES	Tubungan
Layong ES	Maasin
Isian Norte ES	Leon I
Calampitan ES	Igbaras
Igpigus ES	Igbaras

Negros Oriental Province (TEEP)

Name of Elementary School	District
Malangsa	Vallehermoso
Molobolo	Vallehermoso
Maglahos	Vallehermoso
Guihob	La Libertad
Tamao	Jimalalud
Gomentoc	Ayungon
Mabato	Ayungon
Sampiniton	Bantolinao, Manjuyod
Kayotesan	Kauswagan, Manjuyod
Butong	Manjuyod
Candabong	Manjuyod
Tandayag	Amlan
Libertad Ong Calderon	Ajong, Sibulan
San Antonio	Sibulan
Balugo	Sibulan

Cebu Province (NON-TEEP)

Name of Elementary School	District
Balao	Barili I
Lawaan	Dumanjug I
Pawa	Dumanjug II
Madanglog	Vive, Ronda
Polo	Alcantara
Manduyong	Badian
Talayong	Badian
Guiwanon	Looc, Ginatilan
Tangbo	Samboan
Pasil	Santander

Leyte (TEEP)

Name of Elementary School	District
Bachao	Leyte
Canomantag	Barugo I
M. Morales	Libo, Carigara I
Tinaguban	Carigara II
Astorga	Alangalang II
Lukay	Alangalang II
Sta. Cruz	Jaro I
Luntad	San Miguel, Palo I
Guindapunan	Palo II
Campetik	Palo II
Calsadahay	Tanauan I
Maliwaliw	Dagami South
Olmedo	Pongon, MacArthur
Javier CS	Zone I, Javier
Cassidy ES	Ponong, CARIGARA I

Western Samar (NON-TEEP)

Name of Elementary School	District
Jia-An	Jiabong
Canduyucan	San Sebastian
Villareal I Central School	Poblacion, Villareal I
San Roque	Villareal II (Talalora)
Sta. Rita CS	Santan, Sta. Rita I
Salvacion	Basey I
Mabini	Basey I
Basey II CS	Palaypay, Basey II
Kaluwayan	Marabut
Osmena	Marabut

The next step was to sample students. We randomly sampled 15 students from the 1999/2000 grade-6 group, and 20 students from the SY2004/2005 and SY2005/2006 grade-6 group. The former represents pre-TEEP cohort, while the latter, post-TEEP cohort. We also prepared a replacement list for each group (10 students for the 1999/2000 grade-6 group, 15 students for the SY2004/2005, SY2005/2006 grade-6 group). In this process, we did not stratify by gender. Also note that whether they graduated from elementary schools was not known at the time of sampling, and some students did not complete elementary school education.

3.2.3 Survey Instruments

We spent almost 4 months starting in March 2010 discussing and finalizing survey instruments. The instruments have five parts: student tracking questionnaire, household questionnaire, barangay captain questionnaire, PTCA head questionnaire, and school questionnaire.

3.2.4 Survey

After trainings implemented in each division, surveys were launched first in Neuva Vizcaya in late June, followed by Ifugao, Negros Oriental and Cebu, and Iloilo in July. Later Layte and Samar started in mid August, and finally Antique joined in very late August.

I was in Negros Oriental and Cebu in early July, and in Iloilo in mid July to monitor the surveys. Early August I briefly visited Iloilo , and Neuva Vizcaya and Ifugao until mid August. Later I was in Leyte, Samar and Antique. My schedule basically followed survey schedules/progress in each region, except Neuva Vizcaya and Ifugao where I was able to monitor the final stage only (this has some consequence, discussed below)

Survey teams report to Fe Gascon and OPS regularly, which was shared with me too. Especially when teams encounter problems with sample and replacement, I had to solve them immediately.

Except a few cases, the above methodology worked well in the field. Exceptions are: (i) enrollment was small in 1999 (so we exhaust replacement list too), (ii) schools reported the inflated number of students for some reasons (including unreal student names), etc.

As a rule, we replaced one of siblings from the same household if siblings were sampled (keep one student from them with priority on the 1999/2000 group). When we cannot locate student's household or get consent for tracking including being unable to have precise contact information, we were required to replace sample. This process was very tedious since enumerators are visiting different sample respondents. To centralize the replacement process, the role of supervisor (Fe Gascon's team) or team leader (OPS) had to be reinforced in the field.

Due to our municipality selection method, many schools are located in mountainous areas. Not to mention Ifugao (the entire area is mountains), many schools in Iloilo, more than half of schools in Negros Oriental and Cebu, and about half in Antique are in mountain areas. With heavy rain in the rainy season, survey teams had enormous difficulty in accessing schools/barangays and respondents (especially when we need to walk, use motorbike or even horse).

The overall quality control has been successful except some issues specific to Ifugao and Antique. In Ifugao, one enumerator was assigned to one school in a rather isolated location where close monitoring and supervision were challenging. Incidentally, based on the fielded questionnaires, migration rate in both our sample students and their siblings (and household members in 2000-2010) was very low. This issue needs audit visits (planned late September or early October) to verify the migration pattern by rechecking the questionnaires with the respondents, since this tendency – whether true or not – affects our analysis and empirical results.

Antique has faced a situation where teams needed to hire local enumerators recommended from sample municipalities, which showed low quality. Thus more experienced enumerators from Iloilo had to be attached to them to ensure the quality of interviews. Together with school sampling problems, we may need to pay special attention to Antique from this stage, and if necessary, provide enough assistance too.

Survey operations in OPS were relatively stable and robust. The principal investigator, Judith Borja, who took over Nanette Lee, communicates with me very closely to share the field situation.

In OPS team, household and within-division tracking surveys were completed by the middle of September. They immediately started Cebu and Manila tracking surveys in weekends when our sample students can spare their time to cooperate with us. I also attended part of their survey in Manila. The entire survey was completed late October.

In Fe Gascon Team, due to heavy rain, the survey operation tended to be delayed in Western Visayas. The re-sampling process in Antique and remoteness of Valderrama also contributed to the delay. However, the entire survey in Iloilo and Antique was completed by the middle of

October. The regional coordinator who was in charge of this region moved to Iloilo to revisit communities to recollect the data to ensure the quality.

In early October, I organized a special team in Ifugao, consisting of two former enumerators from the province, Fe Gascon and Maritess Tiongco, to verify the information recorded especially in the household roster, students locations, and the sibling section. We have visited three schools in Neuva Vizcaya and 7 schools in Ifugao. We found data manipulation in Ifugao, which were systematic, not restricted only in the 7 schools. After this observation, we decided to spend almost one month to correct the information in Ifugao, by forming a new survey team to revisit all the households in our sample. The correct was started by the middle of November. In Neuva Vizcaya, the team has requested barangay secretaries to collect the contact information of all our sample households and students to verify the information captured in the initial survey. Though this type of problems and our action are unusual, our effort to correct the once-captured data was meaningful especially since the problem critically affected migration rate of our students. The verification work was agreed to be done with non-cost extension. We believed that the corrected migration rate would be comparable with or slightly lower than those from our Visayas sample (this was proven as you see in Section 3.2.8).

Tracking activities by the OPS team were relatively fast mainly because they did a good job at the initial visits to our sample households (e.g., contacting our students from their parents' households to make appointments) and tracking started immediately after the household survey. In contrast, Los Banos team had to spend an extended period of time late 2010 to May 2011. There were several factors to explain this strategic gap. First, they faced a larger number of migrant students (including transition students) from northern Luzon and Western Visayas. Especially in northern Luzon, students were found to be quite mobile since their communities were connected to Manila by land (via highway bus). To a lesser extent, this is true for Western Visayas region. Therefore, migration rate was higher. Second, this team attempted to interview students face to face. Since they are based in Laguna, it is easier to approach our students living in Manila and other parts of Luzon. This process required a longer time. Third, verification work in Ifugao and Neuva Vizcaya (which continued until April 2011) created a substantial delay in starting tracking activities from that area.

3.2.5 Results of verification work

The team headed by Fe Gascon attempted to check all the Ifugao and Neuva Vizcaya questionnaires (households) in a month, but it was not practical since mothers/guardians are temporarily unavailable to interview, and the plan was too ambitious under the team's human resource constraint. Therefore, only 60% of the targeted households were checked and if necessary, corrected in November 2010. The rest was carried to 2011: some periods in January and April. All the verification of the work has been done by April, 2011.

3.2.6 Data editing and cleaning

At OPS, data editing started right after the household and tracking surveys. We have also coded open ending questions/answers at this stage. Therefore coding was first finalized at OPS. Data editing has started at Los Banos in February 2011. Finalized OPS codes were transferred to the Los Banos team, who utilize them in their editing work. In the period of March to June, we have added new answers of open ending questions to redefine the codes. In this process, we have done some optimization works to reclassify answers (therefore, codes) of some questions.

Through the office editing and additional verification phone interviews from Los Banos, we also decided to discard 51 unqualified households from northern Luzon due to unverifiable information recorded in those questionnaires. This was one of the expected outcomes based on my observations in August 2010; this justified my decision to include two schools in Tinoc district.

Cleaning was done at two locations: OPS and Los Banos. OPS has done cleaning almost perfectly (except several cases that I identified to recheck), but Los Banos team faced substantial challenges at this stage since they only used excel. I worked very closely with this team on site as well as from Washington DC. Using Stata, we identified problems patiently, and Los Banos team (especially data manager) checked each problem manually, verifying information recorded in the questionnaires. Starting from duplication problems, we had to check all the data. Though this process was so time consuming, it was quite efficient mainly because our questionnaires were rather simple and the data manager had good expertise in reading data, which helped her understand my inquires.

OPS has a clear ethics policy that rules their data disseminations, including some confidential information from our survey. These included names and addresses. Moreover, we had some misunderstanding on how to handle string information, such as job description (note these information is coded into occupations and industries). For the above reason, we decided to transport all the OPS questionnaires to Los Banos to reenter these marked information.

Data editing and cleaning stage significantly contributed to improving the quality of our data.

3.2.7 Final sample composition

Table 3.3 shows the final sample composition of tracked students. There were 36 cases untracked out of the total of 3487 students. Therefore, the tracking rate is about 98.97%. The final sample size is 3451. As discussed above, Ifugao sample went through a very detailed verification process in the field and office, which resulted in 470 students in the sample.

Table 3.3 Sample composition

Grade-6 school year	Antique	Cebu	Ifugao	Iloilo	Leyte	Negros Oriental	Neuva Vizcaya	Western Samar
1999-2000	244	143	188	143	212	227	139	142
2004-2005	159	107	137	97	158	161	92	96
2005-2006	177	95	145	111	135	145	101	97
Total 3451	580	341	470	351	505	533	332	335

Table 3.4 Mode of interview

Mode	Freq.	Percent	Cum.
Face to face	3,101	89.86	89.86
Phone	350	10.14	100.00
Total	3,451	100.00	

As Table 3.4 shows, in most cases, we used face-to-face interviews (3101 cases; 89.86% of 3451). This is because (i) we tried to interview students face to face in the 3 metropolitan areas and most of the students who came to Luzon (including NCR) from northern Luzon and Western Visayas, and (ii) similarly students in the origin divisions (including those who are in their communities). Teams had a consensus that if possible, it is best to interview face to face to ensure the quality.

3.2.8 Migration behavior and locations of our students

It was found that our population is quite mobile (Table 3.5). 22.11% of our students were not household members at the time of survey. 8.55% were physically away from their communities but still were members of their households (we call them transition students). If we combine these two groups, about 31% of our students were not in their original communities.

Table 3.5 Migration status (tracked students)

Type	Freq.	Percent	Cum.
In-town	2,393	69.34	69.34
Migrant	763	22.11	91.45
Transition	295	8.55	100.00
Total	3,451	100.00	

Table 3.6 disaggregates migration status by division. Migration rates vary across divisions. The proportion of in-town students is the highest in Neuva Vizcaya. This is reasonable since this division has a few relatively large towns that can absorb our students (either for work or study). Migration rate is the highest in Ifugao (if we combine permanent and transition). Note that the migration rate has dramatically changed through the verification work. Permanent migration rate is remarkably high in Antique, Leyte, Negros Oriental and Western Samar, which is highly

correlated with poverty incidence in those divisions. Transition migration rate is the highest in Ifugao. Interestingly, four out of five divisions identified above are TEEP divisions.

Table 3.6 Migration status by division

	Antique	Cebu	Ifugao	Iloilo	Leyte	Negros Oriental	Neuva Vizcaya	Western Samar	Total
In-town	381	236	267	283	344	364	296	222	2,393
	65.69	68.41	56.81	80.63	68.12	68.29	89.16	66.27	69.34
Migrant	167	43	86	57	155	126	33	96	763
	28.79	12.46	18.30	16.24	30.69	23.64	9.94	28.66	22.11
Transition	32	66	117	11	6	43	3	17	295
	5.52	19.13	24.89	3.13	1.19	8.07	0.90	5.07	8.55
Total	580	345	470	351	505	533	332	335	3,451

Next, we tabulate current provinces. Students tend to head to the National Capital Region, its surrounding provinces, Banguet (where Baguio is located) and Cebu (most likely Cebu City).

Table 3.7 Locations of our students (tracked sample)

Province	Freq.	Percent	Cum.
AGUSAN DEL NORTE	1	0.03	0.03
AKLAN	1	0.03	0.06
ALBAY	1	0.03	0.09
ANTIQUE	427	12.41	12.49
BASILAN	1	0.03	12.52
BATANGAS	11	0.32	12.84
BENGUET	64	1.86	14.70
BUKINDNON	1	0.03	14.73
BULACAN	16	0.46	15.19
CAGAYAN	4	0.12	15.31
CAMARINES SUR	1	0.03	15.34
CAPIZ	2	0.06	15.40
CAVITE	40	1.16	16.56
CEBU	389	11.30	27.86
City of Manila, first district	26	0.76	28.62
City of Manila, fourth district	85	2.47	31.09
City of Manila, second district	134	3.89	34.98
City of Manila, third district	38	1.10	36.08
DAVAO DEL SUR	1	0.03	36.11
EASTERN SAMAR	1	0.03	36.14
GUIMARAS	2	0.06	36.20
IFUGAO	350	10.17	46.37
ILOCOS NORTA	1	0.03	46.40
ILOILO	351	10.20	56.60
ISABELA	7	0.20	56.80
KALINGA	1	0.03	56.83
LA UNION	4	0.12	56.94
LAGUNA	25	0.73	57.67
LEYTE	383	11.13	68.80
MAGUINDANAO	1	0.03	68.83
MISAMIS ORIENTAL	2	0.06	68.88
MOUNTAIN PROVINCE	1	0.03	68.91
NEGROS OCCIDENTAL*	25	0.73	69.64
NEGROS ORIENTAL	398	11.56	81.20
NEUVA ECIJA	5	0.15	81.35
NEUVA VIZCAYA	330	9.59	90.94
NORTHERN SAMAR	3	0.09	91.02
PALAWAN	4	0.12	91.14
PAMPANGA	15	0.44	91.57
PANGASINAN	7	0.20	91.78
QUEZON	3	0.09	91.87
QUIRINO	3	0.09	91.95

RIZAL	27	0.78	92.74
ROMBLON	1	0.03	92.77
SIQUIRE	1	0.03	92.79
SOUTH COTABATO	1	0.03	92.82
SOUTHERN LEYTE	1	0.03	92.85
SULTAN KUDARAT	1	0.03	92.88
SURIGAO DEL NORTE	2	0.06	92.94
SURIGAO DEL SUR	1	0.03	92.97
WESTERN SAMAR	238	6.91	99.88
ZAMBALES	2	0.06	99.94
ZAMBOANGA DEL SUR	2	0.06	100.00
Total	3,442	100.00	

Notes: This does not include untracked students, unidentified locations and those who reside abroad, which reduces the sample to 3442 in this table. Some student households of one school in Negros Oriental reside in Negros Occidental crossing the provincial border.

3.2.9 Data quality assessment

The best way to assess the quality of data is to use it for serious analysis. We have conducted several econometric/statistical analyses using our data. Through this process, we discovered and corrected additional errors in the data through close communications with both Cebu and Los Banos teams and if necessary checking the original questionnaires and calling our respondents for further clarification . Empirical findings (after checking their robustness) also confirm that the quality of data is quite high (clean data).

4. Short-term impact analysis

4.1 School-level data

In close collaboration with Dep Ed, especially Research and Statistics Division, we received BEIS (SY2002/03 to SY 2008/09) and NAT (SY002/03 to SY 2008/09) data. In NAT data (school average), we have grade-4 test scores in SY 2002/03, grade-5 in SY 2003/04, and grade-6 in SY 2004/05. Grade-4 in SY 2002/03 to Grade6 in SY 2004/05 is panel data tracking the same cohort in each school.

We needed an intensive programming to transform BEIS data for analysis. The data was originally in Excel. To reorganize school-level data in different divisions/regions for one school year, we needed to spend about 10 hours to run a program. One year data has about 20 different sheets (each one of which contains huge data). In October to November 2010, we completed this conversion for SY 2002/03 to SY 2006/07.

As described above, we supplemented DEDP data by collecting additional information on implementation timings from each TEEP division. School building data was taken directly from

the Dep Ed central office database. In the analysis, we use data on textbook (by grade), training, and school building.

For the short-term impact analysis, we decided to take following strategies. First, since TEEP was introduced to the most depressed provinces, the allocation of TEEP was purposeful. For example, TEEP was concentrated in CAR, a mountainous region in the center of northern Luzon. In Visayas, TEPP divisions was relatively scattered over space. In Mindanao, TEEP divisions were clustered, though it is not as clustered as those in northern Luzon. Under this circumstance, we decided to use Visayas (regions 6, 7 and 8) to use relatively comparable TEEP and non-TEEP divisions in the same regions. It is important to choose relatively homogeneous areas to analyze TEEP impacts.[7]

Second, we use the cohort panel from grade-4 (SY 2002/03) and grade-6 (SY 2004/05) to purge cohort-specific fixed unobservables. In this type of analysis, we have to consider two sorts of unobserved fixed components: school-level and cohort-level. Since the unit of observations in the short-term impact analysis is school, we can always wipe out school-level fixed unobservables, but NAT data structure enables us to difference out cohort-specific effects too. If we pursue the long-term impact analysis using school data (not individual data from our tracking and household surveys), we can use grade-6 NAT score data in different years, handling only school-level unobserved fixed components.

The analysis required a little tedious data management, not to mention BEIS. First, NAT SY 2002/03 (garde-4) data lacked school IDs. The data recorded different IDs, which is understandable since the Dep Ed introduced systematic new school IDs along with BEIS as part of TEEP. BEIS started in SY 2002/03, but NAT SY 2003/03 data was not updated with the new school IDs. This created a difficulty in merging NAT SY 2002/03 and NAT SY 2004/05 (using the new school IDs). To prepare school IDs in NAT, we used BEIS SY 2002/03. However, such merging was not perfect as we can use school names in each division (note that NAT data does not have district information). The merging rate increased if we restrict our sample to elementary schools, by dropping primary schools where grades 1 to 4 students were taught. This restriction makes sense since our analysis used grades 4 to 6 in the cohort panel analysis. Therefore, our analysis pertains to elementary schools in SY 2002/03 that offered grades 1 to 6. (Note that TEEP also contributed to the conversion of primary schools to elementary schools by building new classrooms for grades 5 and 6. However, this effect is not included in our analysis. Moreover, it is

[7] First, we found that TEEP was concentrated in CAR in northern Luzon and also clustered in 3 regions in Mindanao, but seemingly more randomly allocated in Visayas. Our preliminary work showed difference in income class distribution between TEEP and non-TEEP is small in Visayas. Therefore, comparison between TEEP and non-TEEP makes much sense there. In the long-term impact analysis, we included an experimental area from northern Luzon to validate the sample (we did not find any differences in behavior in that region, so it makes sense to pool all the divisions). Our judgment on TEEP/non-TEEP comparability is a critical factor. Second, in both works, we use matching to increase comparability (now in a statistical sense). Short-term impact analysis – we used municipality-level income class. In long-term, we used school data as well as household-level asset (durables plus house and residential and commercial-purpose land) to construct school weights. For the above two reasons, our estimates are highly generalizable.

possible that students from primary schools, not part of our sample, came into grades 5 and 6 in our sample elementary schools, which changes the cohort student compositions at grade 5).

Second, we needed income data on municipalities (or school district) to condition TEEP. The data we used came from Census 2000. Census 2000 defined income category (ranks 1 highest to 5 lowest) to each municipality. For cities, we used rank 1 due to the income threshold used for municipalities. TEEP was implemented in divisions classified as the most socially depressed according to the presidential social reform agenda. However, there are income variations even within a TEEP division, so it is highly likely that we can find municipalities (school districts) that share similar socio-economic conditions in both TEEP and non-TEEP divisions. This setting is quite helpful in our analysis to match similar groups from TEEP and non-TEEP.

To assess the TEEP impact, it was essentially important to have accurate implementation and investment data. For TEEP implementation information, we have Division Education Development Plan (DEDP) data, which was part of the TEEP completion reports. However, we found that the DEDP data do not identify implementation timings of different components and the completeness and quality substantially vary across divisions, though most of the worksheets have useful information. The DEDP data has aggregated TEEP inputs over the period of SY 2000/01 to SY 2004/05.

To overcome this gap, we decided to hire two local consultants with a mission to visit 23 TEEP division offices and find the raw data on TEEP. They started visiting those divisions from late August 2010. Though it took a longer time than we expected, they have completed the data collection in the middle of November. Some divisions had not recorded TEEP inputs as accurately as others, but the data collection was successful.

We covered textbook, training, school based management funds, school building (SBP), school innovation and improvement fund (SIIF), equipment/furniture, and supplementary instructional materials (SIM).

Information on school constructions and renovations was received from the Dep Ed central office in July. For School Building Projects (SBP), we decided to complete this data by gathering information that was missing in some ongoing projects as of March 2005. We also used DEDP data/formats in textbook and teachers' training to identify the implementation timings (for training, we only identified when it started). For other components, we used school years to identify investment amounts.

Our analysis consists of two parts. First, we estimate the average treatment effect. Here we estimate the average impact of TEEP once TEEP is conditioned on the initial municipality incomes. Our basic assumption is that TEEP impacts should be similar between treatment and comparison groups that share similar socio-economic conditions. For this purpose, we use difference-in-difference with propensity score matching (DD-PSM) using kernel estimation. The second approach is component impact analysis. In this analysis, we use TEEP inputs in textbooks,

teachers' training, school building and school based management. Since we cannot use propensity to weight (as we have four treatment variables in the same equation), we use difference-in-difference (DD) controlling the initial conditions.

Table 4.1 shows the mean and standard deviation of mathematics and overall scores of the cohort in SY 2002/03 and SY 2004/05 for TEEP and non-TEEP areas in Visayas regions, separately. TEEP schools have higher average scores than non-TEEP schools in both years.

Table 4.1—Summary of NAT test scores for TEEP and non-TEEP, SY 2002/03 and SY 2004/05

	TEEP				Non-TEEP			
	SY 2002/03		SY 2004/05		SY 2002/03		SY 2004/05	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Overall score	46.975	14.674	63.712	13.431	44.447	13.515	59.795	12.875
Math score	48.390	17.961	66.035	16.624	45.823	16.753	62.208	16.698
Number of observations	1,774		1,774		2,434		2,434	

Source: National Achievement Test database, various years.

The BEIS data provide detailed information on student enrollment and achievements and teachers since SY 2002/03. The data normally disaggregate the information by grade, age, and gender.[8]

As discussed, we obtain income data on municipalities (or school district) from the 2000 Census. Local income level is an important factor that determines school and family environments. Controlling local income levels is crucial because competition between public and private schools matters in the selection of students in the Philippine context. In high-income municipalities (school districts), students from well-off families and with high test scores are likely to be accepted into private schools. Therefore, we expect differences in the ability distribution in public schools between high- and low-income municipalities. If school quality and student ability are complementary, the effect of TEEP on NAT change is expected to be different between high- and low-income districts.

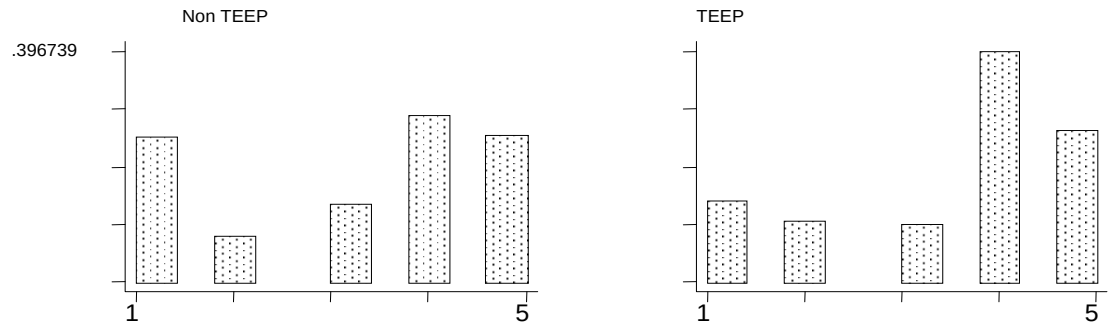
We assigned an income category to each school district based on the 2000 Census. The census defined income category (ranking from 1, highest, to 6, lowest) for each municipality.[9] Note that some municipalities are split into a few school districts. In cities, we ranked school districts as 1

[8] BEIS data needed intensive programming to transform for analysis. The data were originally in Microsoft Excel. The computer program needed about 10 hours to reorganize school-level data in different divisions and regions for one school year.

[9] The income classification of municipalities (municipality income) used in this paper is based on Republic of the Philippines, Department of Finance (2001), Department Order No. 32-01 (effective November 20, 2001) and Census 2000. The income categories for 1,435 municipalities are defined as follows: 1: Philippine peso (PHP) 35 million (M) or more (number of municipalities: 130); 2: PHP 27M or more but less than PHP 35M (140); 3: PHP 21M or more but less than PHP 27M (204); 4: PHP 13M or more but less than PHP 21M (543); 5: PHP 7M or more but less than PHP 13M (401); 6: less than PHP 7M (17).

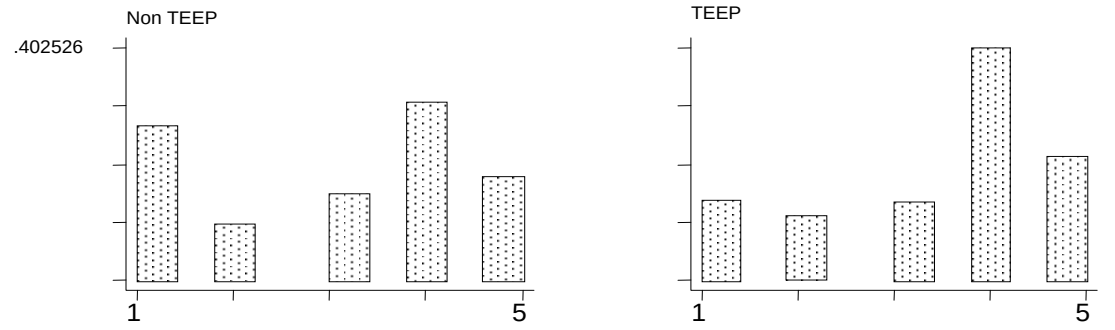
based on the income threshold used for municipalities. TEEP was implemented not randomly but in the divisions identified as socially most depressed in the presidential Social Reform Agenda.

Figure 4.1—Histogram of school districts, by income category for TEEP and non-TEEP groups



Source: census 2000 Municipality Income Classifications

Figure 4.2—Histogram of sampled schools, by income category for TEEP and non-TEEP groups



Source: Census 2000 Municipality Income Classifications.

Figure 4.1 shows the distribution of school districts by income category in TEEP and non-TEEP groups. School districts are concentrated in income categories 1, 4, and 5—that is, the highest income and the two lowest income rankings—for both TEEP and non-TEEP. Although we observe that more school districts are in income category 4 (and fewer in 1) in the TEEP group than in the non-TEEP group, the difference does not look significant. Further, Figure 4.2 shows the distribution of schools in the TEEP and non-TEEP groups. Our basic observation remains valid here. Therefore, it is likely that we can find (and compare) school districts that share similar socioeconomic conditions in both TEEP and non-TEEP divisions.

For TEEP implementation information, we have the Division Education Development Plan data, which was part of the TEEP completion reports. This dataset has aggregated TEEP inputs during SY 2000/01 to SY 2004/05. However, it does not identify implementation timing and inputs of different components of TEEP. Furthermore, the completeness and quality of the data substantially vary across divisions. To overcome this gap in the data, we visited 23 TEEP division offices to find the raw data on TEEP investments. The raw data we collected reveal details of different TEEP investments: textbooks, training, school-based management, school building, school innovation and improvement fund, equipment/furniture, and supplementary instructional materials. For training, we identified the starting date of teacher training and calculated the total number of man-hours spent in training during SY 2000/01 to SY 2004/05 by different categories. For textbooks, we identified investment amounts (quantity and cost by grade and subject) in each school year. Similarly, we sorted school building projects by completion year and identified new construction and renovation cases and their aggregate total values by school.

Table 4.2 describes the initial implementation timing of different TEEP components: school building new construction and renovation, textbooks, and teacher training. The table shows the percentage of schools covered under TEEP in Visayas (our analysis is restricted to this area) from SY 2000/01 through SY 2005/06. In school buildings, we aggregated new construction and renovation projects by their completion timings. In textbooks, we used timing in which textbooks (disaggregated by grade and subject) were distributed to schools. In teacher training, we only used the initial time when training was introduced. Note that training covers a wide range of contents, which principals and teachers studied step-by-step. In many cases, training was conducted at the school district level. This means that instructors visit districts one by one within a division, and therefore it took them a few years to cover all the topics (our data show only total man-hours and the start date). The table shows that by SY 2002/03, about 80 percent of schools had received textbooks and 50 percent had at least one completed school building project. In all schools, the training process had just begun.

Table 4.2—Percentage of TEEP schools in the Visayas region, by the initial implementation timing

	SY 2000/01	SY 2001/02	SY 2002/03	SY 2003/04	SY 2004/05	SY 2005/06
			(percent)			
New construction and renovation projects	6	22	49	63	84	86
Grade 1 textbook distribution	76	76	81	100	100	100
Grade 2 textbook distribution	76	76	81	100	100	100
Grade 3 textbook distribution	76	76	81	81	81	100
Grade 4 textbook distribution	76	76	81	100	100	100
Grade 5 textbook distribution	76	76	81	100	100	100
Grade 6 textbook distribution	69	69	74	100	100	100
Training program of teachers	31	99	100	100	100	100

Source: TEEP investment database (the authors' survey), and Division Education Development Plan database.

4.2 Estimation method

Because the original phase-in plan of TEEP was not followed in practice, we cannot explore the pipeline design to identify the impact of TEEP on school performance. Therefore, we formed a control group based on the schools in the non-TEEP provinces to estimate the counterfactual of the treatment group, which are the schools in the TEEP provinces. Double differences (DD) based on the cohort panel from grade 4 (SY 2002/03) and grade 6 (SY 2004/05) is used to eliminate cohort-specific fixed effects.[10] For details, see Annex 1 Yamauchi and Liu [2011a].

Because the allocation of TEEP was purposive, the initial school conditions are likely to have different distributions in the treatment and control groups. If the initial conditions affect subsequent changes of the outcome variables, DD would give a biased estimate of the TEEP impacts. We use two strategies to deal with the potential bias due to nonrandom program placement. First, we use the sample from Visayas only. As shown in Figure 2.1, TEEP divisions are relatively evenly distributed throughout Visayas compared with the other two macroregions. We therefore expect that the TEEP and non-TEEP provinces are more comparable in Visayas, and hence our extra data collection and cleaning efforts were focused on Visayas. Second, we use propensity score (PS) matching to balance observable cohort characteristics and initial conditions between the treated and the control groups.

Three caveats exist in our method. First, our baseline is not free of contamination. Table 4.1 showed that TEEP had been implemented in all treated schools by SY 2002/03. Thus, the initial level of test scores in the treatment group reflects earlier investments completed before SY 2002/03. Second, it is possible that students from primary schools, which are not part of our sample, came into grades 5 and 6 in our sample elementary schools, which alters the student body at grade 5. Since TEEP also contributed to the conversion of primary schools to elementary schools by building new classrooms and staffing for grades 5 and 6, it is possible that attrition is different in the treated and control groups.[11] Third, as an observational analysis, we cannot eliminate bias due to time-variant unobservables.

4.3 Empirical findings

4.3.1 Average treatment effects

In the estimation, we merged NAT grade 4 in SY 2002/03 and NAT grade 6 in SY 2004/05 using elementary schools in SY 2002/03.[12] Although the selection of TEEP is based on province-level poverty indicators summarized in the Social Reform Agenda, we conjecture that income

[10] Due to delayed preparations at the early stage of TEEP, most of the program schools received investments during or after SY 2002/03.

[11] In SY 2002/03, total grade 5 enrollment was 94.1 percent of the total grade 4 enrollment in TEEP schools on average, compared with 95.4 percent in non-TEEP schools; and the total grade 6 enrollment was 94.6 percent of the total grade 5 enrollment in TEEP schools on average, compared with 95.5 percent in non-TEEP schools.

[12] Our analysis pertains only to elementary schools in SY 2002/03, which offered grades 1 to 6. To maintain a valid cohort, we dropped primary schools, where only grades 1 to 4 are taught.

distributions overlap between TEEP and non-TEEP school districts (see Figures 3.1 and 3.2). In our matching estimation, we control for the interactions of municipality income category and regional dummies, as well as school-level initial conditions including pupil–teacher ratio, grade 4 total enrollment, number of multi-grade classes, and proportion of locally funded teachers. In the Philippine context, local income level not only summarizes broad socioeconomic factors but also proxies the availability of private schools, which affects the competition between public and private schools and therefore the ability distribution of students in public schools (see, for example, Yamauchi 2005). It also controls local labor market conditions.

The first-stage logit regression result is reported in Table 4.3. The dependent variable is 1 if the school is located in a TEEP area and zero otherwise. The results show that income categories, distinguished by regions, significantly explain TEEP placement. Except for income category 5, which is the poorest group, the effect is monotonic. In eastern Visayas, which is omitted as the benchmark case, the effect of income category 5 is negative. In other regions, western and central Visayas, the income effect is monotonic throughout all income classes.

Table 4.3—Logit estimation of TEEP placement

TEEP	Coeff.	
Central Visayas	-2.163***	(0.211)
Western Visayas	-2.518***	(0.226)
Income 2	1.168***	(0.310)
Income 3	1.872***	(0.367)
Income 4	0.306	(0.190)
Income 5	0.142	(0.186)
Central Visayas × Income 2	-1.163***	(0.421)
Central Visayas × Income 3	-1.267***	(0.423)
Central Visayas × Income 4	0.332	(0.259)
Central Visayas × Income 5	-1.977***	(0.388)
Western Visayas × Income 2	-0.610	(0.398)
Western Visayas × Income 3	-1.081**	(0.424)
Western Visayas × Income 4	1.279***	(0.263)
Western Visayas × Income 5	0.954***	(0.312)
Pupil–teacher ratio (both local and national)	-0.00818*	(0.00434)
Grade 4 total enrollment (in ages 6 to 11)	-0.00766***	(0.00141)
Number of multigrade classes	-0.0412	(0.0402)
Proportion of local funded teachers	0.233	(0.595)
Constant	1.294***	(0.212)
Number of observations	4222	
Pseudo R2	0.219	

Source: National Achievement Test database, TEEP investment database (the authors' survey), Division Education Development Plan database, Basic Education Information System database, Census 2000 Municipality Income Classifications.

Note: *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

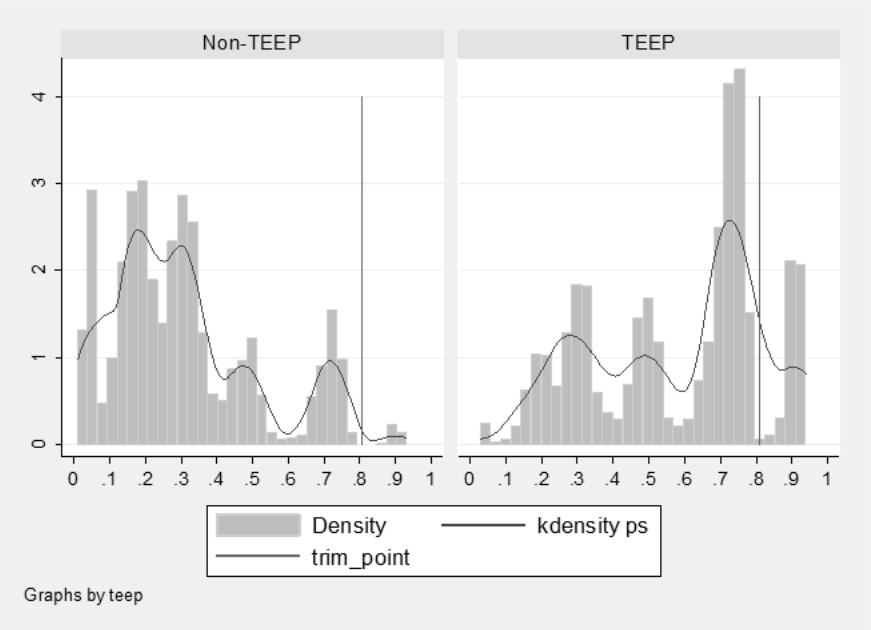
The pseudo R-squared of the logit regression is 0.22, which suggests plausible explanatory power. The PS of each observation is estimated based on the regression. Figure 4.3 plots densities of the estimated PS in the treatment and control groups as well as the cut-point of the PS values above which observations are trimmed. To illustrate the effects of trimming and reweighting, Appendix Table 4.4 displays simple differences of the explanatory variables between the treatment and control groups in the untrimmed sample and the PS weighted and trimmed samples. Although simple differences between the groups are large and statistically significant in the untrimmed sample, trimming and matching based on the propensity score eliminates all significant differences.

Table 4.4 Balance check

variables	Untrimmed sample, simple DD			Trimmed sample, PS weighted regression			Trimmed sample, PS weighted kernel		
	Diff.	s.e.	Sig.	diff.	s.e.	sig.	diff3	se3	sig
Central Visayas	-0.284	0.047	***	-0.003	0.046		-0.012	0.044	
Western Visayas	-0.147	0.050	***	0.000	0.055		0.000	0.058	
Income 2	0.002	0.031		0.002	0.017		-0.003	0.021	
Income 3	0.000	0.040		0.000	0.035		-0.004	0.031	
Income 4	0.104	0.050	**	0.004	0.062		0.024	0.064	
Income 5	0.022	0.039		-0.001	0.054		-0.001	0.049	
Central Visayas × Income 2	-0.024	0.015		0.000	0.010		-0.002	0.009	
Central Visayas × Income 3	-0.026	0.026		-0.001	0.025		-0.002	0.025	
Central Visayas × Income 4	-0.047	0.032		-0.002	0.032		-0.001	0.029	
Central Visayas × Income 5	-0.101	0.020	***	0.000	0.005		-0.002	0.006	
Western Visayas × Income 2	-0.032	0.019		0.000	0.014		-0.004	0.015	
Western Visayas × Income 3	-0.040	0.027		0.000	0.025		-0.004	0.025	
Western Visayas × Income 4	0.021	0.039		0.000	0.047		0.007	0.044	
Western Visayas × Income 5	-0.008	0.014		-0.001	0.014		0.002	0.014	
Pupil–teacher ratio	-2.215	0.758	***	-1.075	0.847		-1.282	0.841	
Grade 4 total enrollment	-7.381	1.323	***	0.716	1.194		0.584	1.098	
Number of multi-grade classes	0.134	0.049	***	-0.039	0.076		-0.042	0.083	
Proportion of local funded teachers	-0.005	0.003		-0.001	0.004		0.000	0.004	
Number of observations	4222			3963			3963		

Source: National Achievement Test database, TEEP investment database (the authors' survey), Division Education Development Plan database, Basic Education Information System database, Census 2000 Municipality Income Classifications.
Notes: DD: Double difference, PS: Propensity score, s.e.: Standard errors, diff: mean-difference, *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Figure 4.3 Plot of estimated propensity scores for schools in non-TEEP and TEEP areas



Source: National Achievement Test database, TEEP investment database (the authors' survey), Division Education Development Plan database, Basic Education Information System database, Census 2000 Municipality Income Classifications.

In Table 4.5, we report the estimation results on the average treatment effect (ATE) of TEEP. We examine changes in overall and mathematics NAT scores from grade 4 in SY 2002/03 to grade 6 in SY 2004/05.[13] Panel 1 shows the simple DD results for the overall test and mathematics test scores. The effects on both scores are small in magnitude and insignificant statistically. Panels 2 and 3 show the results using DD and PS matching (weighted regression) and DD and PS matching (kernel), respectively. The two methods give close results, which suggests that TEEP has significant impacts on both overall and mathematics scores. The magnitude is about 4 overall and 5 for mathematics. In other words, TEEP attributes to an increase of about 6 percent in the overall test score and 8 percent in the mathematics score on average.[14] The impact is not trivial over the two-year period. If the impact can continue at the same rate, the total effect of TEEP over six years (if students are exposed to TEEP in the entire elementary school period) would be a score increase of about 12 to 15 points. This magnitude of performance improvement is substantial. We note that the DD and PS matching estimates of the TEEP impacts are larger than the simple DD estimates, which implies that the endogenous allocation of TEEP creates downward bias in the estimates if the program allocation is not taken into account. That is, it is likely that TEEP schools (and school districts) would tend to have a lower trend in NAT than non-TEEP schools if TEEP were not in place.

[13] Mathematics is the only common subject that was tested by all schools in the two grades. Overall score is the summation of scores of all the subjects being tested. We used percentage scores.
[14] This is computed by dividing the estimated ATE of TEEP by the counterfactual average score of the trimmed treatment group in SY 2004/05.

Table 4.5—Impacts of TEEP on school performance

	Untrimmed sample, simple DD				
	Treated diff	**Control diff**	**DD**	**s.e.**	**sig.**
Overall score	16.737	15.348	1.389	0.874	
Math score	17.645	16.385	1.260	1.090	
Number of observations	1,774	2,434			
	Trimmed sample, DD+PS weighted regression				
Overall score	16.074	12.139	3.934	1.129	***
Math score	16.961	11.719	5.242	1.473	***
Number of observations	1,541	2,408			
	Trimmed sample, DD+PS weighted kernel				
Overall score	16.074	12.260	3.813	1.172	***
Math score	16.961	11.961	5.000	1.442	***
Number of observations	1,541	2,408			

Source: National Achievement Test database, TEEP investment database (the authors' survey), Division Education Development Plan database, Basic Education Information System database, Census 2000 Municipality Income Classifications.

Notes: DD: double difference; PS: propensity score; diff: mean-difference; s.e.: standard errors; *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

4.3.2 Component effects

The previous analysis suggests that TEEP, as a whole, has a significant effect on school performance. Because TEEP is a combination of several components, in this section we explore how each component contributes to school performance. To do so, we specify the empirical model as

$$\Delta H = \alpha + \beta_1 \Delta Textbook + \beta_2 \Delta Training + \beta_3 \Delta Building + z\gamma + \varepsilon,$$

where ΔH is the change in human capital (measured by test scores) from SY 2002/03 to SY 2004/05. $\Delta Textbook$, $\Delta Training$, and $\Delta Building$ are TEEP investments in textbooks, teacher training, and building, respectively, that are expected to benefit the cohort under study.[15] Investments in textbooks include those for grades 4, 5, and 6 separately. Investments in training include instruction training and subjective training of teacher. Investments in building refer to the number of new school constructions and new renovations; z is a vector of the initial district- and school-level conditions including the interactions of municipality-level income categories and regional dummies, pupil–teacher ratio, grade 4 enrollment, number of multi-grade classes, and proportion of local funded teachers. We note that the initial human capital and TEEP investments are potentially complementary (and thus not separable), but we assume that the initial school conditions are sufficient to control such heterogeneities in the intervention effect.

The results are presented in Table 4.6, both for the entire sample and for the TEEP-only sample. The findings are summarized as follows: First, in the textbook effect, earlier stage investments

[15] For example, grade 4 textbook refers to the textbooks distributed to grade 4 in SY 2002/03. The grade 4 textbook distributed to grade 4 in SY 2003/04 is not counted because it did not benefit our cohort.

seem very important in determining later stage outcomes. Grade 4 textbook affects student outcomes from grade 4 to grade 6 onward. This finding is consistent with the recently well established view on the cumulative process of human capital accumulation. Second, new classroom construction significantly helps improve their performance. The effect of renovations is also significant, although it has a much lower magnitude. Third, instructional training seems to have a greater positive effect on student performance than subject-wise training (mathematics, English, and so forth). The latter has a negative effect on student performance, at least in the short run, probably because teachers have to use their teaching time to receive training.

This analysis has some reservations. First, since our sample students (cohorts) are at grade 4 in SY 2002/03, we focus on textbooks for grades 4 to 6 distributed at TEEP. These students (cohorts) could have used TEEP textbooks at lower grades, but the impacts of the textbooks are already reflected in their NAT scores at SY 2002/03 (grade 4). Second, although we have information on school building project contract values, we use the number of new constructions and renovations because the contract value aggregates both types and we also conjecture that the impacts are different between new constructions and renovations. These conjectures were supported in preliminary analyses.

Finally, in this study, we did not explicitly assess school-based management, mainly because we did not find appropriate input measures and variations. The batch plan was not strictly implemented, especially in the first and second batch groups (that is, they were mixed in reality, depending on the updated preparedness at the division level). This soft component is thought to improve the overall effectiveness of physical investments and teacher training.

Table 4.6—Estimation results of component analysis, dependent variables being change in mathematics score and overall score

	All Sample		TEEP only	
	Overall Score	Math Score	Overall Score	MathScore
Grade 4 textbooks (peso/pupil)	0.0418***	0.0142**	0.0333***	0.0129**
Grade 5 textbooks (peso/pupil)	-0.00726	-0.000362	-0.00575	-0.00112
Grade 6 textbooks (peso/pupil)	-0.00289	-0.00243	-0.00340	-0.00251
Instructional training (man-hours/pupil)	0.487**	0.327*	0.427**	0.267*
Subject training (man-hours/pupil)	-0.849***	-0.590*	-0.619**	-0.406
New constructions (number in SY 2003/04)	5.756***	5.316***	5.390***	5.010***
New renovations (number in SY 2003/04)	1.490***	1.199**	1.116***	0.884**
Central Visayas	7.111**	-3.695	3.154	-3.224
Western Visayas	-0.613	-19.41***	-0.254	-14.18***
Income 2	4.176	2.893	4.082	2.474
Income 3	-1.293	-2.525	-0.591	-1.398
Income 4	-0.654	-0.946	-1.027	-1.509
Income 5	2.168	1.181	1.441	0.780
Central Visayas × Income 2	-1.530	-2.906	-0.736	-4.757
Central Visayas × Income 3	-1.758	-2.143	-1.156	-1.751
Central Visayas × Income 4	0.394	-4.268	0.703	-3.634
Central Visayas × Income 5	0.0249	-0.552	0.328	-1.276
Western Visayas × Income 2	-0.623	8.271*	0.0490	6.239
Western Visayas × Income 3	1.083	16.73***	0.597	11.69***
Western Visayas × Income 4	1.006	13.64***	2.310	11.89***
Western Visayas × Income 5	2.199	10.76***	2.551	9.895***
Pupil teacher ratio	-0.118**	-0.128*	-0.0990**	-0.156**
Grade 4 total enrollment	0.0473***	0.0576***	0.0464***	0.0613***
Number of multi-grade classes	-0.456	-0.117	-0.504*	0.160
Proportion of local funded teachers	-11.90*	-6.336	-8.641	-9.587
Constant	15.52***	21.67***	15.25***	20.98***
Number of observations	3905	1471	3905	1471
R-squared	0.059	0.088	0.060	0.113

Source: National Achievement Test database, TEEP investment database (the authors' survey), Division Education Development Plan database, Basic Education Information System database, Census 2000 Municipality Income Classifications.

Note: Pesos are in Philippine pesos (PHP). *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

4.3.3 Returns to test score

Next we show returns to NAT (and NEAT) score. In Table 4.7, log of wage in the latest job (available in the tracking survey) is regressed on grade-6 NAT (or NEAT) score and other conventional variables. Though parameter estimates are biased in this cross-sectional estimation, the results suggest some important mechanism of the test score impact in the long run.

First, the returns to NAT (or NEAT) are quite large in the above result. Wage change due to an increase of NAT by 15 score points if evaluated at the average log wage is approximately P 1000 per month. Therefore, the TEEP impact on NAT is translated into about annual wage increase of P 9,600 to 12,000.

Second, the results in Columns 1 and 2 show that potential ability bias in the OLS estimates of returns to schooling and experience is not large (compared to the estimates without test score).

However, more interestingly, third, test score augments returns to schooling and experience through complementarities between test score and schooling and experience. That is, returns to schooling are greater if grade-6 test score is higher. Similarly, returns to experience increases with grade-6 test score. These findings suggest that an increase in NAT score, due to TEEP intervention, potentially raise lifetime earnings of the students, by augmenting returns to subsequent schooling and labor market experience.

Table 4.7 Returns to national achievement test scores

Dependent: log monthly earnings Job:	Latest	Latest	Latest
Test score (Grade 6)	0.0022	0.0023	-0.0073
	(2.48)	(2.56)	(2.02)
Years of schooling	0.0901	0.0866	0.0515
	(8.60)	(8.33)	(2.31)
Yrs of sch * score			0.0007
			(2.19)
Experience		0.0456	0.0175
		(6.55)	(1.35)
Exp * score			0.0005
			(1.96)
Age started		0.0985	0.1081
		(1.72)	(1.88)
Age started squared		-0.0008	-0.0012
		(0.45)	(0.68)
Age	0.5139		
	(7.36)		
Age squared	-0.0110		
	(6.48)		
Female	-0.1912	-0.1775	-0.1814
	(5.26)	(4.82)	(4.80)
Province (current) fixed effects	yes	yes	yes
Number of observations	1817	1791	1791
Number of provinces	48	48	48
R squared (within)	0.1926	0.1835	0.1872

Notes: Numbers in parentheses are absolute t values with Huber robust standard errors. Current age is age started first job plus experience (years).

4.4 Summary

This paper provided evidence from the Philippines that both physical and soft components of public school education investments significantly increased student test scores, by about 12–15 score points in the National Achievement Test (NAT) with the six-year exposure. Our study also showed that the performance in mathematics is more positively responsive to education reform and investments than other subjects.

An increase in NAT can significantly increase lifetime earnings of the students at later stages. Empirical results show that an increase in NAT score by 12 to 15 points raises annual earnings by approximately 9,600 to 12,000 Peso (in terms of 2010 price).

Second, we also found evidence that early-stage investments improve student performance at later stages in the elementary school cycle. The distribution of grade 4 textbooks is shown to

increase subsequent student test scores more than grade 5 or grade 6 textbooks do. This is not surprising, due to the cumulative nature of knowledge.

5. Long-term impact analysis

5.1 Tracking data

In this section we describe the data we use in our analysis. The data come from the survey conducted in eight education divisions in the Philippines from July 2010 through April 2011. The survey aimed to gather household and individual data to assess the impact of a large school-based intervention implemented in 23 poor education divisions (provinces) in the period of 2001 to 2006. For this purpose, the survey includes 4 intervention and 4 non-intervention divisions. An intervention division is paired to an adjacent non-intervention division in the same area so that the pair shares similar socio-economic conditions. In the above method, our sample is demarked into four areas: (i) Ifugao and Neuva Vizcaya, (ii) Antique and Iloilo, (iii) Negros Oriental and Cebu, and (iv) Leyte and Western Samar. Figure 2 maps our sampled provinces in the Philippines.

In each division, first, relatively poor municipalities (school districts) were chosen. Municipalities of census-2000 income classes ranked 3 to 5 (the highest income is rank 1 and the poorest is rank 6) were chosen from the adjacent area (near the division border) of an intervention and a non-intervention divisions (see Republic of the Philippines, 2001). However, in Ifugao division, all of our school districts are taken from income classes ranked 4 and 5, which created imbalance with Neuva Vizcaya division where some of the school districts are ranked 3.

Second, schools are randomly sampled from the list of elementary schools in school year SY2002/03 satisfying three criteria: (i) total enrollment being larger than 120, (ii) mono-grade (at least one class for each grade), and (iii) complete (having grades 1 to 6). That is, schools had on average at least 20 students in each grade. In an intervention division, 15 schools were randomly sampled from the basic list satisfying the above mentioned conditions. Similarly, 10 schools were randomly sampled in a non-intervention division. In Antique (an intervention division), however, we decided to add 2 more schools since we found that 2 schools were severely damaged in flash floods caused by a typhoon in 2006 (they were relocated in safer higher grounds). Therefore, we have 17 schools in Antique. In Neuva Vizcaya, it was difficult to have student lists from one school at the early stage, so this school was omitted from our sample but we increased the number of students from two large schools (the number of sample students being the same).

Third, we collected lists of students enrolled at Grade 6 in SY 1999/00, SY2004/05 and SY2005/06. SY1999/00 is a pre-intervention cohort, while both SY2004/05 and 2005/06 are cohorts that were exposed to a school intervention if they lived in TEEP divisions. The process required a few months in each division. We randomly sampled 15 students from SY1999/00 Grade 6 (pre-TEEP cohort), while 20 students all together from SY2004/05 and 2005/06 Grade 6 (TEEP cohort). The sampling was done regardless of gender and age. Delays in entering school and repetitions create variations in age even in the same cohort. Note that since the listed students are those who were

enrolled at that time, some of our sample students might not have graduated from their elementary schools.

Table 5.1 Sample student distributions

Grade-6 school year	Antique	Cebu	Ifugao	Iloilo	Leyte	Negros Oriental	Neuva Vizcaya	Western Samar
Pre-TEEP Cohort:								
1999-2000	244	143	188	143	212	227	139	142
TEEP Cohort:								
2004-2005	159	107	137	97	158	161	92	96
2005-2006	177	95	145	111	135	145	101	97
Both Cohorts:								
Total 3451	580	341	470	351	505	533	332	335

Table 5.1 shows the composition of our sample households and students. We have the total of 3451 students in our sample. TEEP divisions and cohorts (school years of 2004/05 and 2005/06) are over-sampled. Among TEEP divisions, Ifugao shows smaller numbers in each grade-6 sample year due to the decision to drop some unreliable and unverified information in the second visit in the division.

Data collection has two components: household survey and student tracking survey. In the household survey, we gathered information on household rosters in 2010 and 2000, schooling and work histories of biological siblings (of our sample students), household income (2010) and asset holding (2010 and 2000), parents' participation in school governance for each sibling, and public service and infrastructure access (2010 and 2000). On assets owned in 2000, only quantities were captured in each category (we asked both values and quantities of assets in 2010). To impute the value of asset holdings in 2000, we used the current prices calculated from the values and quantities of assets in 2010. The survey was supplemented by barangay (community) leader, PTCA head and school surveys.

In the student tracking survey, we tracked our sample students to collect information on their schooling and work histories in detail as well as marriage, anthropometry and illness. Either face-to-face or phone interview was adopted. There were two stages in tracking activities. First, the teams tracked students who reside within their original divisions. This was done immediately after the household survey. Second, in case of out-division tracking, the teams attempted to schedule face-to-face interviews with students who reside in National Capital Region (NCR; Manila), Baguio and Cebu City. For students who reside in other provinces, we basically used phone interviews. However, the teams tried to visit students who reside near or within the province of Laguna and in between northern Luzon and Manila to conduct face-to-face interviews.

We did not track oversea migrants, and omitted several cases that could not be interviewed even through phones.[16]

Although our sample has different age groups (cohorts) due to its sampling design, the majority of our students completed four years of high school. This is highly expected since even TEEP students (grade-6 in SY 2004/05 and 2005/06) are expected to reach age 17 in the survey period (note that children graduate from high school at age 15/16 if they had no delay and repetition in schooling).

The tracking survey captured information on schooling and work histories in detail. Since our sample students were grade-6 enrollers in particular school years, we omit questions on their schooling histories up to that stage. Detailed information starts from high school entry and onward. Similar to the sibling section of the household survey, we collected information on school (name and school ID), school type (public or private), age started, graduated or not, age graduated, age stopped if not graduated, reasons for stopping and whether still in school. At the college level, we also captured course majored first and degree attained.

On their transition from school to work, the survey asked when they left their parents' households for the first time and whether they returned permanently or temporarily, the reasons, and whether they currently live with their parents. In work history, details were identified for each job. Many of our sample students have experienced several works, which often started when they were still in school. The survey collected information on job description, occupation type, employment type, industry, when started and ended, types of reference, payment types and monthly earnings.[17]

In this survey, we categorized students in three types based on their migration and household membership status: (i) in-town, (ii) migrant and (iii) transition. In the case of in-town, students stay in the same school area. They may or may not live with their parents or guardians. In (ii) migrant students, they live away from their town and also are not members of their parents/guardian households. Finally the third group of students categorized as transition students is those who live physically away from their towns, but still belong to their original households. For example, students who are temporarily staying in Manila to look for job, or who recently started living in a boarding house to attend college in distance. This group may come back to their households or permanently migrate, which was still uncertain at the time of our survey.

In our sample 69.34% of the students live in their original school areas, 22.11% are migrants and 8.55% are transitory students. If we add migrant and transitory students, more than 31% of our

[16] There were 36 cases untracked out of the total of 3487 students. Therefore, the tracking rate is about 98.97%.
[17] In family works in agriculture, hunting and forestry, we asked them to estimate average annual income (dividing total annual family incomes from harvests by the number of members who worked, and converted into monthly figure).In the first job, we have 178 cases of reported monthly earnings out of 196 cases in this category of employment type and industry. In 18 cases, they could not estimate monthly earnings.

sample students are physically living away from their parents or guardian households. In the analysis, we assign transition students to migrants.

In the analysis assessing the dynamic impact of the school intervention, we use years of schooling completed, high school choice, repetition at high school, college entry, migration choice, and earnings from the latest job.

Table 5.2 Summary statistics of the dependent variables, separated by gender and cohort

	Non-TEEP cohort				TEEP cohort			
	Female		Male		Female		Male	
	Mean	S.d.	Mean	S.d.	Mean	S.d.	Mean	S.d.
Years of schooling	10.866	2.279	10.310	2.450	9.604	1.370	8.929	1.642
High school repetition	0.026	0.170	0.217	0.567	0.060	0.255	0.279	0.678
College entry	0.517	0.500	0.430	0.495	0.413	0.493	0.279	0.449
Migration	0.392	0.489	0.277	0.448	0.197	0.398	0.084	0.277
Log(salary)	8.147	0.750	8.322	0.768	7.598	0.478	7.640	0.801

Source: Tracking Survey.

Table 5.2 report the mean and standard deviation of the dependent variables, separated by gender and cohort. The table suggests that, on average, females outperformed males in all of the three indicators on school performance for both TEEP and non-TEEP cohorts. For the non-TEEP cohort, females on average have about 10.9 years of schooling, then number of high school repetitions is about 0.026, and 51 percent entered college. While for males, the number of school years, number of high school repetitions, and percentage entering college are 10.3, 0.127, and 42.4%, respectively. For the TEEP and younger cohort, females have 9.6 years of schooling, 0.06 high school repetitions, and 41% college entry rate, compared with 8.9 years of schooling, 0.279 high school repetitions, and 27.8% college entry rate for males. Females are also more likely to migrate, with migration rate being 39.7% for non-TEEP cohort and 20% for TEEP cohort, compared with 28.5% and 8.6%, respectively, for males. Although females outperformed males in schooling and migration, the monthly salary of females is merely 84% of males' for the non-TEEP cohort and 96% of males' for the TEEP cohort.

Table 5.3 Sample means of variables for face-to-face interviews and phone interviews

Variable	Face-to-face	Phone	sig
Years of schooling	9.79	9.72	
Repetitions in high school	0.15	0.13	
College entry	0.40	0.29	***
Migration	0.16	0.73	***
TEEP provinces	0.60	0.67	***
Females	0.50	0.60	***
Log(consumer durables)	11.08	10.76	***
Mother's years of schooling	7.48	6.45	***
Mother's age	49.16	49.98	*
Size of siblings	5.52	6.22	***
Age	19.58	20.18	***
Number of observations	2724	330	

Source: Tracking Survey.

Table 5.3 compares the means of key variables between the two modes of interview: face-to-face and phone. Except years of schooling completed and the number of repetitions at high school, all these variables show statistically significant differences. Migration rate is particularly higher in phone interviews due to the above mentioned method of our survey. Note that most of face-to-face interviews were done within division. In our analyses, we include indicators of interview mode to control potential differences in unobserved characteristics correlated with the mode.

We take necessary measures of controlling potential factors that induced differentiated trends between TEEP and non-TEEP divisions. Two possibilities are considered in our analysis. First, potential changes in secondary school quality might have been different between the two groups. Second, progress in social development in general might have been different, which potentially create differentiated trends between the two groups.

In our TEEP divisions, there was a simultaneous project on secondary school education, Secondary Education Development Improvement Project (SEDIP). Since the implementation was supposed to start in 2000 (and end in 2006), we could have differenced out the impact on both pre-TEEP and TEEP cohorts, but due to a substantial delay in its implementation, SEDIP in our divisions started in late 2002. Therefore, it is necessary to control potentially different trends of high school quality. Table 5.4 compares the hypothetical number of students per academic classroom in high schools in SY2002/03 and SY2007/08 by our sample divisions. We used the previous year's numbers of students enrolled in the first year, being multiplied by 4 to obtain the hypothetical size of total enrollment. This is the size of enrollment if there is no drop-out in subsequent years. We observed large numbers of new classrooms in Neuva Vizcaya (non-TEEP) and Antique and Leyte (both TEEP). Interestingly, we observe that this measure of school quality has converged between adjacent TEEP and non-TEEP divisions.

Table 5.4 Estimated students per class room			
Province	SY 2002/03	SY 2007/08	% Change
Ifugao	99.31	58.57	-41.03%
Neuva Viscaya	93.37	53.59	-42.61%
Antique	156.31	59.12	-62.18%
Iloilo	49.92	58.7	17.58%
Negros Oriental	152.49	75.04	-50.79%
Cebu	68.97	72.78	5.53%
Leyte	54.26	71.3	31.40%
Western Samar	36.11	79.91	121.34%

Source: BEIS.

Table 5.5 Availability of different type of infrastructure in student's residence village (averaged in school areas)									
	2000			2010			(2010-2000)		
Variable	TEEP	Non-TEEP		TEEP	Non-TEEP		TEEP	Non-TEEP	
Electricity	0.865	0.838		0.988	0.972		0.123	0.134	
Piped water	0.422	0.349		0.642	0.595		0.22	0.245	
Paved road	0.388	0.451		0.518	0.709	***	0.129	0.258	*
Modern irrigation	0.197	0.124		0.294	0.238		0.097	0.114	
Cell phone service	0.611	0.624		0.951	0.95		0.34	0.326	
Internet café	0.02	0.031		0.075	0.068		0.055	0.037	
Market	0.077	0.081		0.079	0.111		0.002	0.03	*
Public elementary	0.917	0.795	***	0.891	0.804	**	-0.026	0.009	
Private elementar	0.018	0.038		0.154	0.154		0.136	0.117	
Public high school	0.228	0.192		0.312	0.308		0.084	0.116	
Private high schoc	0.04	0.032		0.123	0.121		0.083	0.089	
Health center	0.784	0.596	***	0.766	0.619	**	-0.018	0.023	
Public library	0.024	0.02		0.046	0.031		0.022	0.011	
Post office	0.036	0.017		0.026	0.038		-0.01	0.02	**
Number of observ	62	39		62	39		62	39	

Source: Household Survey.

TEEP divisions were also identified in Social Reform Agenda (SRA). There were other government efforts in SRA, such as those supporting agrarian reforms in agricultural communities. These government programs, possibly simultaneously introduced in our sample communities, imply that (i) income level increased faster and (ii) access to social services and infrastructures improved better than non-TEEP counterparts. These changes also affect schooling decisions. Therefore, it is necessary to control changes occurred in economic conditions in these communities. In Table 5.5, we compare measures of access to public services and infrastructures between TEEP and non-TEEP areas. The household survey asked each household (mother or guardian) whether they had access to 14 types of public services or infrastructures in their barangays in 2000 and 2010. We took average of their responses to compute the proportion of households that had access to each type. The table shows that except public elementary school, health center and paved road, we do not confirm significant differences in 2000 and/or 2010.

Interestingly, trends in the period of 2000 to 2010 are marginally significantly different in paved road, access to market and post office.

In our analysis, we control potential effects of other SRA programs by using (i) provincial-level poverty incidence rates (available from the National Statistical Coordination Board), and (ii) the household-level measures of the average access to public services and infrastructures (described above). First, the average of poverty incidence rates in 2000 and 2003 (2006 and 2009) was used to represent the effect of economic conditions on pre-TEEP cohort (TEEP cohort). Second, we take average of access indicators at the household level (not barangay). Then, the average access measure of 2000 is used for TEEP cohort, while the average of 2000 and 2010 (representing the situation around 2005) is used for TEEP cohorts. We include these variables in regressions to check robustness of our key results. For controlling potential effects of high school quality change, we use the hypothetical numbers of students per academic classrooms in SY 2002/03 and 2007/08 for pre-TEEP and TEEP cohorts, respectively.

5.2 Empirical framework

Our data tracked the school and labor market performance of TEEP cohort (students enrolled in Grade 6 in SY2004/05 or SY2005/06) and non-TEEP cohort (students enrolled in Grade 6 in SY1999/2000) in both treatment and control schools. This data structure allows us to use double differences (DD) to identify the gender-specific impacts of TEEP. The dependent variables fall into three categories: students' school performance, migration, and return in labor market. School performance is indicated by years of schooling, number of repetitions in high school, and college entry. For details, see Annex 2 Yamauchi and Liu [2011b].

As discussed, the placement of TEEP is not random as TEEP tends to target poorer provinces. Our estimates will be biased if we omit some variables that jointly affect outcomes and TEEP placement. To deal with this problem, we estimate a selection function and weigh observations with the estimated propensity scores of schools to obtain double robustness.[18] To further check robustness of our empirical results, we also run the regressions using trimmed sample which trimmed off the observations with extreme propensity scores.

We explicitly control potential factors that induce trends differentiated between TEEP and non-TEEP divisions. Since we control school fixed effects in all estimations, we include controls specific to pre-TEEP and TEEP cohorts in each school. These include poverty incidence rates, infrastructure access measures, and high school quality measures. We define these variables in the next section. In addition, we include an indicator of interview mode: face-to-face or phone interview, to control unobserved characteristics that are potentially correlated with the mode choice in the tracking survey.

[18] Double robustness means, if the main regression is miss-specified however the selection function is correctly specified, the estimates based on the reweighted regression are still consistent.

5.3. Empirical findings

5.3.1 Schooling attainment

To provide appropriate weights for the analysis, we estimate a logit selection function of TEEP schools in which we control for initial school quality and poverty level of the community of school location. School quality is proxied by pupil-teacher ratio, total student enrollment in Grade 6, and total classes provided by the school. These variables are based on the 2002/03 Basic Education Information System (BEIS) data. Community level poverty is proxied by the mean of the logarithm of the value of consumer durables of the sampled households from the school. We also include school location dummies and their interactions with pupil-teacher ratio and community poverty. The results are reported in Annex 2 (Table A1). Although only three variables are statistically significant, the overall explanatory power is favorable (the psuedo R-squared is 0.144).

Tables 5.6, 5.7, and 5.8 report the regression results of the three schooling outcomes, years of schooling, number of repetitions in high school, and college entry, respectively. In each table, we report the result for the untrimmed and un-weighted sample (full sample) and trimmed and weighted sample. In either sample, we estimate two specifications, with and without cohort-specific controls (poverty incidence, infrastructure index, number of students per classroom in high school). We also include the mode of interview (if phone interview) in Models 2 and 4 to control the choice of interview methods that are potentially correlated with unobserved characteristics.

Table 5.6 Regression results of years of schooling, for full sample and trimmed and weighted sample

	Full Sample				Trimmed and Weighted Sample			
	Model 1		Model 2		Model 3		**Model 4**	
TEEP cohort*TEEP province	-0.41	0.76	-0.862	1.57	-0.427	0.73	**-0.734**	**1.25**
Female	1.419**	2.44	1.457**	2.53	1.892**	3.27	**1.920****	**3.32**
Female*TEEP cohort	-0.183	1.15	-0.184	1.15	-0.2	1.23	**-0.195**	**1.18**
Female*TEEP cohort*TEEP province	0.281**	2.01	0.289**	2.07	0.337**	2	**0.343****	**2.05**
Log(consumer durables)	0.426***	7.6	0.391***	6.64	0.463***	7.96	**0.442*****	**7.33**
TEEP cohort	0.91	1.28	0.593	0.84	0.969	1.47	**0.806**	**1.27**
TEEP cohort* Log(consumer durables)	-0.362***	6.13	-0.325***	5.42	-0.375***	7.29	**-0.353*****	**7.02**
Female*Log(consumer durables)	-0.0691	1.38	-0.0726	1.47	-0.110**	2.31	**-0.113****	**2.37**
TEEP cohort*TEEP province *Log(consumer durables)	0.0046	0.1	0.047	1.01	-0.00406	0.08	**0.0236**	**0.48**
Age	1.402***	4.11	1.467***	4.2	1.433***	3.72	**1.503*****	**3.79**
Age-squared	-0.0442***	4.82	-0.0458***	4.89	-0.0455***	4.39	**-0.0472*****	**4.43**
Mother's years of schooling	0.127***	11.28	0.126***	11.23	0.134***	10.58	**0.132*****	**10.53**
Mother's age	0.0200***	5.15	0.0197***	5.14	0.0198***	4.98	**0.0197*****	**5.04**
Number of siblings	-0.0275**	2.05	-0.0277**	2.07	-0.0277**	2.15	**-0.0284****	**2.22**
Poverty incidence rate			0.0410***	3.79			**0.0346****	**3.16**
Infrastructure index			0.452	1.59			**0.634****	**2.22**
Number of students per classroom in high school			-0.00171	0.89			**-0.00143**	**0.73**
If phone interview			-0.0269	0.26			**-0.11**	**0.82**
Other variables and school fixed effects	yes		yes		yes		**yes**	
Number of observations	2962		2962		2761		**2761**	

*** 1%, **5%, *10% significance. Absolute t values are shown next to parameter estimates.

Poverty incidence and number of students per classroom are at the province level. The infrastructure index is generated by the average of the dummy variables indicating access to each of the 14 types of public services or infrastructures (see Table 4) at the household level. To make them corresponding to the high school period for either cohort, we assign the average of poverty incidence rates in 2000 and 2003 to the pre-TEEP cohort and the average rate of 2006 and 2009 to the TEEP cohort. Similarly, we use the infrastructure index of 2000 the pre-TEEP cohort while the average of 2000 and 2010 (representing the situation around 2005) for TEEP cohorts, and use the hypothetical number of students per academic classrooms in SY 2002/03 and 2007/08 for pre-TEEP and TEEP cohorts, respectively.

We use the logarithm of total value of consumer durables, log (consumer durables), to proxy for liquidity constraint. We interact Female and log (consumer durables) with a variety of variables to fully capture gender-specific TEEP effects and the effects of liquidity constraint.

The estimation results are mostly consistent across the four models/regressions. Our interpretation is based on the results of Model 4 for each outcome. The findings are summarized as follows. First, Female is significant in the functions of years of schooling completed and high school repetitions, pointing to a female advantage in school performance. This is consistent with our observation in Table 2. The interaction term of Female and TEEP cohort is insignificant at any conventional levels, suggesting female advantage is not changed between cohorts. The variable Female × TEEP cohort × TEEP province is statistically significant (at the 5% level for years of schooling completed and high school repetitions, and 10% level for college entry) and its coefficients have the predicted signs for all the three indicators. This result points to an increased female advantage due to TEEP intervention. The gender-specific TEEP effect is non-trivial in magnitude: the change of female-male gap due to TEEP is estimated to be 0.34 school years, -0.14 repetitions in high school, and 9% college entry rate, ceteris paribus.

Table 5.7 Regression results of number of repetitions in high school, for full sample and trimmed and weighted sample								
	Full Sample				Trimmed and Weighted Sample			
	Model 1			Model 2	Model 3		**Model 4**	
TEEP cohort*TEEP province	0.118	0.61	0.0804	0.4	0.29	1.13	**0.241**	**0.95**
Female	-0.455***	3.47	-0.453***	3.49	-0.383***	2.85	**-0.380*** **	**2.86**
Female*TEEP cohort	0.0572	1.25	0.0582	1.27	0.0359	0.75	**0.0376**	**0.78**
Female*TEEP cohort*TEEP province	-0.122**	2.27	-0.121**	2.23	-0.143**	2.32	**-0.143** **	**2.3**
Log(consumer durables)	-0.0159	1.42	-0.0204*	1.77	-0.00563	0.42	**-0.00873**	**0.63**
TEEP cohort	0.205	1.14	0.157	0.85	0.275	1.28	**0.253**	**1.14**
TEEP cohort* Log(consumer durables)	0.00305	0.2	0.00812	0.51	0.00534	0.29	**0.00886**	**0.47**
Female*Log(consumer durables)	0.0235**	2.08	0.0232**	2.07	0.0204*	1.66	**0.0200***	**1.65**
TEEP cohort*TEEP province *Log(consumer durables)	-0.00452	0.26	1.34E-05	0	-0.0174	0.79	**-0.0127**	**0.58**
Age	0.0391	0.44	0.0524	0.59	-0.0218	0.2	**-0.00405**	**0.03**
Age-squared	2.89E-05	0.01	-0.00032	0.14	0.00193	0.67	**0.00147**	**0.49**
Mother's years of schooling	-0.00139	0.4	-0.00155	0.44	0.0014	0.26	**0.00134**	**0.25**
Mother's age	-0.00049	0.35	-0.00054	0.38	0.000217	0.13	**0.00019**	**0.11**
Number of siblings	0.0029	0.65	0.00278	0.63	0.0057	0.82	**0.00559**	**0.81**
Poverty incidence rate			0.00527*	1.97			**0.00634** **	**2.15**
Infrastructure index			0.0573	0.45			**-0.0133**	**0.08**
Number of students per classroom in high school			-5.5E-05	0.11			**-0.00019**	**0.39**
If phone interview			0.00434	0.15			**0.00127**	**0.04**
Other variables and school fixed effects	yes		yes		yes		**yes**	
Number of observations	2776		2776		2585		**2585**	

*** 1%, **5%, *10% significance. Absolute t values are shown next to parameter estimates.

Second, the variable Log (consumer durables) being positive and significant at the 1% for years of schooling and college entry is in line with our expectation that liquidity constraint plays an important role in schooling attainment. The interaction of Log (consumer durables) and TEEP cohort is significant at the 1% level in the years of schooling equation, suggesting that liquidity constraint is likely to be more binding at later stages of education. The interaction term of female and Log (consumer durables) is significant and its coefficient has the opposite sign to the coefficient of Log (consumer durables) when the dependent variable is years of schooling and number of repetitions in high school. This result suggests that females are less affected by liquidity constraint, which increases years of schooling completed and reduces high school repetition.

Table 5.8 Regression results of college entry (linear probability model), for full sample and trimmed and weighted sample								
	Full Sample				Trimmed and Weighted Sample			
	Model 1		Model 2		Model 3		**Model 4**	
TEEP cohort*TEEP province	-0.23	1.18	-0.306	1.55	-0.355*	1.67	**-0.395***	**1.89**
Female	0.0442	0.34	0.054	0.42	0.0979	0.72	**0.1**	**0.73**
Female*TEEP cohort	-0.035	0.77	-0.0372	0.81	-0.0357	0.71	**-0.0357**	**0.71**
Female*TEEP cohort*TEEP province	0.0983**	2.28	0.0992**	2.31	0.0893*	1.82	**0.0892***	**1.82**
Log(consumer durables)	0.0561***	4.58	0.0544***	4.17	0.0651***	5.32	**0.0640*****	**4.97**
TEEP cohort	-0.25	1.27	-0.241	1.19	-0.222	1.16	**-0.217**	**1.11**
TEEP cohort* Log(consumer durables)	-0.0216	1.28	-0.021	1.19	-0.0259	1.55	**-0.0253**	**1.46**
Female*Log(consumer durables)	0.0059	0.52	0.00536	0.47	0.00128	0.11	**0.00117**	**0.1**
TEEP cohort*TEEP province *Log(consumer durables)	0.0157	0.89	0.0203	1.15	0.0257	1.32	**0.0281**	**1.47**
Age	0.303***	3.77	0.297***	3.54	0.275***	2.77	**0.277*****	**2.77**
Age-squared	-0.00933***	4.48	-0.00915***	4.22	-0.00876***	3.41	**-0.00880*****	**3.4**
Mother's years of schooling	0.0340***	10.79	0.0339***	10.89	0.0324***	9.18	**0.0324*****	**9.23**
Mother's age	0.00371***	3.37	0.00370***	3.35	0.00369***	2.97	**0.00368*****	**2.91**
Number of siblings	-0.0129***	4.05	-0.0128***	3.94	-0.0124***	3.5	**-0.0123*****	**3.36**
Poverty incidence rate			0.00275	0.89			**0.00259**	**0.82**
Infrastructure index			-0.00935	0.11			**-0.00107**	**0.01**
Number of students per classroom in high school			-0.00056	1.32			**-0.0003**	**0.55**
If phone interview			-0.0499*	1.88			**-0.0353**	**1.05**
Other variables and school fixed effects	yes		yes		yes		**yes**	
Percentage of predicted value out of [0,1]	5.10%		5.00%		6.60%		**6.20%**	
Number of observations	2962		2962		2761		**2761**	

*** 1%, **5%, *10% significance. Absolute t values are shown next to parameter estimates.

Third, the interaction term of TEEP cohort and TEEP province, is insignificant for years of schooling and repetitions in high school, and marginally significant (but has a negative sign), which suggests that TEEP barely have any impacts on males.

Fourth, an increase in poverty incidence is significantly positively correlated with increased years of schooling and repetitions at high school. Access to public services and infrastructures is positively related to years of schooling. Increased poverty seems to decrease opportunity costs of schooling, which increases schooling attainment. The positive effect on repetitions at high school may be also due to the same reason, i.e., poor-performing students are more likely to stay at school with lower opportunity cost of schooling. Improved access to public services and infrastructures should be positively correlated with income level, which seems to support child schooling.

The effects of other variables are: having educated and older mother and fewer siblings all contribute to better education performance; age has a positive and decreasing effect on school performance; and cohort-specific controls have some explanatory power in schooling attainment.

5.3.2 Migration

The results on the migration are reported in Table 5.9. We use the same model specification as that for the schooling outcomes, except that the mode of interview (if phone interview) is not included because it is endogenous here.

We interpret the results based mainly on Model 4. The major findings are as follows. First, the variable Female is significant at the 1% level and has estimated coefficient of 0.38, indicating that being female increases the chances of migration by 38 percentage points, ceteris paribus. The interaction of Female and TEEP cohort being insignificant suggests females' inclination to migrate does not change between cohorts. The interaction of Female and Log (consumer durables) is negative and significant at the 5% level, suggesting poorer females are more likely to migrate. This is reasonable given that the opportunity cost of migration is lower for the poorer.

Table 5.9 Regression results of migration (linear probability model), for full sample and trimmed and weighted sample

	Full Sample				Trimmed and Weighted Sample			
	Model 1		Model 2		Model 3		**Model 4**	
TEEP cohort*TEEP province	0.222	1.6	0.149	1.07	0.244*	1.65	**0.137**	**0.94**
Female	0.394***	3.71	0.400***	3.75	0.365***	3.11	**0.375*** **	**3.3**
Female*TEEP cohort	-0.0369	0.95	-0.0373	0.97	-0.0111	0.26	**-0.0131**	**0.32**
Female*TEEP cohort*TEEP province	0.0607*	1.91	0.0608*	1.91	0.0524	1.51	**0.0514**	**1.51**
Log(consumer durables)	0.0072	0.62	0.00282	0.23	-0.00173	0.11	**-0.00144**	**0.09**
TEEP cohort	0.0149	0.09	-0.0168	0.1	-0.0383	0.2	**0.00651**	**0.03**
TEEP cohort* Log(consumer durables)	0.00321	0.25	0.00766	0.55	0.0078	0.48	**0.00625**	**0.36**
Female*Log(consumer durables)	-0.0250***	2.85	-0.0254***	2.89	-0.0239**	2.43	**-0.0246** **	**2.56**
TEEP cohort*TEEP province *Log(consumer durables)	-0.0277**	2.24	-0.0210*	1.72	-0.0276**	2.09	**-0.0223***	**1.7**
Age	-0.111	1.52	-0.101	1.34	-0.128*	1.71	**-0.131***	**1.85**
Age-squared	0.00360*	1.81	0.00334	1.63	0.00417**	2.03	**0.00429** **	**2.2**
Mother's years of schooling	-0.00562**	2.11	-0.00554**	2.09	-0.00461	1.66	**-0.00456**	**1.63**
Mother's age	0.000941	1.01	0.000892	0.96	0.00129	1.11	**0.00128**	**1.11**
Number of siblings	0.0114***	3.97	0.0114***	3.94	0.00970***	2.72	**0.00956*** **	**2.65**
Poverty incidence rate			0.00579**	2.23			**0.00395**	**1.34**
Infrastructure index			-0.0486	0.61			**-0.0523**	**0.68**
Number of students per classroom in high school			-0.00026	0.57			**-0.00096**	**1.62**
Other variables and school fixed effects	yes		yes		yes		**yes**	
Percentage of predicted value out of [0,1]	6.3%		8.0%		7.6%		**9.2%**	
Number of observations	2962		2962		2761		**2761**	

*** 1%, **5%, *10% significance. Absolute t values are shown next to parameter estimates.

Second, the interaction of TEEP cohort and TEEP province is insignificant (only marginal significant in Model 3), suggesting that TEEP's effect on male migration is small. The interaction of Female, TEEP cohort, and TEEP province is insignificant in Models 3 and 4, though marginally significant in Models 1 and 2. The partial effect of TEEP on females is significant in the models without cohort-specific controls (Models 1 and 3) but insignificant in the models with cohort-specific controls (Models 2 and 4). Therefore, TEEP's effect on females' mobility is ambiguous. Interestingly, the interaction of TEEP cohort, TEEP province, and Log (consumer durables) is negative and significant, suggesting that TEEP seemed to increase the migration propensity of students from relatively poor households possibly to search better schooling and work opportunities.

5.3.3 Labor market earnings

As explained earlier, we take a two-step method to estimate wage equation. The first-step probit regression results are reported in Table 5.10. The results suggest that the probability of wage being

observed is positively correlated with age (in a decreasing manner), being female in the TEEP cohort, and height (which captures physical endowment). It is intuitive that older and/or taller individuals were more likely to participate in labor force in this relatively young sample.. The positive coefficient of Female × TEEP cohort is consistent with that females are more likely to work while in school in the younger cohort (note that schooling is controlled).

The results also suggest that the probability of wage being observed is lower if the individual is more schooled, female with TEEP, has educated mother, and in wealthier families. Schooling and working are competing for our sample so it is reasonable that higher educated students were less likely to participate in labor force. This inclination is stronger in the TEEP cohort. TEEP reduced females' likelihood of labor force participation, which is consistent with that female TEEP beneficiaries spent more time studying. Also, students from relatively wealthier households are less likely to work while schooling.

Table 5.10 Probit regression results of selection equation of wage (dependent variable: wage being observed)

	Full Sample			Trimmed Sample		
	Coef.	t		Coef.	t	
Years of schooling	-0.065	2.63	***	-0.056	2.2	**
age	0.969	4.03	***	0.963	3.87	***
age squared	-0.023	3.52	***	-0.023	3.38	***
TEEP cohort*TEEP province	0.359	0.86		0.108	0.25	
Years of schooling*TEEP cohort*TEEP province	-0.033	0.71		-0.009	0.19	
TEEP cohort	0.291	0.64		0.566	1.2	
Years of schooling*TEEP cohort	-0.169	4.07	***	-0.197	4.61	***
Female*TEEP cohort*TEEP province	-0.245	1.81	*	-0.251	1.79	*
Female	-0.292	0.86		-0.237	0.68	
Female*Years of schooling	0.021	0.7		0.018	0.59	
Female*TEEP cohort	0.578	3.91	***	0.566	3.72	***
Mother's years of schooling	-0.044	4.9	***	-0.039	4.22	***
Mother's age	-0.006	1.69	*	-0.005	1.55	
Log(consumer durables)	-0.161	7.84	***	-0.175	8.23	***
Height	0.005	1.77	*	0.005	1.73	*
Number of students per classroom in high school	0.002	1.48		0.001	1.24	
_cons	-7.1	3.19	***	-6.988	3.02	***
Number of observations	2962			2761		
Pseudo R-squared	0.237			0.238		

*** 1%, **5%, *10% significance. Absolute t values are shown next to parameter estimates.

Table 5.11 Regression results of logarithm of latest monthly earning, for full sample and trimmed and weighted sample

	Full Sample				Trimmed and Weighted Sample			
	Model 1		Model 2		Model 3		Model 4	
Years of schooling	0.0864***	6.02	0.0873***	5.61	0.0894***	5.39	0.0832***	4.5
age	0.519***	3.19	0.496*	1.95	0.557***	3.66	0.757***	2.92
age squared	-0.0125***	2.97	-0.0119*	1.94	-0.0139***	3.55	-0.0185***	2.98
TEEP cohort*TEEP province	0.0424	0.15	0.0339	0.11	0.156	0.59	0.191	0.71
Years of schooling*TEEP cohort*TEEP province	-0.0204	0.62	-0.0196	0.58	-0.031	1	-0.0344	1.09
TEEP cohort	-0.00543	0.02	-0.0291	0.08	-0.23	0.85	0.0105	0.03
Years of schooling*TEEP cohort	-0.0463	1.52	-0.0408	0.74	-0.0313	1.09	-0.0811	1.37
Female*TEEP cohort*TEEP province	0.234**	2.13	0.240**	2.16	0.272**	2.26	0.224*	1.84
Female	-0.495***	3.2	-0.488***	2.94	-0.333*	1.83	-0.388**	1.99
Female*Years of schooling	0.0280*	1.9	0.0275*	1.75	0.0165	0.99	0.0209	1.18
Female*TEEP cohort	0.122	1.24	0.107	0.78	0.0646	0.58	0.189	1.21
Mother's years of schooling	0.00956	1.39	0.0105	1	0.0165**	2.02	0.00905	0.85
Mother's age	-0.00026	0.12	-0.00013	0.06	-0.00169	0.8	-0.00277	1.27
Log(consumer durables)	0.0455***	3.6	0.0488	1.6	0.0474***	3.74	0.0169	0.48
Height	0.00569*	1.92	0.00557*	1.79	0.00613*	1.78	0.00698*	1.97
Number of students per classroom in high school	-0.0006	0.68	-0.00062	0.7	0.000581	0.52	0.000799	0.69
If phone interview	0.233***	4.62	0.232***	4.6	0.242***	3.18	0.244***	3.24
Inverse Mills Ratio			-0.0445	0.12			0.376	0.92
School fixed effects	yes		yes		yes		yes	
Number of observations	1716		1716		1592		1592	

*** 1%, **5%, *10% significance. Absolute t values are shown next to parameter estimates.

In the second stage, we estimate a Mincerian equation that includes inverse Mills ratio from the first stage as a regressor. The results are reported in Table 5.11. Not surprisingly, years of schooling completed is significant at the 1% level and one more year of schooling contributes to about 8% increase in wage for males, on average. This estimate is consistent with the literature. The interaction terms, Years of schooling × TEEP cohort and Years of schooling × TEEP cohort × TEEP province, are both insignificant, indicating that return to schooling does not change much between cohorts or by TEEP.

The variable, TEEP cohort × TEEP province, is insignificant again, suggesting that TEEP did not have a direct effect on males' wage. However, it is interesting to find that the interaction term of Female, TEEP cohort, and TEEP province is positively significant. This suggests that TEEP reduced the salary gap between females and males if females had lower salary than male. In other words, TEEP reduced female disadvantage in the labor market. We note that this gender effect of TEEP occurred after controlling for years of schooling. Therefore, the overall TEEP effects on females' earnings relative to males can be much larger, once we take into account that TEEP significantly enhanced female advantage on schooling attainment.

The estimates for other variables are reasonable: age has a positive and decreasing effect on wage; and better physical endowment is also related to a higher salary. Phone interview is positively correlated with wages, which is reasonable since phone interview was used for out-migrants who reside in provinces away from their origins and the three metropolitan areas.

5.4 Summary

We examined long-term impacts of improved school quality at the elementary school stage on subsequent schooling investments and labor market outcomes using unique data from a recent survey that tracked students from the areas that experienced (and did not experience) a large scale school intervention in the Philippines. We find that improved school quality enhanced female advantage (or decreased female disadvantage) in subsequent schooling investments, migration and labor market earnings. That is, females study more (relative to males), and tend to migrate and earn more if they receive high-quality educational investments at the early stage.

6. Cost-benefit analysis: Aggregate Impacts of TEEP

The objective of this section is to estimate aggregate impacts of TEEP using individual-level parameter estimates. Estimation is based on parameters estimated from our tracking survey data and the Philippine Labor Force Survey (2009, October round).

We introduce following notations:

(i) Cohorts: e (measured by labor market experience)

(ii) Total number of students exposed (assume time invariant): $n_j, j = m, f \quad e = 0,1,2,\ldots,T$

(iii) Changes in years of schooling and (log) wages: Δs_j, $\Delta \alpha_j$

These are estimated from the long-term impact analysis (estimated in 2010/11 about 5 years after TEEP completion). As a potential problem, changes in both s and α are not completed among TEEP cohorts (Grade-6 in SY2004/05 and 2005/06: age 17/18 in 2010). Therefore, Δs_j and $\Delta \alpha_j$ are potentially underestimated.

(iv) Distribution of s: $f(s)$; pre-TEEP distribution: $F_{j,0}(s)$; post-TEEP distribution: $F_{j,1}(s)$; change (difference) in the distribution: $\Delta F(s) = F_{j,1}(s) - F_{j,0}(s)$

Average years of schooling from LFS 2009/Oct, age 20-29 (pre-TEEP cohorts) are 10.50877 for females, and 9632535 for males. In the simulation, we ignore S=16 group when deriving the average wage.

j	f	m
$f(0)$	1.15	1.16
$f(3)$	5.40	10.09
$f(6)$	5.90	9.57
$f(8)$	10.86	14.19
$f(10)$	29.79	28.09
$f(12)$	24.95	23.02
$f(14)$	21.87	13.81
$f(16)$	0.09	0.07

(v) Wage profile:$w_j(e,s)$

Note that $w_j(e,s)$ is non-linear (exponential)

We specify the expected wage:

$$w_j(e,s) = prob_j[s,e]\exp\left(\alpha_j + \beta_j s + r_j(e)\right)$$

$$\begin{aligned}\Delta w_j(e,s) &= prob_j\left[s+\Delta s_j, e\right]\exp\left(\left(\alpha_j+\Delta\alpha_j\right)+\beta_j\left(s+\Delta s_j\right)+r_j(e)\right)\\ &\quad -prob_j[s,e]\exp\left(\alpha_j+\beta_j s+r_j(e)\right)\end{aligned}$$

where experience effect: $r_j(e) = \gamma_{1j} * a(e) + \gamma_{2j} * a(e)^2$ and $a(e) = 20 + e$ (for simplicity, they are assumed to start working at age 20).

Note that Δs_j can have both level effect (returns to schooling) and growth effect (experience). However, LFS does not support complementarity of schooling and experience (in contrast to the results in Annex 3, using data from the early stage of labor market experience).

Next, we configure the above parameters. From Annex 2 (and Section 5), we have

$$\Delta\alpha_f = 0.224, \Delta\alpha_m = 0$$
$$\Delta s_f = 0.343, \Delta s_m = 0$$

Employment equation, estimated from LFS (linear probability model; age 20-49), supports the specification:

$$prob_j[s,e] = 0.0273 * s_f \quad \text{if } j = f$$
$$= -0.1808 + 0.0116 * s_m + 0.0308 * a(e) - 0.00046 * a(e)^2 \quad \text{if } j = m$$

Wage equation, estimated from LFS (age 20-49), offers the following parameters:

$$\ln\ w_j\ (e,s) = \ \alpha_j + \beta_j s + r_j(a(e))$$

Intercept: $\alpha_f = 5.758707 < \alpha_m = 6.385438$
Returns to schooling: $\beta_f = 0.1587313 > \beta_m = 0.0950345$
Returns to (proxy) labor market experience:

$$\gamma_{1f} = 0.048539,\ \gamma_{2f} = -0.0005535$$
$$\gamma_{1m} = 0.0622503,\ \gamma_{2m} = -0.0007078$$

We make some important assumptions for simulations. First, we assume that students work for T years after age 20. Therefore, they have 8 years after completing elementary school. Aggregate income change is derived as

Aggregate income (next $T + 8$ years)

$$\Delta Y(year = 0 + 8) = \sum_{j=f,m} n_j \sum_s f_j\ (s_j) \Delta w_j(0, s_j)$$
$$\Delta Y(year = 1 + 8) = \sum_{j=f,m} n_j \sum_s f_j\ (s_j) \left[\Delta w_j(0, s_j) + \Delta w_j(1, s_j)\right]$$

.....

$$\Delta Y(year = T + 8) = \sum_{j=f,m} n_j \sum_s f_j\ (s_j) \sum_{e=0}^{T} \Delta\, w_j(e, s_j)$$

Internal rate of return is derived as a solution for:

$$0 = -(TotalCost\ USD) + \sum_{e=0}^{T} \frac{\Delta Y(year=e+8)*(1/43)}{(1+r)^{e+8}}$$

where grade-6 students in SY 2005/06 (age 12) become age 20 after 8 years and then work for T years.

Finally, we use actual numbers of grade-6 enrollment in TEEP divisions in SY 2005/06. The total enrollment in SY 2005/06 was 1,729,421 in the 23 divisions. We simply divide this figure by 6 to get Grade-6 enrollment. Moreover, since our analysis shows that the impact was significant only for females, we assume that a half of the above figure is female enrollment. We use total costs of 221 Million USD for TEEP, 136.9 Million USD for SEDIP, and the exchange rate of 43 Peso/USD.

Table 6.1 demonstrates that the calculated internal rates of return are remarkably high. It depends, of course on the time horizon within which returns are captured. However, we conjecture that even 20 years of working is also a conservative assumption to derive total returns to TEEP investments.

Table 6.1 Internal rate of return

Programs	Horizon	IRR
TEEP	20 years	0.186
TEEP	10 years	0.159
TEEP + SEDIP	20 years	0.162
TEEP + SEDIP	10 years	0.119

Note: Horizon means working years in labor market after age 20. For simplicity, it is assumed that SEDIP was also introduced in the 23 TEEP divisions, not 15 SEDIP divisions.

In the above computation, we used the expected earnings by incorporating employment probability. A preliminary analysis showed that increased schooling generally raises the probability of being employed in the labor market (this is used in above the simulation). Similarly, we looked at the income change for the student who would hypothetically attain the average level of schooling before TEEP. Income variability due to the distribution of schooling attainment was not incorporated. That is, we simulated income changes that can occur to the average individual. Since only the aggregate income change (by cohort) matters to the computation of internal rate of return, such simplification is not consequential.

One can observe that because of the return structure, whether we aggregate TEEP and SEDIP does not significantly affect the internal rate of return. The most important factor is time horizon to capture the returns. This is largely because, as time goes, additional cohorts who were exposed to TEEP enter the labor market, which increases the total number of former students who experience income changes.

However, some of TEEP investments such as textbooks and school buildings may face depreciations over time, which probably requires replacement investments at some stages. Similarly, teachers' training needs continued efforts to refresh and renew training materials and methods. New teachers have to receive a larger amount of training too. These considerations necessitate us to reconsider the time horizon for capturing the returns, and the way to

incorporate additional (probably cyclical) investments and the costs. The assumption that the stock of education capital formed through TEEP investments in the period of 2000 to 2006 sustains for 20 years might make us overestimate the internal rate of return. However, as the above estimates indicate, variations of the total cost seem not to substantially influence the rate estimate.

7. Disseminations and Capacity Building

In September 2011, preliminary results were presented at JICA headquarter, the University of Philippines (UP), the Philippine Institute for Development Study (PIDS), the Philippine Department of Education (central office) and Cornell University. First, some useful detailed comments stimulated additional analyses and further refinement of the current study. Second, the Dep Ed seminar had 2 undersecretaries of education, 15 provincial superintendents, TEEP consultants, commentators from PIDS and UP, and representatives from JICA and World Bank. Key results of this evaluation study were well communicated to the stakeholders.

We also had a training workshop at Dep Ed central office on December 10, 2010. The objective of this workshop was to introduce impact evaluation methods Dep Ed statisticians. An example was prepared from TEEP. In this occasion, we also had attendees from World Bank, JICA Manila Office, Asian Development Bank, and the University of Philippines.

8. Conclusion

In the short-term impact analysis, we found significant impacts of TEEP on students' learning achievements. Our estimates show an increase of 12 to 15 score point with a 6-year exposure to TEEP. The long-term impact analysis demonstrates significant impacts biased in favor of females. Years of schooling completed significantly increased among females relative to makes. TEEP also significantly increased females' migration and labor market earnings. Though the positive impacts were directed toward females, our study also shows a remarkably high rate of return in TEEP since the program covered a larger population in 23 provinces and they are expected to work in labor markets for years.

Our study indicates that public investments in elementary education likely have positive longer-term impacts on schooling and labor market outcomes. If so, social returns to an early-stage investment can be greater than what the current study seems to show. This argument justifies large public investments to improve school quality at the early stage of public education, because the cumulative benefits are gradually realized at later stages in the education system and labor markets.

The competition between public and private schools is a unique feature of the Philippine education system due to the historical dominance of private institutions. If publicly subsidized and high-quality education is available, we also expect the inflow of good students into the public school system in the long run.

Our findings from the long-term impact analysis are also consistent with the observation that returns to schooling are significantly higher among females than males, especially after high school completion. Labor market imbalance, represented by significant gender gap in returns to schooling, seems to create a unique situation where the impact of improved school quality is biased in favor of females. This result is also supported by the behavior of parents, who were found to prioritize schooling investments in their daughters in the face of financial constraints. Our example demonstrates the importance of linking education and labor markets when predicting the impact of any intervention in schooling.

Acknowledgement

We are grateful to Yolanda Quijano, Irene Robles, Ishidra Afunggol, Vanessa Arness, Keitaro Aoyagi, Hiroaki Asaoka, Yasuharu Shimamura, Fe Gascon, Violy Cordova, Pamela Castanar, Leonarda Ebron, Aida Papag, Keijiro Otsuka, Judith Borja, Nanette Lee, Lorna Perez, Delia Carba, Abet Bas, Elizabeth King, Agnes Quisumbing, Juliet Abunyawan, Felisberta Sanchez, Ali Subandoro, Surajit Baruah, Maureen Bugatti, Joceyln Binumnya, the Bureau of Elementary Education, the Research and Statistics Division, and Antique, Cebu, Ifugao, Iloilo, Leyte, Negros Oriental, Nueva Vizcaya, Western Samar division offices of the Philippine Department of Education, and the Office of Population Study of the University of San Carlos for collaboration, support and guidance throughout this project.

IFPRI team

Futoshi Yamauchi*, Yanyan Liu*, Marites Tiongco, Shalini Roy, Surajit Baruah, Yisehac Yohannes, Ali Subandoro, Lynette Aspillera and Nelly Tioco (* in charge of impact evaluation analysis)

References [see Annex 1 and 2 for more details of the literature]

Ramos, F. V. 1995. *The Social Reform Agenda: Winning the Futu*re. Quezon City, Philippines: Republic of the Philippines.

Republic of the Philippines, Congress. 1998. *Republic Act 8425* (Social Reform and Poverty Alleviation Act).

Republic of the Philippines, Department of Finance. 2001. *Department Order No. 32-01* (effective November 20, 2001).