MATHEMATICA
Policy Research, Inc.

MEMORANDUM

P.O. Box 2393
Princeton, NJ 08543-2393
Telephone (609) 799-3535
Fax (609) 799-0005
www.mathematica-mpr.com

TO: Rebecca Tunstall and Orlando Martínez

FROM: Larissa Campuzano, Lorenzo Moreno, Randall Blair

DATE: 11/3/2009
ESVED-213

SUBJECT: Final Impact Evaluation Design for Technical Middle School Activity - **Revised**

This memorandum presents the final impact evaluation design for the technical middle schools strengthening activity of the Human Development project. The document builds on our previous memoranda ESVED-104 and ESVED-103, as well as discussions with you and other stakeholders in Fondo del Milenio (FOMILENIO), the Ministry of Education (MINED), and Consortium for International Development in Education (CIDE).

EXECUTIVE SUMMARY

The government of El Salvador, together with the Millennium Challenge Corporation (MCC), is implementing a program that seeks to address poverty and economic development in the Northern Zone of El Salvador. The program consists of two projects: (1) Human Development and (2) Productive Development. Under the Formal Technical Education Sub-Activity of the Human Development project, MCC is providing support to strengthen 20 selected general and technical middle schools in key municipalities in the Northern Zone (thereafter "strengthening activity"). This support includes improving the array of degree granting and non-degree granting vocational training and skills courses for youth; training teachers in the use of advanced instructional technologies; linking formal education with private sector needs; capital improvements (laboratories and workshops); and purchasing needed equipment. Over 9,000 students are expected to benefit from these activities, which will be implemented from 2009 to 2012.

MCC has contracted Mathematica Policy Research to design and conduct the impact evaluation of the middle school strengthening activity. The objective of the evaluation is to answer the following research question: What is the impact of strengthening 20 technical middle schools on students' educational and labor market outcomes?

Based on extensive consultations with MCC, FOMILENIO, and MINED, we chose as the final evaluation design for the strengthening activity a *matched comparison design*—a quasi-experimental design in which the 20 middle schools selected for the intervention were matched to 20 schools with similar demographic characteristics. To measure the educational outcomes of the students, Mathematica will use student-level data collected by MINED for the school years 2008 through 2012. To measure income, employment, and post-secondary educational outcomes not available in MINED's administrative data files, Mathematica will use survey data from the

Encuesta de Seguimiento de Estudiantes (ESE). The ESE is being implemented by CIDE with Mathematica's technical assistance. Baseline data collection will take place during October and November of 2009. Another round of data collection will take place in October and November of 2013. The proposed quasi-experimental design allows us to assess the impact of the intervention on the students who attended the schools in 2010 (first year of the intervention) through 2012 (third year of the intervention), regardless of the grade in which they were enrolled.[1]

A. DESCRIPTION OF THE INTERVENTION

Technical middle schools in El Salvador serve grades 10, 11, and 12. These middle schools could offer two types of degrees: general (for which the students need to complete grades 10 and 11); and technical (for which the students need to complete grades 10, 11, and 12).[2] The intervention will strengthen 20 technical middle schools in the Northern Zone. While the intervention was scheduled to begin in 2009, 2010 is considered the first full year of implementation since most of the activities undergone in 2009 pertain to planning the implementation. The strengthening activities will include: (1) improving the array of vocational training and skills courses for youth, (2) supporting capital improvements (laboratories and workshops), (3) purchasing needed equipment, (4) training teachers in the use of advanced instructional technologies, and (5) linking formal education with private sector needs. The intervention will benefit students from both the general and technical specializations offered by the middle school. These actions intend to improve enrollment, continuation, and graduation rates in participating middle schools. The final goal of the intervention is to improve the incomes and employment opportunities of youths in the Northern Zone.

The rest of this memorandum describes the evaluation design in detail, including the research questions, outcomes, data collection, and the methods we propose for evaluating the impact of the intervention.

---

[1] Mathematica's current contract for the evaluation ends in September 2010. For this reason, all analysis and deliverables scheduled for completion after this date are outside of Mathematica's current scope of work. Detailed calculations on the statistical power available to detect the impacts of the intervention on outcomes will be provided once we analyze the baseline administrative and survey data.

[2] Although the term "middle school" denotes in the U.S. a pre-secondary education, in El Salvador it denotes "secondary" education. Thus, throughout this document we refer to middle schools as interpreted in El Salvador.

B. KEY RESEARCH QUESTION

The objective of the middle school activity impact evaluation is to assess whether the intervention improves educational and labor market outcomes for the target population of students attending the 20 intervention schools. Specifically, the study is designed to answer the question:

- What is the impact of strengthening 20 technical middle schools on students' educational and labor market outcomes?

To measure the impact of the intervention on students who attended the 20 selected middle schools, we need to compare *what happened* to these students after the intervention was implemented with *what would have happened* to these students if their schools had not received the intervention. This last scenario, the counterfactual, cannot be observed. Therefore, our objective is to approximate it by finding a group of schools that were not selected for the intervention but were similar to the selected 20 middle schools before the intervention. The experience of this comparison group will serve to approximate what would have happened to the group of schools that received the intervention.

C. IMPACT EVALUATION DESIGN AND IMPLEMENTATION

The proposed evaluation design is a matched comparison group approach using propensity score methods.[3] The difference in outcomes between what we observed in the intervention group and what we observed in the selected comparison group represents our impact estimator. We used propensity score matching to identify a comparison group with observable characteristics similar to those of the intervention group before the intervention. The limitation of this method, as with any quasi-experimental method, is that we cannot guarantee that the intervention and the comparison groups are similar on unobserved characteristics at baseline.

In the remainder of this section, first we describe how the intervention group was selected by FOMILENIO. Then, we describe the selection of the comparison group.

[3] Propensity score methods are discussed in Rosenbaum and Rubin (1983, 1985), Dehejia and Wahba (1999, 2002), and Smith and Todd (2005).

1. Selection of the Intervention Group

MINED identified 75 middle schools in the Northern Zone that were eligible to receive the intervention. FOMILENIO contracted CIDE's services to develop the criteria on which 20 of the 75 technical middle schools would be selected for the intervention. Once FOMILENIO, MINED, and CIDE agreed on the final criteria, CIDE constructed a ranking score for each of the 75 eligible schools. A high ranking score reflects that a school demonstrated a high level of need according to the selection criteria, while a low score reflects that a school demonstrated a low level of need.[4]

An additional concern among stakeholders was to attain a wide geographic distribution of the intervention throughout the Northern Zone. Therefore, the procedure agreed upon by FOMILENIO, MINED, and CIDE was to select the two highest ranked schools in each of the 11 micro regions of the Northern Zone. Through this procedure, wide geographic distribution was attained and preference was given to the schools that had scored highest on the selection criteria in each micro region. Given that this procedure would have selected 22 schools, 2 micro regions had only one school selected for the intervention and 9 micro regions had 2 schools selected for the intervention. The list of selected schools is provided in Appendix D.

2. Characteristics of the Potential Comparison Group

The 55 schools that were eligible for the intervention but were not selected to receive it were candidates for our comparison group; we refer to them as the potential comparison group. Appendix A compares the characteristics of the intervention and potential comparison groups according to data from MINED's Censo Matricular 2006 and 2007, and primary data collected by CIDE to select the 20 intervention schools. The table presents means and standard deviations of the treatment and potential comparison groups, along with the standardized mean difference between the two groups.[5] We found that the mean characteristics of the intervention group were significantly different from those of the potential comparison group. Therefore, our objective was to use a propensity score matching to identify a comparison group of 20 schools among these 55 schools that had school-level characteristics similar to those of the intervention group.

---

[4] CIDE's deliverable, dated August 17, 2008, describes the selection criteria and the construction of the ranking score.

[5] The standardized mean difference is the difference of the sample means in the intervention and potential comparison groups as a percentage of the square root of the average of the sample variances in the intervention and potential comparison groups. The advantage of presenting the difference in means this way is that, unlike statistical tests of equivalence, this quantity is not affected by the sample size. Negative values reflect that the average of the comparison group is higher than the average of the intervention group.

a. Selection of the Comparison Group

Propensity score matching uses a propensity score (that is, the estimated probability of selection in the intervention) to assess the similarity among schools on the right-hand side variables included in the regression equation that estimates the probability of selection. Once the propensity score for each school has been estimated, there are several algorithms that can be used to select the comparison group. Given that the number of potential comparison schools is small (55 schools), we used the nearest-neighbor algorithm (without replacement) to select the comparison schools. This algorithm assigns each intervention school to a comparison school whose propensity score is closest to the propensity score of the intervention school (that is, the school that produces the smallest arithmetic difference in scores), and has not been selected previously. Once a comparison school is matched to an intervention school, it is taken out of the pool of potential comparison schools. Using this algorithm, we matched each intervention school to a unique comparison school, for a total of 40 schools (20 intervention schools and 20 comparison schools).

We estimated the propensity score with a probit model. The model's set of independent variables included variables that are correlated with the probability of selection into the intervention and, most importantly, variables that are most closely related to the outcomes we intend to measure (education and labor market outcomes). The variables come from the MINED's Censo Matricular 2006 and 2007, and from primary data collected by CIDE to select the intervention schools. The list of variables included in the models is reported in Appendix B, along with the estimated regression coefficients for the propensity score model. We also considered forming the comparison group by other methods; for example, in each micro region, we could have selected the school(s) in CIDE's ranking right below the schools that were selected for the intervention group. However, this group was not balanced because the schools in that comparison group were considerably different from schools in the intervention group. Therefore, we discarded that option.

b. Characteristics of the Matched Comparison Group

As shown by balancing tests in Appendix C, our selected comparison group is more similar to the intervention group in (observable school) characteristics than the potential comparison group was (Appendix A). In general, the mean characteristics of the intervention group are similar to those of the comparison group. We can see that the differences between the intervention and comparison groups tend to be small and not statistically significant for most variables. However, we found three statistically significant differences. Specifically, the scores that measure external and internal management capacity of the intervention and comparison

groups differ by a statistically significant margin (at the 5 percent level).[6] The ranking score calculated by CIDE also differs between the intervention and comparison groups by a statistically significant margin (at the 10 percent level), as well as the propensity score. As expected, the treatment group has a higher estimated probability of being selected than the comparison schools (also significant at the 5 percent level).

As noted above, CIDE gave high importance to geographic distribution across the 11 micro regions of the Northern Zone when selecting schools in which to intervene. Table 1 presents the geographic distribution of the 75 eligible schools, the 20 intervention schools, and the 20 schools selected as the comparison group. The comparison schools were selected in 8 of the 11 micro regions.[7] None of the potential comparison schools in the 3 regions without a comparison were similar to any of the intervention schools; therefore, they were not selected by the matching procedure.

TABLE 1

GEOGRAPHIC DISTRIBUTION OF TREATMENT AND COMPARISON SCHOOLS

Region	Micro Region	Eligible Schools	Intervention Schools	Comparison Schools
Cabañas				
	Ilobasco	4	1	1
	Sensuntepeque	6	1	2
	Subtotal	10	2	3
Metapán-La Palma				
	La Palma-San Ignacio-Citala	4	2	0
	Metapán	3	2	0
	Subtotal	7	4	0
Norte del Oriente				
	Gotera-Chapeltique	5	2	1
	Manantiales Del Norte	5	2	0
	Osicala-Perquín	8	2	5
	Santa Rosa De Lima	10	2	2
	Subtotal	28	8	8

[6] The variables are: subcriterio_capacidad_de_gestion and subcriterio_gestion_interna__600, created by CIDE.

[7] The three microregions that do not have comparison schools selected are: La Palma-San Ignacio-Citala; Metapán; and Manantiales Del Norte.

Region	Micro Region	Eligible Schools	Intervention Schools	Comparison Schools
Valle Alto del Lempa-Chalatenango				
	Alto Lempa Norte	7	2	2
	Alto Lempa Sur	9	2	2
	Chalatenango	14	2	5
	Subtotal	30	6	9
Total		**75**	**20**	**20**

Appendix D lists the names of the 75 eligible schools, the value of the ranking score calculated by CIDE, and whether the school was selected for the intervention or comparison group. Note that the procedure for selecting the comparison schools took into account this ranking, as well as other characteristics—including the number of students registered, retention rates, scores from the Prueba de Aptitudes para Egresados de Educacion Media (PAES)[8], and teacher characteristics. Therefore, our method did not select the 20 potential comparison schools with the highest ranking. However, 14 of the 20 comparison schools we selected would have also been selected by taking the 20 potential comparison schools with the highest ranking as the comparison schools. This process selected the 20 comparison schools that were as similar as possible to the 20 intervention schools. However, an important step is assessing the similarity of the intervention and comparison groups on the student-level indicators discussed below. When those data are available, we will conduct a baseline analysis in order to confirm the success of the matching procedure.

D. OUTCOME INDICATORS AND DATA SOURCES

The outcome indicators for the impact evaluation hinge on the availability of data from administrative and survey sources. Although the intervention will be delivered at the school level, the goal is to improve outcomes at the student level. Therefore, the impact evaluation will use student-level data to construct outcome indicators. Two types of outcome indicators are of interest to the stakeholders: (1) educational outcomes such as enrollment, progression and continuation in school, academic achievement, and middle school graduation, which could be collected from administrative records; and (2) labor market outcomes such as employment, income, or continuation in post-secondary education, which would need to be collected from

[8] The PAES is a mandated exam administered to every year to all 11th-grade students across El Salvador. Students must pass the exam to be granted their middle school diploma, or *bachillerato*. The PAES tests students on mathematics, language, social science and science competencies.

survey data. Next we discuss the outcome indicators and data sources for these two types of outcomes. Table 2 summarizes the outcome indicators and the data sources.

1. Outcomes from Administrative Data (MINED)

MINED collects data on the schools at the beginning of the school year through the Censo de Matrícula Inicial, and at the end of the year through the Censo de Matrícula Final. Given the high cost of collecting data, our impact evaluation design makes the best and most efficient use of the available administrative data.

As noted, first we need data on indicators before the intervention is implemented (baseline data). The middle school intervention was scheduled to begin in 2009. Although most activities conducted in 2009 pertained to planning the intervention, we cannot consider 2009 a baseline year because intervention activities had already started. Thus, the baseline data will consist of student-level data for students enrolled in the 40 selected schools in 2008 in all grades (10, 11, and 12). Data from previous years (2007 and 2006) can be used as additional baseline data in some cases. However, one limitation of the data collected by MINED prior to 2008 is that it lacks student identifiers, so it is infeasible to track students over time (see memo ESVED-104). As a result, the evaluation will use student-level data for 2008 as the baseline and, in some cases, will complement it with school-level information from the previous years.

Second, we need data on indicators after the intervention is fully implemented (post-intervention data). Because 2009 was mostly a planning year, we do not consider it the first full year of implementation. The post-intervention student-level data will come from the students attending the 40 selected schools in all grades (10, 11, and 12) in 2010 (first year of full implementation of the intervention), in 2011 (second year of the intervention), and in 2012 (third year of the intervention).

Outcome indicators that we will build from administrative data are:

1. **Enrollment:** Starting in 2008, student-level data are available. We can, therefore, construct a student-level variable of enrollment for each grade level. A student will be considered enrolled if he or she is registered in grades 10, 11, or 12.[9]

2. **Grade completion:** Starting in 2008, we will construct a student-level variable of grade completion for students who were enrolled in school at a given grade. We will consider

[9] For 2007 and 2006, the only available data are the number of students enrolled in each grade at the school level.

that a student completed grade 10, 11, or 12 if he or she was registered in that grade and completed it.[10]

3. **Continuation in school:** Starting in 2008, we will construct a student-level variable of continuation to grades 11 or 12 for students who were enrolled in grades 10 or 11 in the previous school year. A student registered in grade 10 or 11 in a certain year will be considered continuing in school if the student is registered for the next grade (11 or 12) in the next year.[11]

4. **Academic achievement:** Starting in 2008, student-level PAES test scores will be available.[12] We will construct student-level variables of students' language, mathematics, social science and science test scores for 11th grade students in 2008, as well as students in 11th grade during 2010, 2011, and 2012.

5. **Middle school graduation:** Data on graduation from middle school are likely to be available at the student level from MINED data starting in 2008.[13]

2. Outcomes from Survey Data

The administrative data collected by MINED do not include outcome indicators for post-secondary education and labor market outcomes. Thus, we will collect these data through the Encuesta de Seguimiento de Estudiantes (ESE). The specifications for this survey were provided to you in a separate document, ESVED-076.

Outcome indicators that we will collect from survey data are:

---

[10] Data on the number of students registered in each grade and the number of students who completed each grade at the school level are available for 2006 and 2007.

[11] To define this outcome indicator, we need data on registration for two consecutive years. For example, to define the continuation in school of 10th graders in 2008, we need data on the students registered in 10th grade in 2008, and data on students registered in 11th grade in 2009. Unlike the other indicators, it is not possible to construct an equivalent or similar variable at the school level since we cannot track the students for 2007 and before.

[12] Data on the average score of the PAES at the school level are available for 2007 and 2006.

[13] This has not been confirmed by MINED, but our preliminary conversations indicated that it was possible to include data for this outcome variable in the file they will provide to us.

1. **Middle school graduation:** Graduation from middle school for students who were enrolled in their last grade of middle school at the beginning of the school year.[14] An important consideration is that the last grade of middle school will be grade 11 for those students registered in the general specialization, and grade 12 for those students who were registered in the technical specialization.

2. **Employment:** Employment one year after students attended the last year of middle school (grade 11 for students in the general specialization, and grade 12 for students in the technical specialization).

3. **Income:** Income one year after students attended the last year of middle school (grade 11 for students in the general specialization, and grade 12 for students in the technical specialization).

4. **Post-secondary education:** Post-secondary education one year after students attended the last year of middle school (grade 11 for students in the general specialization, and grade 12 for students in the technical specialization).

CIDE will collect baseline data from students attending the last grade of middle school in 2008 through the baseline ESE.[15] These students will be interviewed in October and November 2009, one year after they attended the last middle school grade.[16] Through the post-intervention ESE, post-intervention data will be collected from students attending their last year of middle school in 2012. These students will be interviewed at the end of 2013, one year after attending

[14] MINED could provide data for this outcome. However, it is unclear whether the data will be available to us. Thus, we have included this outcome in the survey instrument if MINED cannot provide the data.

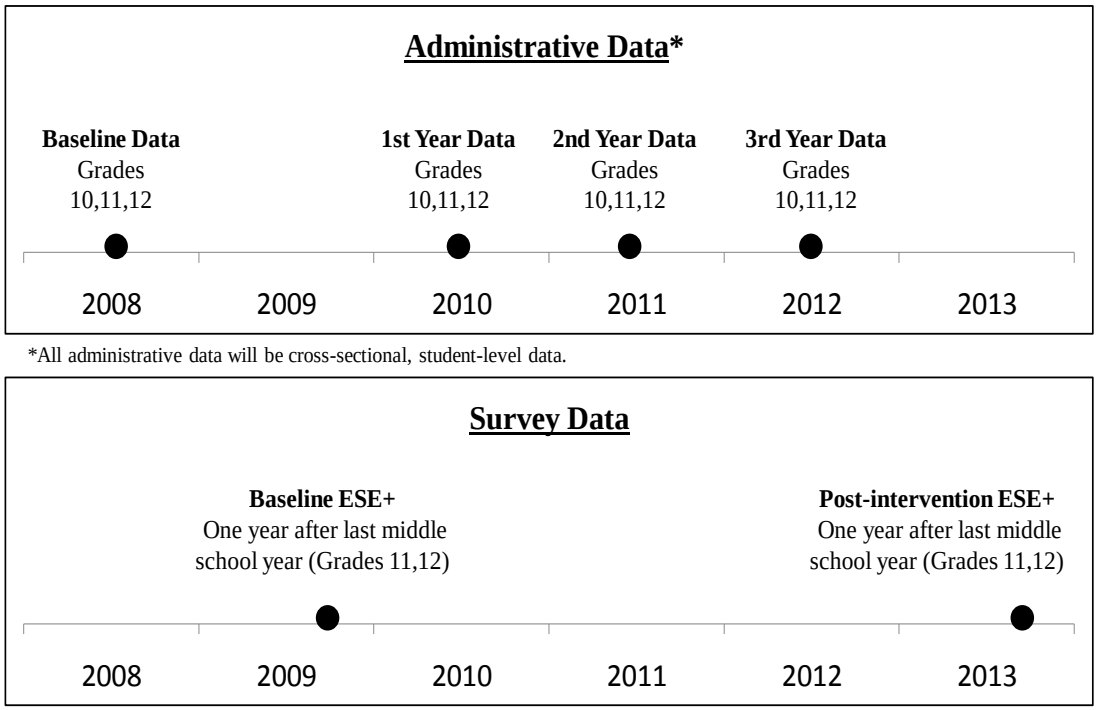
[15] Note that two types of students are included in the survey; those that are in the general track in 11th grade in 2008, which is their last year of middle school, and those that are in the technical track in 12th grade in 2008, which is their last year of middle school.

[16] Baseline administrative data from 2008 will be available for students taking the baseline ESE in 2009. These administrative data (such as test scores) will allow us to control for various student characteristics when estimating the impact of the intervention.

the last year of middle school.[17] This group will include students in the technical track registered in 12th grade in 2012, and students in the general track registered in 11th grade in 2012.[18]

FIGURE 1

DATA COLLECTION TIMELINE



[17] Administrative data from the third year of implementation (2012) will be available for students taking the post-intervention ESE in 2013. These administrative data (such as test scores) will allow us to control for various student characteristics when estimating the impact of the intervention.

[18] If the first year of intervention implementation (2010) falls short or is delayed, another survey can be conducted in 2014 on the students that will register in the last year of middle school in 2013. This will allow the evaluation to have outcome data for students that attended the technical middle schools in 2011, 2012, and 2013, which would allow the study to identify the effects of the intervention on students who completed all three years of technical middle school under the improved programs.

TABLE 2

DESCRIPTIONS AND DATA SOURCES OF OUTCOME INDICATORS

Outcome Indicator	Description	Data Source
Enrollment	Students' registration in grades 10, 11, or 12	MINED
Grade completion	Students' completion of grades 10, 11, or 12	MINED
Continuation in school	Students' registration in the subsequent grade (11 or 12) for students registered in grades 10 or 11	MINED
Academic achievement	Students' PAES test scores in grade 11	MINED
Middle school graduation	Students' graduation from middle school (grade 11 or 12)	MINED
Middle school graduation	Students' graduation from middle school (grade 11 or 12) for the survey sample	ESE
Employment	Students' employment one year after they attended middle school	ESE
Income	Students' income one year after they attended middle school	ESE
Post-secondary education	Students' post-secondary education one year after they attended middle school	ESE

E. IMPACT ESTIMATION

As we explained before, the matching procedure allowed us to select a comparison group of schools with baseline characteristics that are somewhat similar to those of the intervention group (see Appendix C). However, there are some characteristics that are significantly different in the intervention and comparison groups. Because the baseline student-level data are not yet available, we have not analyzed the similarity of baseline student outcomes between the two study groups. We plan to use 2008 student-level data to assess the equivalence of the intervention and comparison groups at baseline.[19] If this analysis shows that equivalence is not

[19] Note that to construct the continuation-in-school outcome for 2008, it will also be necessary to have the data for students registered in 2009.

attained in some outcome variables, we will control for any initial differences in the regression framework. An additional advantage of this framework is that the statistical precision of the impact estimates is improved by controlling for covariates such as students and school baseline characteristics in a regression model.

1. Statistical Analysis

The impact analysis will rely on a regression specification that compares students in schools in the intervention group to students in schools in the comparison group, controlling for idiosyncratic differences in the two groups. The main analysis will estimate the model presented below separately for each middle school grade (10, 11, and 12), and for each outcome indicator. The basic model can be expressed as follows:

$$(1)\ y_{is} = \alpha + \beta' x_{is} + \gamma z_s + \lambda T_s + \eta_s + \varepsilon_{is}$$

where y_{is} is the outcome of interest for student *i* in school *s*; x_{is} is a vector of baseline characteristics of student *i* in school *s*; z_s is the vector of baseline characteristics in school *s*; T_s is an indicator equal to one if school *s* is in the treatment group and zero if it is in the comparison group; η_s is a school-specific error term, a school "random effect"; and ε_{is} is a random error term for student *i* in school *s*. The vector of student baseline characteristics x_{is} will include time-invariant characteristics such as age or gender, and time-variant characteristics such as academic achievement. The vector of school baseline characteristics z_s will include time-variant school characteristics, such as number of students registered in the school in a certain year and number of teachers in a certain year, among others. Note that our main source of baseline information is the students who attended middle school in 2008, so we will construct school-level averages of these student-level outcomes and use them also as school baseline characteristics, such as the school average of student progression from grade 10 to grade 11.

The parameter estimate for λ is the estimated impact of the activity on the outcome of interest. The model presented in equation 1 takes into account the nested structure of the data; in this case, students are nested or clustered into schools. This type of model is referred to as a hierarchical linear model or mixed model and can be estimated with standard statistical packages.[20]

The model in equation (1) is designed to answer the general research question, "How do outcomes among students in treatment schools differ from outcomes among students in

[20] We will also conduct sensitivity analysis of alternative specifications of equation 1.

comparison schools, controlling for initial differences between treatment and comparison schools?" For outcomes from administrative data, student outcomes in treatment schools during 2010, 2011, and 2012 will be compared to student outcomes in comparison schools during 2010, 2011, and 2012, respectively, controlling for initial differences during 2008. For outcomes from survey data, student outcomes in treatment schools during 2013 will be compared to student outcomes in comparison schools during 2013, controlling for initial differences during 2009.

2. Statistical Power

An important consideration in any evaluation is to assess the size of the impact estimates that the evaluation will be likely to detect (statistical power). The sample size is critical in determining the size of the impact estimates that the evaluation will be likely to detect. Because this intervention is to be implemented at the school level, the number of schools participating in the study (20 treatment schools and 20 comparison schools) will limit the size of the intervention effect that the evaluation will be able to detect. The analysis proposed uses student-level data in order to increase the statistical power of the evaluation. However, because the students are clustered within schools and are affected by the same school environment, the students cannot be considered statistically independent and clustering at the school-level needs to be accounted for. The size of the student sample that will be used in the analysis depends on the data source; we will first discuss the sample from MINED's administrative data and then discuss the ESE survey data.

For educational outcome indicators constructed with MINED's administrative data, we anticipate that at each grade level data on approximately 50 students will be available for each school. Our preliminary estimates on statistical power, therefore, assume that data on 20 treatment schools, 20 comparison schools, and 50 students (per grade) in each school are available. Other factors that affect the statistical power of the study are the portion of the total variation of the outcome indicator of interest that lies between schools (intra-cluster correlation), and the size of the correlation between baseline data and follow-up data. We made conservative assumptions on these factors[21] and estimated that the evaluation will be able to detect differences in graduation rates between the intervention and comparison groups of about 13 percentage points.[22]

[21] We assume an intra-cluster correlation of 0.2, a student-level R^2 of 0.3, and a school-level R^2 of 0.5.

[22] We used an initial graduation rate of 73 percent to calculate the mean and standard deviation of this outcome indicator. This rate is based on graduation rates for technical middle school that appear in MCC's Monitoring and Evaluation plan at http://www.mcc.gov/mcc/bm.doc/el-salvador---me-plan.pdf

We drew the sample for the baseline data collection with the goal of obtaining the best school-level estimates of baseline income, employment, and graduation status for the 40 study schools (20 intervention schools and 20 comparison schools). These school-level estimates will be used as school-level covariates that adjust for school-differences and offer gains in precision in a multi-level model when estimating the intervention impacts, as the one described above. The target sample size for the baseline ESE survey was 600 students. To gain maximum precision for each school, we selected 15 students from each school to be included in the sample.[23] We selected the sample using the school as the stratum, while implicitly stratifying by degree type (general degree or technical degree), and gender. This ensures the general proportionality among these variables without making the cells for strata too small.

The sample size for the student survey follow-up, which will collect labor market outcomes in 2013, has not been defined yet. We plan to use baseline survey data to refine our estimate of the intra-cluster correlation of the income and employment outcomes. Once we have estimates for these parameters, we will calculate the sample size required to detect relevant impacts on the income and employment indicators.

F. REPORTING PLANS

In summer 2010, we will provide a short memorandum summarizing the baseline findings from MINED student-level data from 2008, and baseline outcomes of the student survey collected at the end of 2009. This date assumes MINED and CIDE will provide baseline data by January 2010, including 2008 data and complete registration information for 2009, which are necessary to construct the continuation-in-school indicator.

In summer 2011, the evaluator will submit an interim report summarizing the findings after one year of program implementation. Educational outcomes will be based on data from MINED administrative records, which correspond to student-level data from 2010. This date assumes MINED will provide the evaluator with data by May 2011, including 2010 data and complete registration data for 2011, which are necessary to construct the continuation-in-school indicator. The main focus of this document will be to estimate the impact of the first year of full implementation of the intervention on students' educational outcomes.

In summer 2012, the evaluator will provide a second interim report summarizing the findings after two years of program implementation. Educational outcomes will be based on data from MINED administrative records, which correspond to the student-level data from 2011. The

[23] The schools vary dramatically in size (from about 9 to 400). For the one school with 9 students, we selected the 9 students and gave one extra student to each of the 6 largest schools.

main focus of this report will be to quantify the impact of the first two years of full implementation on students' educational outcomes. This date assumes MINED will provide the evaluator with data by May 2012, including all 2011 data and complete registration information for 2012, which are necessary to construct the continuation-in-school indicator.

In summer 2014, the evaluator will provide a final evaluation report for the technical middle school activity summarizing the findings after three years of program implementation. The main focus of this report will be to quantify the impact of three years of full implementation on students' educational and labor market outcomes. Educational outcomes will be based on data from MINED administrative records, which correspond to student-level data from 2012. Labor market outcomes will come from the survey to be conducted in 2013. The submission date of this final report assumes that post-intervention ESE data is available by the end of 2013.

TABLE 3

SCHEDULE OF DELIVERABLES

Deliverable	Main Focus	Tentative Due Date
Baseline Analysis	Statistical comparison of treatment and comparison students at baseline based on administrative and survey data	Summer 2010
Interim Report 1	Impact estimation after one year of full implementation based on administrative data	Summer 2011
Interim Report 2	Impact estimation after two years of full implementation based on administrative data	Summer 2012
Final Report	Impact estimation after three years of full implementation based on administrative and survey data	Summer 2014

cc: S. Alfaro (FOMILENIO), V. Ruddy (MCC-EL Salvador), V. Crowder (MCC-DC), M. Induni, File

APPENDIX A

MEAN CHARACTERISTICS OF INTERVENTION AND POTENTIAL COMPARISON SCHOOLS

Variable[a]	Intervention	Potential Comparison	Standardized Mean Difference (% of a Standard Deviation)
Score: Access to Poor Populations	0.56	0.47***	80.9
	(0.08)	(0.14)	
Score: Basic Services	0.70	0.65	40.9
	(0.12)	(0.11)	
Score: Infrastructure and Space	0.41	0.33***	73.3
	(0.11)	(0.12)	
Score: Learning Environment	0.49	0.43	39.8
	(0.13)	(0.12)	
Score: Human Resources	0.41	0.33	34.2
	(0.24)	(0.22)	
Score: Growth Potential	0.33	0.26*	51.6
	(0.11)	(0.14)	
Score: External Management Capacity	0.43	0.31***	96.5
	(0.1)	(0.14)	
Score: Internal Management Capacity	0.49	0.38***	75.1
	(0.16)	(0.14)	
Score: Economic Development	0.50	0.49	8.0
	(0.15)	(0.13)	
Score: Psychological/Social Risk and Prevention Measures	0.39	0.34	22.3
	(0.24)	(0.24)	
Ranking Score (Weighted sum of scores)	41.23	34.79***	110.1
	(6.5)	(5.11)	
General Matriculation 2007	163.85	94.58*	38.8
	(238.81)	(81.66)	
Vocational Matriculation 2007	279.20	122.33***	62.1
	(303.66)	(187.88)	
PAES Mathematics Score 2006	5.13	5.15	-3.3
	(0.64)	(0.83)	
PAES Social Sciences Score 2006	5.46	5.42	6.0
	(0.62)	(0.61)	
PAES Language Score 2006	5.85	5.48*	43.3
	(1.01)	(0.64)	
PAES Sciences Score 2006	5.25	5.30	-7.9
	(0.58)	(0.59)	
PAES Global Score 2006	5.30	5.19	15.0
	(0.81)	(0.72)	
Number of Students Tested: PAES 2006	166.10	81.87**	59.0
	(177.23)	(96.57)	
PAES Mathematics Score 2007	5.57	5.14*	41.0
	(1.21)	(0.85)	

Appendix A *(continued)*

Variable[a]	Intervention	Potential Comparison	Standardized Mean Difference (% of a Standard Deviation)
PAES Social Sciences Score 2007	6.01	6.14	-23.0
	(0.39)	(0.71)	
PAES Language Score 2007	5.82	5.63	43.3
	(0.40)	(0.50)	
PAES Sciences Score 2007	5.78	5.72	8.5
	(0.86)	(0.65)	
PAES Global Score 2007	5.83	5.56	33.4
	(0.85)	(0.76)	
Number of Students Tested: PAES 2007	105.15	57.38**	49.9
	(122.12)	(58.52)	
Years of Teacher Experience: 2007	10.82	10.51	10.6
	(2.91)	(3.02)	
Total Number of Teachers: 2007	15.20	8.51***	69.8
	(12.43)	(5.41)	
2nd Year General Retention Rate 2006-2007 (Method A)	1.08	0.86*	37.2
	(0.80)	(0.26)	
2nd Year Vocational Retention Rate 2006-2007 (Method A)	0.69	0.67	11.6
	(0.18)	(0.18)	
3rd Year General Retention Rate 2006-2007 (Method A)	0.11	0.08	23.7
	(0.17)	(0.14)	
3rd Year Vocational Retention Rate 2006-2007 (Method A)	0.87	0.87	0.3
	(0.08)	(0.16)	
2nd Year General Retention Rate 2006-2007 (Method B)	0.92	0.80	19.1
	(0.83)	(0.34)	
2nd Year Vocational Retention Rate 2006-2007 (Method B)	0.55	0.46	26.7
	(0.32)	(0.34)	
3rd Year General Retention Rate 2006-2007 (Method B)	0.10	0.07	20.6
	(0.17)	(0.14)	
3rd Year Vocational Retention Rate 2006-2007 (Method B)	0.69	0.57	31.5
	(0.36)	(0.44)	
Propensity score	0.53	0.17***	152.6
	(0.29)	(0.17)	
Sample Size	**20**	**55**	

Source: Scores calculated by CIDE from data collected from the 75 eligible middle schools; matriculation data and PAES results, MINED; retention rates and propensity score calculated by Mathematica.

Note: Standard deviations in parentheses.

[a]Retention rates are calculated as the ratio of students matriculated in grade X+1 in 2007 divided by the students registered in grade X in 2006. This calculation assumes that no new students have matriculated in grade X+1. Under Method A, schools that did not have general or vocational registered students in 2006 were given a missing value. Under Method B, schools that did not have general or vocational registered students in 2006 were given zeros.

*Significant at the 0.10 level.
**Significant at the 0.05 level.
***Significant at the 0.01 level.

APPENDIX B

ESTIMATION OF THE PROPENSITY SCORE

Table B.1 presents the estimated coefficients from the propensity score regression model. The dependent variable was a binary indicator with a value of 1 if the school was selected for the intervention and 0 if the school was not selected. A probit specification was used to model the propensity score, which can be interpreted as a school's estimated probability of being selected for the intervention. We included three variables used by CIDE in selecting the intervention schools (*Score: Infrastructure and Space*, *Score: Access to Poor Populations*, and *Sum of Scores*). We also included variables from the Censo de Matrícula collected by MINED that are correlated with educational outcomes (*Years of Teacher Experience in 2007*, *Total Number of Teachers in 2007*, *Matriculation in the Middle School 2007*, *General Matriculation 2007*, *Vocational Matriculation 2007*, *PAES Global Score 2007*). Finally, we calculated retention rates from the second year of technical middle school programs by dividing the number of students registered in the second year in 2007 by the number of students registered in the first year in 2006. We did a similar calculation for general middle school programs (bachillerato general).

TABLE B.1

RESULTS FROM PROBIT MODEL USED TO ESTIMATE THE PROPENSITY SCORE

Variable[a]	Coefficient
Score: Infrastructure and Space	3.11*
	(1.84)
Score: Access to Poor Populations	2.89
	(2.06)
Years of Teacher Experience: 2007	-0.07
	(0.08)
Total Number of Teachers: 2007	0.16*
	(0.08)
Matriculation in the Middle School 2007	0.00
	(0.04)
PAES Global Score 2007	0.08
	(0.25)
General Matriculation 2007	0.00
	(0.04)
Vocational Matriculation 2007	0.00
	(0.03)
2nd Year Vocational Retention Rate 2006-2007 (Method B)	-0.21
	(0.79)
2nd Year General Retention Rate 2006-2007 (Method B)	0.19
	(0.32)
Sum of Scores	0.07
	(0.05)
Constant	-6.35***
	(2.17)

[a]Retention rates are calculated as the ratio of students matriculated in grade X+1 in 2007 divided by the students registered in grade X in 2006. Under Method A, schools that did not have general or vocational registered students in 2006 were given a missing value. Under Method B, schools that did not have general or vocational components were given zeros for these components.

*Significant at the 0.10 level.
**Significant at the 0.05 level.
***Significant at the 0.01 level.

APPENDIX C

MEAN CHARACTERISTICS OF INTERVENTION AND SELECTED COMPARISON SCHOOLS

Variable[a]	Intervention	Selected Comparison	Standardized Mean Difference (% of a Standard Deviation)
Score: Access to Poor Populations	0.56	0.57	-3.7
	(0.08)	(0.11)	
Score: Basic Services	0.70	0.65	46.4
	(0.12)	(0.11)	
Score: Infrastructure and Space	0.41	0.36	50.0
	(0.11)	(0.13)	
Score: Learning Environment	0.49	0.43	41.5
	(0.13)	(0.12)	
Score: Human Resources	0.41	0.40	6.0
	(0.24)	(0.20)	
Score: Growth Potential	0.33	0.32	7.1
	(0.11)	(0.15)	
Score: External Management Capacity	0.43	0.32**	88.4
	(0.10)	(0.16)	
Score: Internal Management Capacity	0.49	0.38**	74.5
	(0.16)	(0.14)	
Score: Economic Development	0.50	0.51	-3.2
	(0.15)	(0.12)	
Score: Psychological/Social Risk and Prevention Measures	0.39	0.38	4.8
	(0.24)	(0.27)	
Ranking Score (Weighted sum of scores)	41.23	37.96*	55.9
	(6.50)	(4.93)	
General Matriculation 2007	163.85	113.95	28.0
	(238.81)	(90.25)	
Vocational Matriculation 2007	279.20	199.60	31.5
	(303.66)	(265.14)	
PAES Mathematics Score 2006	5.13	5.21	10.8
	(0.64)	(0.89)	
PAES Social Sciences Score 2006	5.46	5.58	19.2
	(0.62)	(0.43)	
PAES Language Score 2006	5.85	5.55	35.4
	(1.01)	(0.67)	
PAES Sciences Score 2006	5.25	5.25	-0.3
	(0.58)	(0.55)	
PAES Global Score 2006	5.30	5.25	7.0
	(0.81)	(0.7)	

Appendix C *(continued)*

Variable[a]	Intervention	Selected Comparison	Standardized Mean Difference (% of a Standard Deviation)
Number of Students Tested: PAES 2006	166.10	123.65	29.7
	(177.23)	(135.44)	
PAES Mathematics Score 2007	5.57	5.20	35.6
	(1.21)	(0.84)	
PAES Social Sciences Score 2007	6.01	6.26	44.1
	(0.39)	(0.63)	
PAES Language Score 2007	5.82	5.71	24.1
	(0.40)	(0.48)	
PAES Sciences Score 2007	5.78	5.85	9.4
	(0.86)	(0.72)	
PAES Global Score 2007	5.83	5.69	17.7
	(0.85)	(0.68)	
Number of Students Tested: PAES 2007	105.15	76.95	29.5
	(122.12)	(83.73)	
Years of Teacher Experience: 2007	10.82	10.46	12.1
	(2.91)	(3.28)	
Total Number of Teachers: 2007	15.20	11.35	40.2
	(12.43)	(6.81)	
2nd Year General Retention Rate 2006-2007 (Method A)	1.08	0.81	46.1
	(0.80)	(0.21)	
2nd Year Vocational Retention Rate 2006-2007 (Method A)	0.69	0.65	19.6
	(0.18)	(0.14)	
3rd Year General Retention Rate 2006-2007 (Method A)	0.11	0.10	10.0
	(0.17)	(0.15)	
3rd Year Vocational Retention Rate 2006-2007 (Method A)	0.87	0.92	38.0
	(0.08)	(0.10)	
2nd Year General Retention Rate 2006-2007 (Method B)	0.92	0.76	24.0
	(0.83)	(0.27)	
2nd Year Vocational Retention Rate 2006-2007 (Method B)	0.55	0.56	-1.5
	(0.32)	(0.27)	
3rd Year General Retention Rate 2006-2007 (Method B)	0.10	0.09	6.1
	(0.17)	(0.15)	
3rd Year Vocational Retention Rate 2006-2007 (Method B)	0.69	0.78	-21.1
	(0.36)	(0.35)	
Propensity Score	0.53	0.34**	83.3
	(0.29)	(0.17)	
Sample Size	**20**	**20**	

Source: Scores calculated by CIDE from data collected from the 75 eligible middle schools; matriculation data and PAES results, MINED; retention rates and propensity score calculated by Mathematica.

Note: Standard deviations in parentheses.

[a]Retention rates are calculated as the ratio of students matriculated in grade X+1 in 2007 divided by the students registered in grade X in 2006. Under Method A, schools that did not have general or vocational registered students in 2006 were given a missing value. Under Method B, schools that did not have general or vocational registered students in 2006 were given zeros.

*Significant at the 0.10 level.
**Significant at the 0.05 level.
***Significant at the 0.01 level.

APPENDIX D

SCHOOLS' RANKING SCORES, INTERVENTION STATUS, AND COMPARISON STATUS

School		Micro Region	Ranking Score	Selected for Intervention	Selected for Comparison
1	Instituto Nacional De Nueva Concepción	Alto Lempa Norte	49.97	X	
2	Instituto Nacional 14 de Julio de 1875	Gotera-Chapeltique	49.04	X	
3	Instituto Nacional Benjamin Estrada Valiente	Metapán	48.96	X	
4	Instituto Nacional Dr. Francisco Martínez Suárez	Chalatenango	48.85	X	
5	Instituto Nacional De Sensuntepeque	Sensuntepeque	47.65		X
6	Instituto Nacional De Osicala	Osicala-Perquín	46.95	X	
7	Instituto Nacional De Anamoros	Santa Rosa De Lima	46.16	X	
8	Complejo Educativo General Manuel José Arce	Osicala-Perquín	45.82	X	
9	Instituto Nacional La Reina	Alto Lempa Norte	45.48	X	
10	Instituto Nacional Dulce Nombre De María	Chalatenango	45.40		X
11	Instituto Nacional General Juan Orlando Zepeda	Chalatenango	44.56	X	
12	Instituto Nacional De Chapeltique	Gotera-Chapeltique	44.18	X	
13	Instituto Nacional Caserío El Coyolito	Alto Lempa Norte	43.88		
14	Instituto Nacional Republica De Italia	Alto Lempa Norte	42.05		X
15	Instituto Nacional Profesor Francisco Ventura Zelaya	Santa Rosa De Lima	41.98		X
16	Instituto Nacional De Corinto	Osicala-Perquín	41.51		
17	Instituto Nacional De Azacualpa	Chalatenango	41.44		
18	Instituto Nacional Segundo Montes	Osicala-Perquín	41.19		X
19	Complejo Educativo Naciones Unidas	Osicala-Perquín	40.60		X
20	Instituto Nacional De Ilobasco	Ilobasco	40.23		X
21	Complejo Educativo Sotero Laínez	Sensuntepeque	39.93	X	
22	Instituto Nacional De Potonico	Chalatenango	39.44		X
23	Instituto Nacional San Antonio Los Ranchos	Chalatenango	39.38		X
24	Instituto Nacional De El Sauce	Santa Rosa De Lima	39.11	X	
25	Instituto Nacional El Paraíso	Alto Lempa Norte	39.10		X
26	Complejo Educativo Santiago De La Frontera	Metapán	38.86	X	
27	Instituto Nacional De Perquín	Osicala-Perquín	38.37		X
28	Complejo Educativo De Lislique	Santa Rosa De Lima	38.12		
29	Complejo Educativo Florinda De Juárez Alemán	Osicala-Perquín	37.61		X
30	Instituto Nacional De Concepción Quezaltepeque	Chalatenango	37.31		
31	Complejo Educativo Profesor Vidal Umanzor	Santa Rosa De Lima	36.95		
32	Instituto Nacional De Sesori	Manantiales Del Norte (Manorsam)	36.86	X	
33	Instituto Nacional De Nueva Esparta	Santa Rosa De Lima	36.80		X
34	Instituto Nacional De Nombre De Jesús	Chalatenango	36.73		X
35	Complejo Educativo De Bolívar	Santa Rosa De Lima	35.93		
36	Instituto Nacional De San Simón	Osicala-Perquín	35.84		X
37	Instituto Nacional De La Laguna	Chalatenango	35.74		X
38	Instituto Nacional San Miguel De Mercedes	Chalatenango	35.42		
39	Instituto Nacional De San Ignacio	La Palma-San Ignacio-Citala	35.38	X	
40	Complejo Educativo De Poloros	Santa Rosa De Lima	35.20		
41	Instituto Nacional De Aguilares	Alto Lempa Sur	35.15	X	
42	Complejo Educativo Jacinto Castellanos Palamo	Sensuntepeque	35.02		
43	Instituto Ciudad Barrios	Manantiales Del Norte (Manorsam)	34.39		

Appendix D *(continued)*

School		Micro Region	Ranking Score	Selected for Intervention	Selected for Comparison
44	Instituto Carolina	Manantiales Del Norte (Manorsam)	34.05	X	
45	Complejo Educativo Cantón El Tule	Alto Lempa Sur	33.98	X	
46	Instituto Nacional Doctor Salvador Antonio Navarrete	Sensuntepeque	33.75		
47	Instituto Nacional De Arcatao	Chalatenango	33.65		
48	Instituto Nacional Profesor José Luis López	Santa Rosa De Lima	33.62		
49	Instituto Nacional San Francisco Lempa	Chalatenango	33.56		
50	Instituto Nacional De Sociedad	Santa Rosa De Lima	33.53		
51	Instituto Nacional De El Paisnal	Alto Lempa Sur	33.30		
52	Instituto Nacional De Guatajiagua	Gotera-Chapeltique	33.14		
53	Instituto Nacional De San Pablo Tacachico	Alto Lempa Sur	32.81		
54	Instituto Nacional San José Las Flores	Chalatenango	32.72		
55	Instituto Nacional Del Divisadero	Gotera-Chapeltique	32.59		
56	Instituto Nacional Caserío Aguacayo	Alto Lempa Norte	32.38		
57	Complejo Educativo Eben-Ezer	Alto Lempa Norte	31.88		
58	Instituto Nacional De Suchitoto	Alto Lempa Sur	31.64		
59	Instituto Nacional De La Palma	La Palma-San Ignacio-Citala	31.58	X	
60	Instituto Nacional De San Isidro	Sensuntepeque	30.90		
61	Instituto Católico San Pablo Apóstol	Alto Lempa Sur	30.79		X
62	Complejo Educativo Ciudad De Toronto	Metapán	30.67		
63	Instituto Nacional De Victoria	Sensuntepeque	30.52		X
64	Instituto Nacional De Yamabal	Gotera-Chapeltique	29.93		X
65	Complejo Educativo Caserío Las Américas	Alto Lempa Sur	29.76		X
66	Instituto Nacional De Jutiapa	Ilobasco	29.69	X	
67	Complejo Educativo Caserío Los Almendros	Alto Lempa Sur	28.90		
68	Instituto Nacional Cantón Las Pilas	La Palma-San Ignacio-Citala	28.60		
69	Instituto Católico Wojtila	Ilobasco	28.58		
70	Complejo Educativo Doctor Guillermo Ungo	Alto Lempa Sur	28.08		
71	Instituto San Luis De La Reina	Manantiales Del Norte (Manorsam)	27.91		
72	Instituto Nacional De San Rafael	Chalatenango	27.72		
73	Instituto Nacional Cristóbal Iglesias De Tejutepeque	Ilobasco	27.67		
74	Instituto San Gerardo	Manantiales Del Norte (Manorsam)	25.74		
75	Instituto Nacional De Citala	La Palma-San Ignacio-Citala	25.65		

Source: FOMILENIO, email communication from Orlando Martínez (October 22, 2008). Ranking score calculated by CIDE.

REFERENCES

Dehejia, R., Wahba, S., 1999. Causal effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 94, 448, 1053-1062.

Dehejia, R., Wahba, S., 2002. Propensity Score Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics*, 84,1, 151–161.

Rosenbaum, P. R., Rubin, D. B., 1983. Reducing Bias in Observational Studies using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, 79, 387, 516-524.

Rosenbaum, P.R., Rubin, D. B., 1985. Constructing a Control Group using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician*, 39, 1, 33-38.

Smith, J.A., Todd, P.E., 2005. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? *Journal of Econometrics*, 125, 1-2, 305-353.