Rewriting the Algorithm: School Committees, Misreporting, and Machine Learning in a Social Program Targeting*

Arturo Aguilar
ITAM

Vicente Lopez
ITAM, Banco de México

June 23, 2025
WORK IN PROGRESS - PLEASE DO NOT CITE

Abstract

The selection of beneficiaries in public policy is a central issue in economics, as it involves the efficient allocation of scarce resources. This paper analyzes key targeting strategies using PROBEMS, a cash transfer program for high school students in Mexico. We examine three dimensions of beneficiary selection. First, we compare two allocation mechanisms: a centralized system based on a proxy-means test and a decentralized school-based ranking process. We find that school rankings prioritize academic merit over poverty, with a moderate risk of resource capture. Second, we examine how misreporting affects the centralized system. It not only distorts poverty indicators but also changes the weight assigned to specific assets in the proxy-means algorithm. Third, we apply machine learning methods to assess how targeting would differ if the program prioritized outcomes rather than need, as is typical with proxy-means tests. Relative to the centralized proxy-means test, we estimate that recipient assignments would differ by 40% under school-based rankings, 49% using corrected data, and 39% with outcome-based machine learning targeting. These results underscore the tradeoffs between accuracy, fairness, and manipulation risk when targeting social programs.

Keywords: Targeting, education, cash transfers, school committee.
JEL classification: I38, I22, I28, H75.

*Arturo Aguilar is deeply indebted with the people from the Ministry of Education for their support gathering data and following the intervention implementation. We thank Sergio Cárdenas for his comments to this draft. We would like to thank Pavel Alonso, Luis Enrique Madrid, Rocío Madrigal and Waldo Soberón. Maria Fernanda Pérez and César Sordia has given superb research assistance. We also thank seminar participants at ITAM, UCSD, CIDE and Sobre Mexico conference. Arturo Aguilar acknowledges support from the Asociación Mexicana de Cultura. All errors are our own. Corresponding author (Arturo Aguilar): arturo.aguilar@itam.mx.

1 Introduction

Adequately selecting recipients of a program or public policy is key to its successful implementation. Across a range of domains, programs rely on broad concepts to guide targeting: poverty-alleviation initiatives aim to reach low-income individuals; agricultural programs focus on potential productivity; loan providers assess repayment capacity; and disaster relief efforts prioritize vulnerable households. Targeting recipients based on such concepts is often challenging.

In the educational domain, programs, and policies that focus on students are often designed to prevent school dropouts or improve learning (see Ganimian and Murnane (2016) and Glewwe and Muralidharan (2016) for comprehensive reviews). In terms of targeting, scholarship and cash transfer programs tend to be needs or merit-based (Barrera-Osorio and Filmer, 2016) and some condition the delivery of transfers upon school attendance (Parker et al., 2008).

This paper contributes to the targeting literature by providing evidence of different components of the selection process. We employ the case of PROBEMS, a Federal stipend (cash transfer) program for public high school students in Mexico. We begin our analysis by stressing the importance of efficiently selecting recipients by showing that if the transfers are randomized among applicants, no statistically significant impacts on our main outcomes of interest (test scores, attendance, or dropouts) are found.

The first targeting feature that we study is the comparison of two selection mechanisms that have been widely discussed in the targeting literature: a centralized system that employs a proxy-means test (PMT) to select low-income students to a community (school)-based targeting (CBT) approach. The centralized system gathers the inputs for the PMT by requiring students to fill out an application form that asks for household demographic and socioeconomic characteristics (mainly asset holdings). The CBT process employs a committee ranking of applicants. Such committees exist throughout the country at the school level and are formed by school authorities, teachers, and parents. Our results show that school committees tradeoff wealth for merit since they select applicants that look richer but also have a better high-school entry GPA. Committees also select indigenous and students with disabilities in lower proportions. Also, we found a higher proportion of last name matches between the committee members and students who benefited from the CBT selection process. Finally, employing an RCT that determined the targeting mechanism to be used at the school level, we found that students who benefited from the CBT process obtained higher test scores at the end of the school year (0.18σ). No significant differences in dropouts were found.

The targeting literature, most extensively developed in the context of social protection programs, has rich evidence of the different strategies often implemented to select recipients. These strategies encompass proxy-means tests, local (community)-based rankings, self-selection, geographical targeting, and universal benefits (Alatas et al., 2016; Banerjee et al., 2019; Banerjeea et al., 2024; Hanna and Karlan, 2017). The papers that, similar to us, contrast PMT to CBT stress the tradeoff between more and better information that is available to local committees to the risk of resource capture. Basurto et al. (2017) show how the risks involved in this tradeoff can be counterbalanced with a more productive use of resources, which, if ultimately are shared, does not necessarily lead to a reduction in welfare. In the context of an agricultural and food distribution program, they find that deferring the recipient selection to local chiefs yields a grade of nepotism that is compensated with higher productivity. Alderman (2002) finds that involving communes in the delivery of social assistance improves the distribution of resources to the poor beyond what would be expected from a centralized proxy indicator created with the use of a questionnaire by accessing difficult-to-capture information. Alatas et al. (2012) compare the use of a community-based targeting approach to a centralized (and a hybrid). Their findings suggest that even though the community approach delivers worse selection results, this might be explained by the fact that the objective function (i.e. selecting the poor) could be conceived differently by the community. Further work contrasting local versus centralized (proxy-means) based has been implemented to distribute loans (Vera-Cossio, 2021), provide seed funds for microentrepreneurs (Hussam et al., 2022) and other poverty-alleviation programs (Stoeffler et al., 2016).

In the education literature, this paper also contributes to the work in local governance (or school-based management). Papers in this literature exemplify how the involvement of school committees influences (or not) educational outcomes through their participation in different aspects of school decision-making. Pradhan et al. (2014) study the impact of school committees whose role was to give recommendations on aspects like school expenditure and teacher qualifications by promoting community engagement. They find limited effect on educational outcomes except when a strong link to a village council was implemented. Similarly, Duflo et al. (2015) finds positive effects on student performance when school committees influence the contract renewal of teachers and participate on their evaluation. Khanna (2015) analyzes different levels of decentralization in the decision-making of the use of resources. He finds that decentralization benefits literacy levels given the decisions of school construction. In the present paper's context, Santibañez et al. (2014) evaluate a program (*Quality Schools Program*) that gives school councils decision-making power upon the use of public funds in the school (as opposed to centralized decisions). They find positive effects on learning through the use of a difference-in-difference estimation, which are likely driven by infrastructure improvements.

A second aspect that we analyze is the impact of information misreporting for the centralized targeting. In our context, households are familiar with PMTs to select recipients since the PROBEMS program was established in 2007, and the *Progresa* program (which also employs a similar targeting strategy) has been in place since 1997. In the case of PROBEMS, no verification of the self-reported information is performed. We match individual information for a subsample of the program applications to an independent survey collected a few months before, where respondents had no incentive to misreport. After substituting the information from the survey and simulating a new PMT with the same algorithm, we find that 50.1% of the program recipients would not have received it if the survey information had been employed. This difference results not only from differences in assets reported, but also for the importance that the algorithm gives to the different inputs. Using decomposition methods, we find that 75% of the difference between the original and the simulated PMT is due to levels of assets reported, while the remaining 25% results from the weight given to the different inputs.

This contributes to a developing misreporting literature. Niehaus et al. (2013) did not look directly at information misreporting but argued that if the number of variables used in the PMT increases, the verification cost would rise, making bribes more likely to appear. The paper we found to be most closely related was Banerjee et al. (2020b), which shows the distortionary effects of implementing the PMT in later survey responses about TV and SIM card ownership, without sales of either being really affected. Also, recent papers have looked into the use of mechanism design (Hussam et al., 2022) and modeling the motives to manipulate responses in a machine learning framework (Björkegren et al., 2020) to prevent information misreporting seeking to capture more resources.

A final element that we analyze is how to combine available information to select recipients. By leveraging from the initial RCT that was implemented, we use a causal forest to estimate the *conditional average treatment effect* and determine which recipients benefit the most from the program (Athey et al., 2019; Sverdrup et al., 2024). By using a simulation, we then compare how are these recipients different from those selected by PMT. Our results show a bigger contrast of recipients since out of the PMT recipients, only 39% would have also been selected by the machine learning algorithm. When the model is trained to reduce dropout probability, recipients tend to have higher wealth, a greater proportion are female, a lower proportion are pregnant, indigenous, and have disabilities. Also, ML-selected recipients tend to live closer to school in comparison to those selected by PMT.

This latest analysis contributes to a flourishing and fast developing literature. Haushofer et al. (2022) raised the question if the PMT strategies are really the best equipped to serve the purposes set forward by many programs. This paper applied to our context would suggest that if the PROBEMS program seeks to reduce school dropout, it is not necessarily the poorest students the ones that would take the most advantage of additional resources for such purpose. Beuermann et al. (2025) also discuss the difficulty of setting up an adequate objective for a targeted program by including the dynamic properties of a concept like poverty. Finally, the use of machine learning methods has gained attention in combination with satellite imagery (Jean et al., 2016; Smythe and Blumenstock, 2022) and with phone usage (Aiken et al., 2022) to better target program recipients (Athey and Wager, 2021).

The rest of the paper is organized as follows: section 2 describes the Mexican high school context, the program under analysis and details the experiment design; section 3.2 describes the data sources employed in the analysis; section 4 describes the differences between the profiles of applicants selected under either targeting assignment; section 5 follows into the impacts on educational outcomes; finally, section 6 concludes.

2 The scholarship program and the high-school context in Mexico

During the 2017-2018 school term, 5.2 million students were enrolled at the upper secondary level in Mexico (10th-12th grade), which represents 79% of the eligible population. Even though public schools (which account for 79% of students) are tuition-free, educational expenditures, which include school fees, supplies, and transportation, are non-negligible. The main challenges faced at this level are school dropouts and quality. Statistics indicate that, considering the previous 10 cohorts, 50% of the entry class will not graduate from high school, with dropouts being concentrated in the first year of high school (23%). The 2011 National Dropout Survey indicates that 63% of students mention economic restrictions as one of the reason for quitting school. The second and third most common reasons mentioned are "failing to pass a grade" and "marriage or pregnancy", mentioned by 42% and 27% of students, respectively.[1]

2.1 Scholarship programs

Since economic factors are one of the primary drivers of student attrition, Mexico has formulated and executed a series of policies aimed at mitigating dropout rates. Between 2012 and 2018, two Federal funded programs[2] stand out for their number of beneficiaries and the amount of financial support provided. Below, we describe them, underlining their differences in the targeting strategy employed on each.

[1]Students were asked to enumerate all the reasons that might have explained dropout. Thus, the percentages do not sum up to 100%.

[2]These programs compare to the well-known PROGRESA program, which established attendance to school as one of the conditions required to receive cash transfers. In contrast with the programs described here, PROGRESA was assigned at the household level. See Parker et al. (2008) for a detailed description and thorough literature review.

Centralized targeting. *PROBEMS* is a program that began in 2007 and was designed to provide economic support to students from poor households, conditional on their school enrollment. During the 2012-2018 period, *PROBEMS* was the largest educational program in terms of number of beneficiaries, summing between 350 thousand and 1.1 million of students per year (equivalent to 8%-23% of enrollment). To sign up, students had to apply through an online platform and fill out a form which, in addition to the applicants personal information, it inquired about demographic and socioeconomic conditions of the student's household.[3] The scholarships were publicized and known to be targeted towards poor individuals. No verification of the applicant's responses nor later audit is performed to beneficiaries. Only, upon submission of the application, the respondent declares to have truthfully reported the information and accepts liability if the information is proven to be misreported. Applications were collected twice during a school year, and an algorithm would calculate an income proxy with the households' self-reported demographic and asset holding information. This proxy is calculated using PCA (principal component analysis) and relates to a PMT (Alatas et al., 2012).[4]

This program provided monthly cash transfers between 34 and 42 USD depending on grade (more for higher grades) and gender (more for women).[5] Selected applicants receive their cash installments through local government offices and must collect their transfers individually at most 45 days after the bi-monthly payment.

Community (school)-based targeting. Established in 2013, the *Grants Against Dropout* program has been awarded to more than one million high-school students. To apply, students must submit at their schools: (1) a letter describing their motives to request the cash support, and (2) a survey with socioeconomic and demographic information, using a very similar form

[3]The questionnaire filled by applicants asks for the number of individuals in the household, latest total monthly income, parent's characteristics (e.g. age and schooling), household assets (e.g. refrigerator, washing machine, computer) and characteristics (e.g. water supply, electricity, dirt floor).

[4]Specifically, the PCA was constructed with: (a) locality-level characteristics, mainly a marginality index and a dummy indicating if the locality is part of a poverty program against hunger (*Cruzada contra el Hambre*); (ii) asset ownership, including TV, cable TV, washing machine, refrigerator, vehicle, boiler, gas stove, mobile phone; (iii) household characteristics, including dirt floor, sewage, electricity, internet, the ratio of rooms per inhabitant; and (iv) applicant characteristics, including being pregnant, having a disability and being indigenous.

[5]Between 650 and 800 Mexican pesos at the 2018 exchange rate.

as PROBEMS.[6] The applications are revised by a school committee that ranks students according to their evaluation of *dropout risk*. Committees are composed of seven members, which include the school principal, the vice-principal, three teachers and two parents. Rural schools hold smaller committees composed of three members: the school principal, a teacher, and a parent. The committee submits a package of applications to the Federal Ministry of Education. The package includes the information delivered by the students and the applicants' ranking established by the committee. The package is electronically filed and must include (i) names, positions, IDs, and signatures of the committee members and (ii) a ranked list of the applicants. The ranked list must include the names, address, email, telephone, grade, secondary GPA, risk of dropout (ranked between 1 and 5), and per capita household income of the applicants.[7]

To form their ranking, committees receive guidelines from the *Ministry of Education* about relevant dropout risk factors, such as socioeconomic, gender, and pregnancy status. However, the guidelines establish also that the committee has the ability not to follow those suggestions and to order students as they best consider. This gives committees the flexibility and ability to use private information and to determine on their own the relevant risk factors of students on their specific context (Alderman, 2002). The Ministry of Education then decides based on the available budget, the number of applications, and the number of grants awarded. The cash award and form of distribution is the same as *PROBEMS*.

In any of the previous programs, if the applicants or committee falsify information to get resources, the scholarship is suspended and any cash disbursement that has been paid is requested back.[8] In addition, Mexican laws establish that falsifying documents to appropriate public resources can be punished (if proven and prosecuted) with 4 to 8 years of prison. This would be applicable either to the applicant and/or committee.

[6]The survey completed by the students is widely used by the Ministry of Education and follows an almost identical format and question-wording as PROBEMS.

[7]The rules of the program indicate that committee members cannot nominate a direct or indirect relative to be a program recipient.

[8]The rules of the program indicate that the authority of the program (i.e. Ministry of Education) can request access to the beneficiary's household and documents to make a verification of the truthfulness of the information in their application (Mexican Diario Oficial de la Federación, 2017).

2.2 Conceptual framework

The scholarship programs previously described have a well-established goal: to prevent students from dropping out of school. This facilitates evaluating the program since administrative information is regularly collected to report dropouts as part of the national education indicators. Preventing dropouts is not a simple task. Assuming that not dropping out is a desirable goal (which might not be the case in the context of low educational quality or DTO capture of schools), there are heterogeneous reasons for leaving school and, as a result, different required interventions to prevent it. Monetary restrictions might be the main driver, albeit the cash incentive required to prevent dropouts might be heterogeneous (Filmer and Schady, 2011). In other cases, the lack of motivation might be the principal cause, and having a role model, teacher or peer support, or access to information about the benefits of schooling could be the best interventions (Nguyen, 2008; Avitabile and de Hoyos, 2018). Marriage, pregnancy, and unequal gender roles also might be important mechanisms.

Cash transfers and scholarships are among the predominating programs. Ultimately, monetary incentives might prove useful in preventing dropouts, even if they arise from different causes. However, to be effective, they must be successful identifying those students in the margin of dropping out for whom the monetary transfer would make a difference (de Janvry and Sadoulet, 2006; Haushofer et al., 2022). If sufficient, the cash could prevent students from working, give them the resources required to buy school supplies, pay for childcare, or even motivate them to study harder to avoid failing. Thus, the challenges faced by these programs include (i) gathering relevant and reliable information, (ii) being able to successfully use such information to identify students in the margin of dropping out, and (iii) determining the cash transfer amount required to prevent such dropout. This paper focuses on the two first aspects since the amount of resources in the scholarship programs is fixed.

The CBT targeting strategy might improve on the information component since school committees might have access to more detailed and hard-to-access information than what the centralized scheme achieves through the forms and questionnaires employed (Alderman, 2002). Having the ability to freely rank students following any procedure or judgment that they see fit, they might also have an advantage on a more efficient use of information(Alatas et al., 2012). Committees are able to combine the information in a more complex way than what a fixed algorithm would do. The flexibility enjoyed by the school committee enables them to target students based on more than one dimension (poverty, motivation, pregnancy, etc.). Also, since the committee is aware of the transfer amount, they might incorporate this in discriminating students for whom the transfer amount is not sufficient (Hanna and Karlan, 2017). The downside faced by the committee is that it might be more vulnerable to elite capture (Basurto et al., 2017) and, given that gathering information is costly, the available information might not be evenly distributed among students, and its quality might be related to students' characteristics, committee composition and effort. For example, an applicant with higher social skills might be more likely to be well-known to the committee.

In contrast, the centralized targeting strategy might enjoy a higher credibility. If the rules are clear, this strategy is more transparent and can be audited, making it less vulnerable to manipulation and corruption (Banerjee et al., 2020a). A disadvantage is that it collects information from a fixed questionnaire, which is vulnerable to misreporting (Niehaus et al., 2013; Hussam et al., 2022), particularly if the respondents have an incentive to lie and know that their likelihood of being punished is low. Also, this system follows an algorithm or set of rules which are easier to assess.

3 Empirical Strategy

In this section, we describe the setup for the empirical analysis and describe the main data sources employed.

3.1 RCTs and simulations

There is a wide literature on the effect of scholarships and cash transfers on educational outcomes with wide ranging effects (Glewwe and Muralidharan, 2016; Ganimian and Murnane, 2016). Even in the Mexican context, there is mixed evidence regarding the impacts that result from cash transfers on schooling outcomes (Behrman et al., 2005; Dustan, 2020). Even previous work from Peña (2013) and de Hoyos et al. (2024) found non-significant effects from the PROBEMS program. The first paper employed a regression discontinuity design, while the second randomized ties in the PROBEMS algorithm. These strategies, despite being very clean from the internal validity perspective, lacked some generalization power.

Simple RCT. To tackle the question of the potential impacts of the PROBEMS program we started by reserving a sample of $50,001$ applicants and implementing a simple randomized experiment, where treatment would receive the cash transfer and control would not. *Table A.1* shows an adequate balance for the final sample.[9] The results from this simple analysis, shown in *Table 1*, are consistent with de Hoyos et al. (2024). No significant effects are found for any of the measured outcomes, with point estimates showing the desired impact on dropouts (-0.3 pp), but not on test scores (-0.01σ).[10]

Targeting RCT. Following the literature suggestions to strengthen the targeting strategy, the next step consisted of implementing a larger experiment where the two possible targeting strategies were contrasted: (i) community (school)-based targeting (CBT) and (ii) proxy-means (or centralized) targeting (PMT). Random assignment was done at the school level: 497 schools were assigned to CBT and 502 to PMT.[11] A total of $61,512$ applications was received, which means that, on average, each school submitted 61 applications. *Table 2* shows that the randomization achieved an adequate balance. The randomization took place after the applications were received, which means that school committees were not aware ex-ante of the experiment. Therefore, the committee ranking and the PMT ranking is available for both treatment groups. The number of scholarships to be awarded per school (hereon n_s) was determined using the PMT algorithm.[12] The proportion of applicants in a school that received awards was, on average, 55% and ranged between 1% and 100%.

[9] *Table A.2* proves that the attrition was not random, but also not different in many dimensions.

[10] *Figure A.1* shows the overall distribution of test scores, which does not seem to be affected either by the program.

[11] *Table A.3* also shows the attrition in this experiment, which also displays differences in few covariates.

[12] The PMT was calculated using the socio-economic and demographic information submitted by the applicants. The Ministry of Education defined that, given their budget, they would be able to award a total of $24,691$ scholarships. The algorithm simulated the assignment of those scholarships to those applicants with the lowest PCA scores.

If a school was randomly selected to be in the PMT treatment, the highest n_s-th ranked applicants, according to the PCA score, were awarded a scholarship. If a school was randomly assigned to the CBT treatment, the highest n_s-th applicants following the committee ranking receive the scholarship. *Figure 1* illustrates the assignment process using a school with four applicants and two scholarships assigned to the school ($n_s = 2$). Applicants are numbered according to their PMT ranking; therefore, if the school is randomly assigned to the PMT treatment, applicants 1 and 2 would receive the scholarship. Meanwhile, if the school is selected to the CBT group, applicants 1 and 3, being the highest two according to the committee rank, would receive the scholarship. Following this illustration, we define four types of individuals akin to the *potential treatment status* definition employed in the LATE literature (Angrist and Pischke, 2009): (i) *always receivers*, individuals that are awarded the scholarship regardless of the targeting method used (such as individual 1 in *Figure 1*); (ii) *never receivers*, whom never receive the scholarship regardless of the method (such as individual 4); and (iii) *PMT (CBT) benefactors*, applicants that would receive the scholarship only if the PMT (committee) targeting is employed. In our targeting experiment, 24.4% of applicants are always receivers, 42.6% are never receivers, and 16.5% of applicants are PMT and CBT benefactors each.[13]

The previous numbers imply that, out of the pool of students receiving the scholarship, 40% are sensitive to the targeting method employed. As we describe in the previous subsection, such differences in assignment might result from different mechanisms. In addition to the targeting process, we will provide evidence to assess the importance of two other critical mechanisms.

Misreporting. The centralized targeting strategy heavily depends on the reliability of the responses given to the questionnaire from which the PMT is calculated. Given that applicants know that the scholarship is needs-based, they have the incentive to misreport the questions that they suspect are most likely to be used to estimate their poverty level (Martinelli and Parker, 2009). If verification is perceived to be unlikely, the incentive to under-report assets increases (Niehaus et al., 2013). Households might learn or create a belief about verification likelihood based on past applications (Camacho and Conover, 2011). Interestingly, only 1% of households

[13] Given that PMT and CBT benefactors perfectly substitute each other in the assignment process, their proportion in the sample is, by definition, the same.

report having no assets in the applications.

In this paper, we are able to match applicant's responses to a previous independent survey in which applicants respond to questions about some of the same assets reported in the scholarship application. This survey is applied to students whenever they are applying to placement in public high schools in Mexico City. Applicants are well aware that their responses to the survey do not affect their high school assignment and are only requested for statistical purposes. Applicants know that their assignment is only a function of a knowledge test and the school preference reported (Bobba and Frisancho, 2022).

Using the matched subsample, we will simulate the PMT assignment after substituting the asset information available in both data sources. Given that this would result from a simulation, we will be able to characterize applicants in a similar way as in the targeting experiment: (i) *always receivers*, (ii) *never receivers*, (iii) *PMT (honest) benefactor*.[14]

Algorithm efficiency. Using the asset information, the PMT strategy in this program employs a *principal component analysis* (PCA) to create a poverty index. The PCA's objective is to find the weights of the different assets, taking into account the correlation between assets, to maximize the variation of the resulting index. The result is a linear combination of the assets, where having more assets is assumed to be positively related to wealth. Being a needs-based scholarship, the benefit is then assigned to applicants with the highest PCA scores up to the total number of scholarships to be assigned. Alternative procedures used in the literature consider a linear regression, quantile regressions, and recently, machine learning methods (Hanna and Karlan, 2017; Brown et al., 2018). Also, variations in the objective variable to estimate have been explored (Haushofer et al., 2022; Beuermann et al., 2025).

An advantage of our setting is that the objective is easily measured and available through administrative information: school dropouts. We also explore the difference resulting from

[14]Our term *honest* is an oversimplification of an applicant that once the survey information for assets is employed, she would receive the program, but if the original PMT information was used she would not receive it. Honest benefactors are actually displaced by PMT benefactors in the distribution of the program. Therefore, we term them *honest* benefactors since they would receive the program if the algorithm employed the survey information, where respondents have a lower incentive to misreport.

using test scores as our objective.[15]

Nonetheless, the main objective of this section is to employ machine-learning methods to explore a more efficient algorithm to assign recipients (Haushofer et al., 2022). We are interested to learn how much the algorithm could be improved and to what extent the scholarship assignment would change. For that purpose, we follow Athey et al. (2019) and estimate conditional average treatment effects using the simple RCT. After we train the model, we employ the information from the targeting RCT to simulate how different the scholarship assignments would have been if we had followed the machine learning assignment procedure (hereon, *ML assignment*). After the simulation, a similar classification of applicants is defined: (i) *always receivers*, (ii) *never receivers*, and (iii) *PMT (ML) benefactors*.

3.2 Data sources

The main data sources used in this paper include information from the applications and administrative data gathered directly from schools. They have been provided by the Ministry of Education and Mexico City's public high school assignment system (known as COMIPEMS).

Scholarship applications. In total, $72,319$ applications from $1,410$ schools were received. Out of those, $1,001$ schools were selected for the experiment favoring the likelihood of accessing administrative data to collect the outcomes. *Table 2* gives descriptive information coming from the applications. The applications include household socioeconomic information, such as asset holding, services, dwelling characteristics, and applicant's characteristics. Also, all applications include the committee ranking for each applicant and committee members information: names and positions (e.g. teacher, parent, principal) of the committee members.[16]

Education outcomes. Administrative information was collected from schools, including: (i) student dropout (during the school year and across grades), (ii) grades[17] (including general

[15] Being a proxy for learning, this objective might be more debatable.

[16] In the committee application, students can also include a letter of motives for the scholarship. Interestingly, a small proportion of applicants in my sample are reported to send such letter (only 17%). The letters were not made available.

[17] It is important to note that these are not standardized test scores. Each school grades its own students as it sees fit, and one should expect heterogeneity in grades to reflect school quality. The grading scale in Mexico is

GPA, math GPA, and dummies for failing courses), (iii) indiscipline (measured through conduct reports), and (iv) absenteeism. Given the complexity of the high school system administration, collecting the information was challenging, and 37.5% of the schools (which amounts to 38.7% of applicants) did not respond.[18] *Table 3* shows the descriptive statistics for our main outcome variables in both RCTs. Interestingly, dropout rates are lower compared to national statistics.

COMIPEMS application data. The scholarship applicants that attend first grade in a high school located in Mexico City went through the assignment process that is followed in Mexico City's public high schools. This process is administered by the Metropolitan Commission of Public High School Institutions (COMIPEMS for its initials in Spanish). High school assignment follows a *serial dictatorship* mechanism based on self-reported school preferences and a multi-topic standardized test. As part of the process, applicants are required to fill a survey with socioeconomic and demographic information, which is employed exclusively for statistical purposes.[19] These questions do not affect the high-school assignment process and match with some of the items that applicants report in the scholarship application and that are employed in the PMT calculation (6 out of 17 items that are employed in the PMT calculation are available in both datasets).[20] 7.13% of our sample corresponds to Mexico City first-grade applicants, and we were able to match 95.2% of them with the COMIPEMS data using personal identifiers. Our matched sample shows that 88.3% of respondents report different asset holdings on both surveys. In 85.3% of the cases in which applicants report differently, they indicate having more assets in the COMIPEMS survey. Such direction is consistent with the fact that applicants have the incentive to misreport in the scholarship application and not in the COMIPEMS survey.

between 5 and 10, with 6 generally being the minimum passing grade.

[18] *Tables A.2* and *A.3* in the appendix give evidence of attrition. There is selective attrition reflected in a few variables and is lower in the targeting experiment. Such attrition was not different across treatment groups, as shown in the balance tables.

[19] Two differences between the COMIPEMS survey and the scholarship application questions are: (i) COMIPEMS survey occurs six months in advance and (ii) the questions use slightly different wording, however, both inquire about the same asset holdings.

[20] The PMT has 19 inputs, but we are not counting two inputs that correspond to locality-level information and are not self-reported by the applicants.

4 Effects on beneficiaries' profile

We start our analysis by looking at the differences in beneficiaries' selection that result from using either the centralized or local targeting strategies, employing data less susceptible to misreporting, and using a different (ML-based) algorithm to identify potential beneficiaries for the program.

4.1 Targeting strategy

Figure 1 shows that the targeting method selected is quite relevant for the selection of beneficiaries. The proportion of applicants whose award assignment depends on the targeting strategy (CBT or PMT) amounts to 32% of total applicants (or 40% among the awarded applicants).

Figure 2(a) shows the distribution of the PCA index for the four different types of applicants (according to the definitions established in section 3.1). By construction, there is no overlap between the distribution of applicants receiving and those not receiving the scholarship under the centralized targeting strategy since the assignment is a function of the PCA score. Nonetheless, two interesting facts arise. First, we expected the distribution of *always* and *never receivers* to be more concentrated to the extremes since those applicants would be awarded a scholarship regardless of the targeting strategy. Second, and probably more striking, the distribution of (self-reported) income per household member (*Figure 2(b)*) does not have a clear pattern across the types of applicant. It is true that the *never (always) receivers* have a distribution slightly more concentrated to the right (left), however, the *PMT* and *CBT benefactors* display very similar distributions. After exploring the data, we find correlations in the expected direction, although small in absolute terms (between 0.039 and 0.229).

A similar comparison can be made with the committee ranking. *Figure 2(c)* shows the histogram of the committee ranking values, where (1) corresponds to the rank of the applicant with the highest order of preference (i.e. the individual with the highest dropout risk). It is expected again that the *always receiver* and the *committee benefactor* will have a distribution concentrated to the left. However, it is surprising the considerable overlap between the *PMT benefactor* and the *never receiver*.

Table 4 makes a comparison between the types of applicants using a balance table format. Here, the last two columns display the F-stat and p-value for the joint test of equality of means. The asterisks displayed in the CBT benefactor column result from a test comparing the means between the *CBT* and *PMT benefactors*, which are the group whose scholarship assignment is sensitive to the targeting strategy employed.

Overall, the evidence from the table suggests that the targeting strategy employed seems to reflect a needs versus merit scholarship design, where the *committee targeting* is sacrificing poverty in favor of recipients with higher GPAs. The proportion of assets and services available to *CBT benefactors* in their household is, on average, 23 and 11 percentage points higher than that of *PMT benefactors*. Also, *CBT benefactors* are located in richer and more urban localities. As for the applicants' characteristics, *CBT benefactors* are younger (0.14 years), a lower proportion have disabilities (1.1 percentage points), and a lower proportion self-identifies as indigenous (1.9 percentage points). As for GPA, *CBT benefactors* have a GPA that is 0.05σ higher than that of *PMT benefactors*. *Figure 2(d)* shows also that the *CBT benefactors* first order stochastically dominate *PMT benefactors* in the secondary GPA distribution.[21] Travel distance from children's home to school is not statistically different from zero, but the distance between home and the paypoint location is statistically closer for *CBT benefactors*, suggesting that committees might also be selecting based on their accessibility to payments. Interestingly, pregnancy status of female students does not display a significant difference, despite the fact that this is one of the main reasons to dropout from school for female students.

A concern related to committee targeting relates to the risk of the private capture of resources (Basurto et al., 2017; Vera-Cossio, 2021; Alatas et al., 2019). Even though we have no direct evidence of corruption, we employed the proportion of applicants' last names that match at least one of the committee members. *CBT benefactors* match in a higher proportion (1.6 percentage points). Finally, it was surprising for us that the dropout risk classification (a metric self-reported by committees between 1-5) does not display statistically significant differences.

[21] Further detail about test score distribution can be seen in *Figure A.2*. Nonetheless, committees do not seem to reward effort in the students' applications since the proportion of students submitting a letter of interest with the application is not significantly different between both groups.

Unsurprisingly, *Table 4* also shows that always (never) receivers are the richest (poorest) in almost every single asset holding and household characteristic variable. Almost every single F-test rejects equality between groups, except only for the committee dropout risk.

4.2 Misreporting

As described in *section 3.2*, for a subsample of applicants ($n = 4,175$, 6.8% of the full sample), we are able to match the information with high school application data, where students do not have the incentive to misreport. At first glance, we see that for the six assets that we observe in both datasets, the number of assets reported in the high school application is higher than in the scholarship application (*Table A.4*). The difference ranges from a 48 percentage point difference in washing machine to a 10 percentage point difference in refrigerator. In all cases, the mean difference is statistically significant, and in every case, the direction is consistent with the incentives the scholarship application provides to under-report.

To assess the impact of misreporting in scholarship assignment, we start by substituting the asset information in the scholarship applications with that of the high school data for each applicant (COMIPEMS data). Then, we simulate the *PMT targeting* methodology by recalculating the PCA employed to rank applicants and distributing the same number of scholarships ($n = 1,499$ for this subsample). *Table 5* shows the result of this exercise with a confusion matrix format.[22] Misreporting severely affects the scholarship distribution: 50.1% of the original PMT recipients are sensitive to the source of information.

Substituting the asset information has two impacts on the *PMT targeting*: (i) those individuals that tend to under-report their holding of assets will increase their wealth status proportionally to the asset's weight in the PCA index (determined by the asset's coefficient); (ii) the assets' weights for the PCA index will change increasing the relative importance of those assets that have a higher variance after the information substitution. Following the *Oaxaca-Blinder* decomposition methodology (Fortin et al., 2011), we characterize the difference between the

[22]The observations in the diagonal correspond to those applicants whose outcome is the same regardless of the information used. The number of observations out of the diagonal match by construction since, by keeping the number of scholarships constant, if an applicant is no longer eligible once the information was substituted, she must have been substituted by another applicant.

original PCA score (PCA_s) and the simulated score (PCA_h) as *endowment* (Δ_X) and *structural* (Δ_β).

$$
\begin{aligned}
\Delta &= PCA_h - PCA_s \\
&= X_h'\beta_h - X_s'\beta_s \\
&= (X_h' - X_s)'\beta_h + X_s'(\beta_h - \beta_s) \\
&= \Delta_X + \Delta_\beta
\end{aligned}
\tag{1}
$$

Table 6 shows that the endowment effect is responsible for 75% of the difference. Even thought, misreporting was expected to amount for most of the difference, the structural component is non-negligible. The table employs the linearity of the OB decomposition to display the contribution that each asset has to the structural and endowment effects. The asset in which the reporting difference is the highest (washing machine) amounts for 22.6% of the overall difference. One limitation of this exercise is that we only have information for six of the seventeen items employed in the PCA calculation[23] (two of which correspond to administrative data).

Figure 3 complements the previous information. Subfigure (c) contrasts PCA_s (X-axis) and PCA_h (Y-axis), adding the 45-degree line as a reference. As can be seen, 83.2% of the observations are below the 45-degree line, which is indicative of the benefits that result from misreporting. In these figures, the applicants who report the same information in both sources (application and high school data) are light-colored. As can be seen, in subfigure (c), most of them appear in the upper-left part of the scatterplot, which is indicative of how negatively they are affected for truthfully[24] reporting their assets. Moreover, we plot the endowment and structural components in subfigures (a) and (b), respectively. The endowment subfigure shows how most applicants under-report assets in the scholarship application since 76.6% of the observations are below the 45-degree line. The structural subfigure displays a non-linear

[23]To assess how sensitive the result is to the use of partial truthful information, we compare it to alternative simulations in which only a proportion of the available information is substituted. Results are available upon request.

[24]Truthful reporting is a simplification, which is equivalent to consistently reporting the assets that were substituted in both data sources.

relation which suggests that those applicants with a higher PCA value in the simulation with the COMIPEMS data seem to benefit more from misreporting.

4.3 ML targeting

In this section, we explore an alternative algorithm to select beneficiaries in the centralized targeting. Recent literature has focused on designing optimal targeting rules for treatment allocation under some constraints (Kitagawa and Tetenov, 2018; Athey and Wager, 2021; Sverdrup et al., 2024; Sun et al., 2024). In this context, a targeting rule is defined as a function that maps a set of covariates to a treatment assignment (binary or multi-treatment), prioritizing individuals who would benefit the most from the treatment. We consider a decision rule that ranks future scholarship assignments by sorting applicants in decreasing order according to their conditional average treatment effect (CATE). To estimate the CATE, we implement the causal forest algorithm developed by Athey et al. (2019).

Unlike standard prediction tasks in supervised machine learning, predicting the CATE involves the challenge of training a model in a setting where the outcome of interest cannot be directly observed. To overcome this, Athey et al. (2019) generalize the random forest algorithm of Breiman (2001) to settings beyond conditional mean prediction. In particular, during the tree-growing process, the binary recursive splitting rule of the covariate space is modified to maximize the squared difference in treatment effects between child nodes.[25] After building the forest, the algorithm predicts the CATE for a given sample x using a residual-on-residual regression, weighted by the frequency with which each training sample i appears in the same leaf as x across all trees in the forest.[26]

We train the causal forest using data from applicants in the simple RCT.[27] We estimate two

[25] In the standard random forest, the algorithm selects the split point that maximizes the squared difference between the predicted outcomes in each child node (Breiman, 2001).

[26] It is worth mentioning that while the traditional random forest makes final predictions by averaging the predictions of each tree in the forest (Breiman, 2001), Athey et al. (2019) use the forest to derive adaptive neighborhood weights, which are then used in an estimating equation. In particular, they apply a weighted version of the residual-on-residual regression method from Robinson (1988) to predict CATEs.

[27] The covariates used to train the causal forest are: age, income per capita, disability, pregnancy, indigenous, female, marginality index, rural, violence program, hunger program, overcrowding, secondary GPA, distance to school, and the total number of household deprivations (assets, services, and characteristics). We keep constant the information being received in the scholarship application (even if it is potentially misreported). Although the

versions of the model: one that predicts the CATE on the probability of not dropping out, and another that predicts the CATE on general test scores.[28] We then use these models to predict the CATEs of applicants in the targeting RCT.

Table 7 compares the selection of applicants and shows that only between 48% and 52% of applicants are equally classified under both methods in the different scenarios simulated. *Tables 8* and *9* compare the characteristics of four different types of applicants (following the classification described in section 3.1).[29] When the ML strategy seeks to reduce dropouts, ML beneficiaries tend to be richer than PMT beneficiaries and, in fact, very similar in asset holding and dwelling characteristics to never receivers. Also, ML beneficiaries in comparison to PMT beneficiaries are older, a lower proportion is pregnant, indigenous and report disabilities. The most notable characteristic of ML beneficiaries is that they live closer to school compared to all other types of applicants (their average distance to school is almost half than that of PMT beneficiaries). Interestingly, when the ML targeting is based on improving test scores, the comparison between ML and PMT beneficiaries changes in some dimensions. Notably, ML beneficiaries compared to PMT beneficiaries have a starting GPA 6% higher, a lower proportion have a matching last name with a committee member, age slightly decreases, they live closer to school, and they also hold more assets and have better dwelling characteristics.

5 Impact on schooling outcomes

Few papers in the literature provide evidence of different targeting strategies on the actual outcomes that they seek to improve (Basurto et al., 2017). Here, we employ the administrative information collected from high schools to analyze the results of the different targeting strategies on dropout rates (the objective of the program) and other relevant educational outcomes, like performance (measured with GPA).

algorithm could benefit from additional information, that is beyond the scope of this paper.

[28]For exposition purposes, we trained the dropout-based causal forest using a dummy variable that takes the value of one if the applicant did not drop out.

[29]For this section, (i) always-receivers, (ii) never-receivers, (iii) PMT beneficiaries (those that receive under the original centralized targeting, but not under the ML algorithm, and (iv) ML beneficiaries who opposite to the previous group receive the scholarship under the ML algorithm, but not in the original centralized assignment.

To begin, we estimate the intent to treat (ITT) employing the randomization, which indicated if a school would select recipients using a local-based (CBT) or a centralized (PMT) targeting strategy. For this, we employ a simple regression:

$$Y_{ij} = \tau_0 + \tau_1 CBT_j + U_{ij}, \tag{2}$$

where Y_{ij} represents the educational outcome of applicant i in school j, CBT_j is a dummy indicating if the committee targeting strategy was randomly assigned to school j, and the error is clustered at the school level. In this estimation, τ_1 is the relevant parameter and estimates the impact of the targeting strategies.

The results of equation (2) are shown in panel (A) of *Table 10*. No significant effects are found for the different outcomes employed. It must be emphasized that this ITT does not display the impact of the scholarship distribution but rather the impact of implementing different targeting strategies. Both groups receive scholarships; the main difference between groups is how the recipients are selected, as the previous section emphasized. Table XX in the appendix shows the differential impact of the scholarships on actual recipients by estimating the heterogeneous effects of CBT on those who receive and do not receive a scholarship. However, this comparison would entail a scholarship and a composition effect since scholarship recipients under local and centralized targeting are different.[30]

Therefore, to make a comparison that would isolate the composition effect, a second specification is estimated by distinguishing the profile types and estimating for each profile the effect of the targeting method selected:

$$Y_{ij} = \alpha + \sum_{k=1}^{4} \Big(\gamma_k \, Type(k)_{ij} + \tau_k \, CBT_j * Type(k)_{ij} \Big) + X'_{ij} \, \beta + U_{ij}, \tag{3}$$

where $Type(1)$ to $Type(4)$ are dummies for the different applicants' profiles: *Always receiver*, *CBT benefactor*, *PMT benefactor*, and *Never receiver*. Our paper has the key advantage that,

[30]For the CBT targeting, receiving the award gathers *Always receivers* and *Committee benefactors* whereas for PMT targeting it gathers *Always receivers* and *PMT benefactors*. For *Always receivers*, no direct effect is evident; targeting would possibly affect them only through spillovers from their peers.

given the information being submitted and the simulations we performed, we are able to identify at the individual level the type of recipient that each applicant is.

The results are shown in panel (B) of *Table 10*. Two aspects are relevant to highlight. First, the committee targeting is worse than the PMT targeting with respect to the program's goal, which is to reduce dropout rates. As explained above, results were expected (if any) in the applicants with profiles that are sensitive to the targeting strategy. *PMT benefactors* benefit from their assignment to a PMT school and reduce dropout rates (across school years) by 2.9 percentage points (p-value = 0.107). This effect is equivalent to a 23% reduction with respect to the average dropout rate). In contrast, *Committee benefactors* do not display any significant effects, and their point estimates are negligible. Second, in terms of school performance, GPAs increase when the committee targeting is employed being this effect greater in the *Committee benefactors'* group. This type of applicants increase their general and math GPA in 0.072σ (NS) and 0.188σ, respectively. With respect to the other outcomes, no significant effects on absenteeism and indiscipline are found for either type. These results are robust to the addition of controls as shown in Panel (C).

6 Conclusion

This paper examined how different mechanisms shape the allocation of a large-scale educational cash transfer program in Mexico. We incorporated the use of randomized experiments, administrative data, and machine learning methods. We focus particularly in three key variations to the traditionally used proxy-means procedure: the use of school committee that can leverage specific information difficult to capture through a centralized process; the vulnerability to misreporting in the proxy-means methodology; and how alternative data-driven targeting approaches would reallocate resources. Our results highlight that targeting for need, merit, or impact can produce dramatically different beneficiary pools.

It should be stressed that we defer from making any welfare interpretation to our results. It is possible that school committees differently distribute resources not only because they might incorporate not observable information and have a more flexible way of selecting beneficiaries, but also because they might be distributing resources according to a particular concept of fairness. Also, potential recipients might be misreporting information not only to appropriate resources, but because of a high willingness to take risks in order to avoid abandoning their studies.

The evidence presented in this paper shows how the proxy-means targeting strategy that has been widely employed in policymaking is very sensitive to variations to some of its components, such as the information source, the respondents motives and the function established as a goal for the algorithm.

References

AIKEN, E., S. BELLUE, D. KARLAN, C. UDRY, AND J. E. BLUMENSTOCK (2022): "Machine learning and phone data can improve targeting of humanitarian aid," *Nature*, 603, 864–870.

ALATAS, V., A. BANERJEE, R. HANNA, B. A. OLKEN, R. PURNAMASARI, AND M. WAI-POI (2016): "Self-Targeting: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 124, 371–427.

——— (2019): "Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia," *AEA Papers and Proceedings*, 109, 334–39.

ALATAS, V., A. BANERJEE, R. HANNA, B. A. OLKEN, AND J. TOBIAS (2012): "Targeting the Poor: Evidence from a Field Experiment in Indonesia," *American Economic Review*, 102, 1206–40.

ALDERMAN, H. (2002): "Do local officials know something we don't? Decentralization of targeted transfers in Albania," *Journal of Public Economics*, 83, 375–404.

ANGRIST, J. D. AND J.-S. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press.

ATHEY, S., J. TIBSHIRANI, AND S. WAGER (2019): "Generalized Random Forests," *The Annals of Statistics*, 47, pp. 1148–1178.

ATHEY, S. AND S. WAGER (2021): "Policy Learning With Observational Data," *Econometrica*, 89, 133–161.

AVITABILE, C. AND R. DE HOYOS (2018): "The heterogeneous effect of information on student performance: Evidence from a randomized control trial in Mexico," *Journal of Development Economics*, 135, 318–348.

BANERJEE, A., E. DUFLO, C. IMBERT, S. MATHEW, AND R. PANDE (2020a): "E-governance, Accountability, and Leakage in Public Programs: Experimental Evidence from a

Financial Management Reform in India," *American Economic Journal: Applied Economics*, 12, 39–72.

BANERJEE, A., R. HANNA, B. A. OLKEN, AND S. SUMARTO (2020b): "The (lack of) distortionary effects of proxy-means tests: Results from a nationwide experiment in Indonesia," *Journal of Public Economics Plus*, 1, 100001.

BANERJEE, A., P. NIEHAUS, AND T. SURI (2019): "Universal Basic Income in the Developing World," *Annual Review of Economics*, 11, 959–983.

BANERJEEA, A., R. HANNA, B. A. OLKEN, AND D. SVERDLIN LISKER (2024): "Social Protection in the Developing World," *Journal of Economic Literature*, 62, 1349–1421.

BARRERA-OSORIO, F. AND D. FILMER (2016): "Incentivizing schooling for learning: Evidence on the impact of alternative targeting approaches," *Journal of Human Resources*, 51, 461–499.

BASURTO, P. M., P. DUPAS, AND J. ROBINSON (2017): "Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi," Working Paper 23383, National Bureau of Economic Research.

BEHRMAN, J., P. SENGUPTA, AND P. TODD (2005): "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico," *Economic Development and Cultural Change*, 54, 237–275.

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2013): "Inference on Treatment Effects after Selection among High-Dimensional Controls," *The Review of Economic Studies*, 81, 608–650.

BEUERMANN, D. W., B. HOFFMANN, M. STAMPINI, D. L. VARGAS, AND D. VERA-COSSIO (2025): "Shooting a moving target: Evaluating targeting tools for social programs when income fluctuates," *Journal of Development Economics*, 172, 103395.

BJÖRKEGREN, D., J. E. BLUMENSTOCK, AND S. KNIGHT (2020): "Manipulation-Proof Machine Learning," .

BOBBA, M. AND V. FRISANCHO (2022): "Self-perceptions about academic achievement: Evidence from Mexico City," *Journal of Econometrics*, 231, 58–73, annals Issue: Subjective Expectations Probabilities in Economics.

BREIMAN, L. (2001): "Random Forests," *Machine Learning*, 45, 5–32.

BROWN, C., M. RAVALLION, AND D. VAN DE WALLE (2018): "A poor means test? Econometric targeting in Africa," *Journal of Development Economics*, 134, 109–124.

CAMACHO, A. AND E. CONOVER (2011): "Manipulation of Social Program Eligibility," *American Economic Journal: Economic Policy*, 3, 41–65.

DE HOYOS, R., O. ATTANASIO, AND C. MEGHIR (2024): "Targeting High School Scholarships to the Poor: The Impact of a Program in Mexico," *Economic Development and Cultural Change*, 72, 1747–1768.

DE JANVRY, A. AND E. SADOULET (2006): "Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality," *World Bank Economic Review*, 20, 1–29.

DUFLO, E., P. DUPAS, AND M. KREMER (2015): "School Governance, Teacher Incentives and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools," *Journal of Public Economics*, 123, 92–110.

DUSTAN, A. (2020): "Can large, untargeted conditional cash transfers increase urban high school graduation rates? Evidence from Mexico City's Prepa Sí," *Journal of Development Economics*, 143, 102392.

FILMER, D. AND N. SCHADY (2011): "Does more cash in conditional cash transfer programs always lead to larger impacts on school attendance?" *Journal of Development Economics*, 96, 150–157.

FORTIN, N., T. LEMIEUX, AND S. FIRPO (2011): "Chapter 1 - Decomposition Methods in Economics," Elsevier, vol. 4 of *Handbook of Labor Economics*, 1–102.

GANIMIAN, A. J. AND R. J. MURNANE (2016): "Improving Education in Developing Countries: Lessons From Rigorous Impact Evaluations," *Review of Educational Research*, 86, 719–755.

GLEWWE, P. AND K. MURALIDHARAN (2016): "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications," in *Handbook of the Economics of Education, 2016*, ed. by S. Machin, L. Woessmann, and E. Hanushek, Elsevier, Handbook of the Economics of Education, 653–743.

HANNA, R. AND D. KARLAN (2017): "Chapter 7 - Designing Social Protection Programs: Using Theory and Experimentation to Understand How to Help Combat Poverty," in *Handbook of Economic Field Experiments*, ed. by A. V. Banerjee and E. Duflo, North-Holland, vol. 2 of *Handbook of Economic Field Experiments*, 515–553.

HAUSHOFER, J., P. NIEHAUS, C. PARAMO, E. MIGUEL, AND M. W. WALKER (2022): "Targeting Impact versus Deprivation," Working Paper 30138, National Bureau of Economic Research.

HUSSAM, R., N. RIGOL, AND B. N. ROTH (2022): "Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design in the Field," *American Economic Review*, 112, 861–98.

JEAN, N., M. BURKE, M. XIE, W. M. A. DAVIS, D. B. LOBELL, AND S. ERMON (2016): "Combining satellite imagery and machine learning to predict poverty," *Science*, 353, 790–794.

KHANNA, G. (2015): "The Effectiveness of Decentralized Education Programs: A Multi-Dimensional Regression Discontinuity Approach," Mimeo.

KITAGAWA, T. AND A. TETENOV (2018): "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

MARTINELLI, C. AND S. PARKER (2009): "Deception and Misreporting in a Social Program," *Journal of the European Economic Association*, 7, 886–908.

Mexican Diario Oficial de la Federación (2017): *Reglas de Operación del Programa Nacional de Becas para el Ejercicio 2017*.

NGUYEN, T. (2008): "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar," Policy Research Working Paper WPS 4372, World Bank, Washington, DC, enGender Impact: The World Bank's Gender Impact Evaluation Database.

NIEHAUS, P., A. ATANASSOVA, M. BERTRAND, AND S. MULLAINATHAN (2013): "Targeting with Agents," *American Economic Journal: Economic Policy*, 5, 206–38.

PARKER, S. W., L. RUBALCAVA, AND G. TERUEL (2008): *Evaluating Conditional Schooling and Health Programs*, Elsevier, vol. 4 of *Handbook of Development Economics*, chap. 62, 3963–4035.

PEÑA, P. (2013): "The impact of conditional cash transfers to high school students on dropout rates: An evaluation of Mexico's PROBEMS," Mimeo.

PRADHAN, M., D. SURYADARMA, A. BEATTY, M. WONG, A. GADUH, A. ALISJAHBANA, AND R. P. ARTHA (2014): "Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia," *American Economic Journal: Applied Economics*, 6, 105–126.

ROBINSON, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

SANTIBAÑEZ, L., R. ABREU-LASTRA, AND J. L., O'DONOGHUE (2014): "School based management effects: Resources or governance change? Evidence from Mexico," *Economics of Education Review*, 39, 97–109.

SMYTHE, I. S. AND J. E. BLUMENSTOCK (2022): "Geographic microtargeting of social assistance with high-resolution poverty maps," *Proceedings of the National Academy of Sciences*, 119, e2120025119.

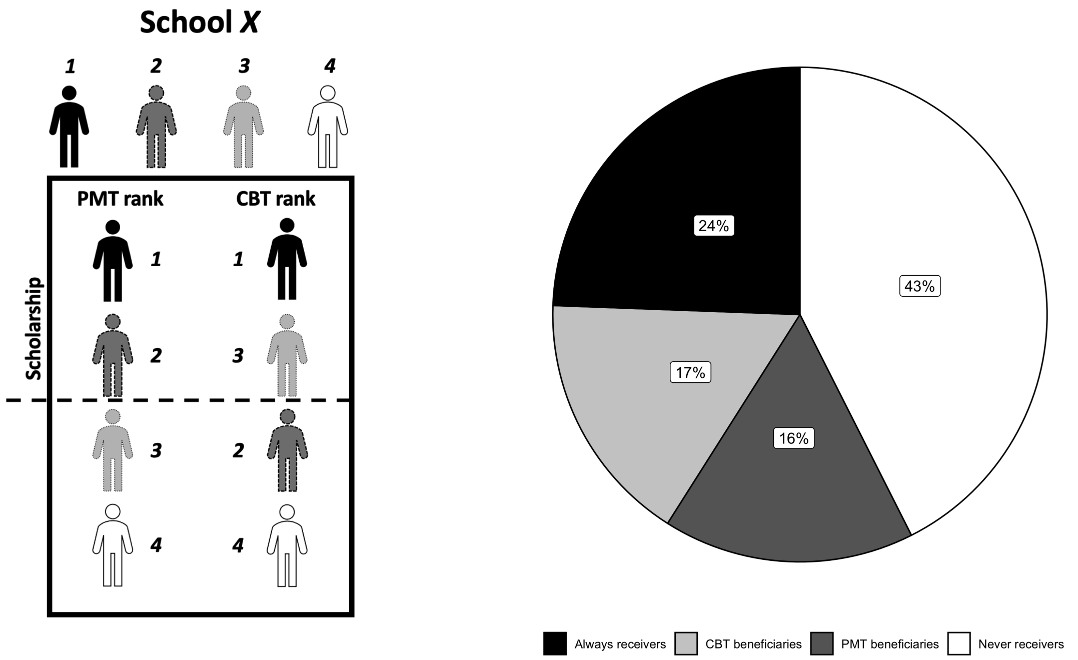
STOEFFLER, Q., B. MILLS, AND C. DEL NINNO (2016): "Reaching the Poor: Cash Transfer Program Targeting in Cameroon," *World Development*, 83, 244–263.

SUN, H., E. MUNRO, G. KALASHNOV, S. DU, AND S. WAGER (2024): "Treatment Allocation under Uncertain Costs," .

SVERDRUP, E., H. WU, S. ATHEY, AND S. W. AND (2024): "Qini Curves for Multi-Armed Treatment Rules," *Journal of Computational and Graphical Statistics*, 0, 1–13.

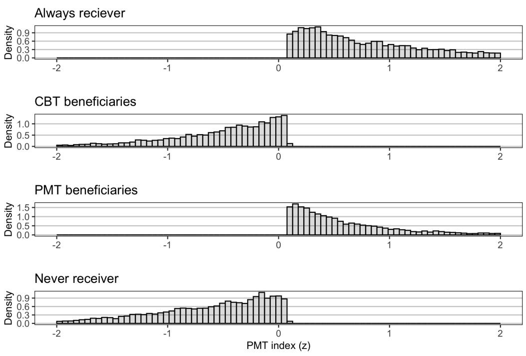
VERA-COSSIO, D. (2021): "Targeting Credit through Community Members," *Journal of the European Economic Association*, 20, 778–821.

Figure 1: Scholarship assignment process

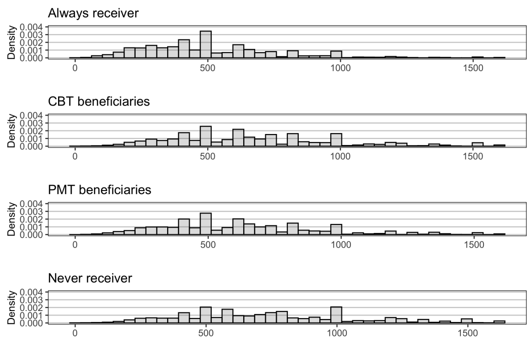


Notes: This figure illustrates the scholarship assignment process under each possible treatment group: (i) centralized or (ii) local-based targeting. The centralized targeting ranks students based on their proxy-means test (PMT) score, while the local-based targeting employs the committee (CBT) rank. In this example, we assume that school X has been assigned 2 scholarships ($n_s = 2$). The PMT ranks applicants 1 as the poorest, followed by applicants 2, 3, and 4, respectively. If the school was assigned to treatment PMT, applicants 1 and 2 would receive the scholarship. The school committee rank does not totally agree with the PMT. The committee places applicant 1 first, followed by applicants 3, 2 and 4. If the school is assigned to CBT treatment, applicants 1 and 3 would receive the scholarship. Applicant 1 is an *always receiver*, applicant 4 is a *never receiver*, while applicants 2 and 3 are *PMT* and *CBT benefactors*. The plot on the right indicates the proportion of each type of applicant in the case of the experiment. Note that the proportion of students 2 and 3 will be identical since one group perfectly substitutes the other.

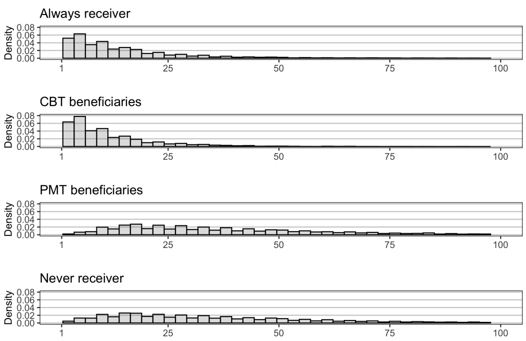
Figure 2: PMT vs CBT Targeting. Beneficiaries Selection



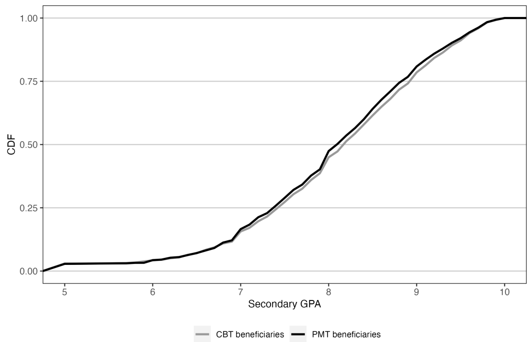
(a) PCA index



(b) Income per capita



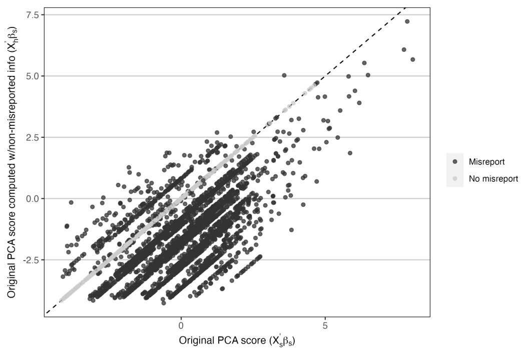
(c) Committee ranking



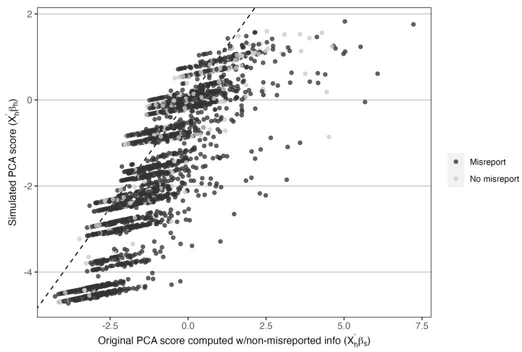
(d) GPA CDF

Notes: This figure shows the distribution of two economic indicators, the committee ranking and the CDF of secondary GPA for the different types of applicants. Panel (a) shows the histogram of the PCA index used to target applicants based on their economic needs. The PCA index is constructed using locality characteristics (marginality index and dummy for participation in the hunger program), household's asset ownership (TV, cable TV, washing machine, refrigerator, vehicle, boiler, gas stove, mobile phone), household's characteristics (dirt floor, sewage, electricity, internet, ratio of individuals per room) and applicant's characteristics (pregnancy, disability, indigenous). Panel (b) shows the histogram of self-reported income per capita. Panel (c) shows the histogram of the ranking employed by the committee for the CBT targeting, where a lower number indicates a higher position in the committee's preferences. Panel (d) shows the cumulative density function of the applicant's GPA in middle school (grades 7th to 9th). The different types of individuals are the following: (1) *Always receiver* is an applicant that receives the scholarship regardless of the targeting strategy; (2) *CBT benefactor* is an applicant that receives the scholarship if the committee rank is used, but does not receive it under the PMT rank; (3) *PMT benefactor* is an applicant that receives the scholarship if the PMT rank is employed, but not according to the committee; and (4) *Never receiver* is an applicant that does not receive the scholarship, regardless of the targeting used. The unit of observation is the applicant. Data employed for these graphs comes from the scholarship applications.

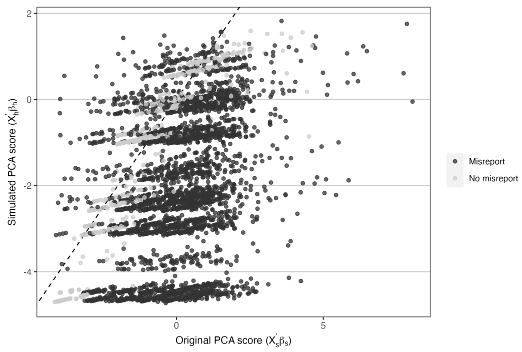
Figure 3: Misreporting. Comparison of the original PCA score and the simulated PCA with high school information



(a) Endowment effect



(b) Structural effect



(c) Overall effect

Notes: This figure shows comparisons between the original PCA score produced with the scholarship application information (which is vulnerable to misreporting) and PCA scores simulated after substituting the high school information. Panel (a) compares the PCA scores by keeping constant the coefficients (β_s) from the original PCA score and varying the information (X_i') employed to calculate the PCA. Panel (b) shows the comparison between the PCA scores by keeping constant the information submitted (X_h') and varying the coefficients employed to calculate the PCAs (β_i). Panel (c) compares the original (PCA_s) and simulated PCA score (PCA_h) after substituting the high school information, where both X_i' and β_i change.

Table 1: Simple RCT. Effect of cash transfers on educational outcomes

	(1) Dropout (*within*)	(2) Absent	(3) Conduct	(4) Overall GPA (z)	(5) Math GPA (z)	(6) Dropout (*across*)
(A) Estimation with controls†						
Cash transfer	-0.003	0.011	-0.002	-0.015	-0.015	-0.002
	(0.002)	(0.125)	(0.002)	(0.010)	(0.010)	(0.007)
(B) Estimation with controls selected by double-selection lasso‡						
Cash transfer	-0.003	0.023	-0.002	-0.010	-0.010	-0.000
	(0.002)	(0.125)	(0.002)	(0.008)	(0.009)	(0.005)
Observations	35,472	33,357	33,357	33,357	33,357	9,939
$\bar{Y}$| cash transfer = 0	0.059	11.324	0.053	0.010	0.009	0.115

Notes: † Only covariates with a p-value of the mean difference between treated and control applicants less than or equal to 0.1 were included as controls in the main estimation (see *Table A.1*). ‡ Selected covariates listed in *Table* A.7. Standard errors clustered at the school level. Asterisks indicate significance at the $^{***}1\%$, $^{**}5\%$ and $^{*}10\%$.

Table 2: Targeting RCT. Balance and descriptive statistics.

Variable	Mean			Difference	p-value
	Overall	PMT	CBT		
(A) Locality characteristics					
Marginality index	-1.259	-1.269	-1.249	0.020	0.533
Rural†	0.258	0.259	0.258	-0.002	0.954
Hunger program†	0.654	0.674	0.634	-0.040	0.488
Violence program†	0.360	0.338	0.383	0.046	0.535
(B) Household assets					
No washing machine†	0.672	0.657	0.687	0.030	0.218
No vehicle†	0.778	0.770	0.785	0.015	0.480
No boiler†	0.695	0.701	0.690	-0.011	0.729
No refrigerator†	0.118	0.115	0.120	0.005	0.715
No mobile phone†	0.227	0.222	0.233	0.012	0.575
No computer†	0.830	0.824	0.836	0.012	0.419
No TV†	0.182	0.169	0.196	0.027	0.193
No cable TV†	0.842	0.829	0.855	0.025	0.222
No stove/grill†	0.116	0.107	0.126	0.019	0.357
(C) Household services					
No sewage disposal†	0.376	0.378	0.375	-0.003	0.941
No electricity†	0.005	0.006	0.003	-0.002	0.091
No internet†	0.808	0.791	0.826	0.034	0.103
No water supply†	0.251	0.251	0.252	0.001	0.958
No gas†	0.111	0.111	0.112	0.001	0.943
(D) Household characteristics					
Dirt floor†	0.042	0.041	0.043	0.002	0.808
Low-quality roofing†	0.020	0.019	0.022	0.003	0.386
Low-quality walls†	0.011	0.010	0.011	0.001	0.794
Inhabitants per room	2.475	2.475	2.474	-0.002	0.964
Overcrowding†	0.278	0.276	0.279	0.003	0.815
(E) Applicant's characteristics					
Age	16.280	16.240	16.320	0.081	0.102
Female†	0.539	0.541	0.538	-0.003	0.821
Disability†	0.038	0.041	0.036	-0.005	0.320
Secondary GPA (MX scale)	8.087	8.086	8.088	0.002	0.961
Income per capita	756.973	785.499	728.036	-57.463	0.066
Indigenous†	0.029	0.036	0.022	-0.015	0.091
Pregnancy†	0.006	0.006	0.006	0.000	0.706
Distance to school (km)	3.583	3.030	4.183	1.153	0.168
Distance to paypoint (km)	3.640	3.739	3.538	-0.201	0.591
Last name match† ‡	0.188	0.192	0.185	-0.007	0.550
Dropout risk	3.919	3.875	3.964	0.089	0.533
Number of applicants	36,094	18,176	17,918		
Number of schools	633	318	315		

Notes: This table shows descriptive statistics and balance of the random assignment considering only applicants with complete information on educational outcomes. PMT corresponds to the schools randomly assigned to use the PMT-based centralized targeting. CBT corresponds to the schools randomly assigned to use the school committee ranking for the targeting. The unit of observation is the applicant. Data comes from the application socio-economic and demographic survey filled out by the students to apply for the scholarship. † indicates that the variable is a dummy. ‡ Complete observations $n = 35,852$. F-stat= 2.060 (p= 0.000). The p-values were calculated using standard errors clustered at the school level.

Table 3: Educational outcomes. Descriptive statistics

Variable	Mean	Standard deviation	Observations
(A) Simple RCT			
Dropout (*within*)	0.058	0.234	35,472
Dropout (*across*)	0.115	0.319	9,939
Indiscipline	0.052	0.223	33,357
Absenteeism	11.339	11.689	33,357
Overall GPA (MX scale)	8.291	0.988	33,357
Math GPA (MX scale)	7.932	1.280	33,357
Failed a course	0.034	0.181	33,357
Failed math	0.109	0.311	33,357
Conduct reports	0.308	1.468	19,108
Cash transfer	0.500	0.500	35,472
(B) Targeting RCT			
Dropout (*within*)	0.078	0.268	36,094
Dropout (*across*)	0.127	0.333	10,074
Indiscipline	0.051	0.219	32,666
Absenteeism	14.082	12.839	32,666
Overall GPA (MX scale)	8.130	1.027	32,666
Math GPA (MX scale)	7.667	1.324	32,666
Failed a course	0.047	0.212	32,666
Failed math	0.156	0.363	32,666
Conduct reports	0.224	1.189	18,158
Cash transfers	0.405	0.491	36,094

Notes: This table shows the descriptive statistics of the educational outcomes for both experiments. The outcomes used are the following: (1) *Dropout (within)* is a dummy variable indicating if a student abandoned school during the school year; (2) *Dropout (across)* is a dummy variable indicating if the student abandoned school between school years; (3) *Indiscipline* is a dummy variable indicating if the student has ever received a conduct report; (4) *Absenteeism* is the total number of days that the student missed classes during the school year; (5) *Overall GPA* is measured between 5 and 10 and represents an average of all the courses taken during the school year; (6) *Math GPA* is measured between 5 and 10 and represent the average of all the math-related courses taken during the year; (7) *Failed a course* indicates if a student has failed a course; (8) *Failing math* is a dummy variable indicating if a student failed its math GPA; (9) *Conduct reports* is the total number of conduct reports received by a student; and (10) *Cash transfer* is a dummy variable indicating if a student received cash transfers as part of the program.

Table 4: Targeting RCT. Descriptive statistics and balance by type of applicant.

Variable	Mean				F-stat	p-value
	Always receivers	CBT beneficiaries	PMT beneficiaries	Never receivers		
(A) Locality characteristics						
Marginality index	-0.980	-1.318***	-1.217	-1.413	104.620	0.000
Rural†	0.373	0.227***	0.314	0.184	49.912	0.000
Hunger program†	0.617	0.655***	0.633	0.682	5.505	0.001
Violence program†	0.216	0.356***	0.364	0.443	7.680	0.000
(B) Household assets						
No washing machine†	0.941	0.533***	0.916	0.479	482.285	0.000
No vehicle†	0.957	0.693***	0.943	0.644	247.827	0.000
No boiler†	0.910	0.584***	0.871	0.547	167.085	0.000
No refrigerator†	0.278	0.022***	0.229	0.020	176.979	0.000
No mobile phone†	0.366	0.135***	0.360	0.133	221.695	0.000
No computer†	0.947	0.775***	0.927	0.746	156.899	0.000
No TV†	0.344	0.071***	0.331	0.076	231.663	0.000
No cable TV†	0.968	0.769***	0.967	0.750	115.358	0.000
No stove/grill†	0.282	0.043***	0.150	0.037	85.310	0.000
(C) Household services						
No sewage disposal†	0.310	0.423***	0.372	0.398	36.062	0.000
No electricity†	0.013	0.000***	0.008	0.000	.	.
No internet†	0.973	0.727***	0.959	0.687	173.891	0.000
No water supply†	0.430	0.164***	0.375	0.135	204.168	0.000
No gas†	0.327	0.005***	0.175	0.004	143.466	0.000
(D) Household characteristics						
Dirt floor†	0.121	0.002***	0.069	0.002	83.132	0.000
Low-quality roofing†	0.054	0.006***	0.025	0.005	36.251	0.000
Low-quality walls†	0.023	0.005***	0.013	0.004	20.784	0.000
Inhabitants per room	2.866	2.184***	2.895	2.201	414.338	0.000
Overcrowding†	0.415	0.176***	0.434	0.179	352.378	0.000
(E) Applicant's characteristics						
Age	16.474	16.127***	16.270	16.232	10.336	0.000
Female†	0.541	0.541	0.555	0.532	2.620	0.050
Disability†	0.037	0.039***	0.050	0.035	5.299	0.001
Secondary GPA (MX scale)	8.044	8.140***	8.086	8.092	4.524	0.004
Income per capita	562.739	776.009***	680.994	890.363	79.359	0.000
Indigenous†	0.098	0.003***	0.022	0.002	19.699	0.000
Pregnancy†	0.009	0.006	0.005	0.005	2.911	0.034
Distance to school (km)	4.790	3.063	3.095	3.275	2.125	0.096
Distance to paypoint (km)	4.382	3.552	3.579	3.291	3.635	0.013
Dropout risk	3.961	3.916	3.900	3.904	0.389	0.761
Last name match†	0.206	0.198**	0.182	0.177	5.318	0.001
Number of applicants	8,813	5,987	5,935	15,359		

Notes: This table shows descriptive statistics by type of applicant. Data of all applicants included in the targeting method experiment. † indicates that the variable is a dummy. The F-statistic and p-value correspond to the joint significance test (Never receiver is the reference group) and were calculated using standard errors clustered at the school level. Asterisks indicate the significance at the ***1%, **5%, and *10% for the mean difference test between CBT and PMT beneficiaries.

Table 5: Misreporting. Changes in eligibility status

Scholarship application PMT ranking	**COMIPEMS simulated PMT ranking** Non-eligible	Eligible
Non-eligible	1,925	751
Eligible	751	748

Notes: This table shows a confusion matrix that compares the scholarship assignment under two alternative scenarios: (A) Scholarship application, which corresponds to the scholarship assignment that results of using the information available in the scholarship applications and following the *centralized targeting* methodology. This is the procedure effectively used in the centralized, or PMT targeting treatment arm; (B) COMIPEMS simulated, which results of simulating the scholarship assignment that would result of following the *centralized targeting* methodology, but substituting the asset information available in the COMIPEMS data, where the applicants have no incentive to misreport. Out of the 17 assets or household characteristics reported by the applicants, six are available in both datasets and were substituted for the simulation. Table A.4 compares the mean values of the six assets available in both datasets. In each scenario, we keep constant the total number of scholarships assigned: $1,499$.

Table 6: Misreporting. Oaxaca-Blinder decomposition

Overall difference	-1.755	
	(0.035)	
	Structural	Endowment
Decomposition	-0.440	-1.315
	(0.041)	(0.045)
Detailed decomposition		
Misreported variables	-0.328	-1.315
	(0.042)	(0.045)
No washing machine	-0.006	-0.397
	(0.017)	(0.016)
No vehicle	-0.162	-0.237
	(0.025)	(0.016)
No refrigerator	0.048	-0.083
	(0.004)	(0.006
No TV	0.029	-0.065
	(0.005)	(0.006)
No cable TV	-0.268	-0.294
	(0.032)	(0.017)
No internet	0.031	-0.239
	(0.009)	(0.011)
Other variables	-0.112	-0.000
	(0.005)	(0.000)

Notes: This table shows the results of the Oaxaca-Blinder decomposition of the mean difference between the original PCA score (PCA_s) and the simulated PCA score after we substitute the information of the assets available in the high school application (PCA_h). The same number of scholarships is kept constant in the simulation $(1,499)$. Equation (1) details how the difference between the PCAs is expressed as the sum of the *structural* and *endowment* effects. The *structural* effect describes the difference that results from the weights given to each asset in the PCA calculation. The *endowment* effect corresponds to the difference resulting in the levels of assets reported. Bootstrapped standard errors in parentheses $(10,000$ iterations).

Table 7: Selection algorithm. Comparison between PMT and ML targeting

(A) ML target: minimize dropout probability		
	ML algorithm ranking	
Original PMT ranking	Non-eligible	Eligible
Non-eligible	22,028	14,793
Eligible	14,793	9,898

(B) ML target: maximize general GPA		
Original PMT ranking	Non-eligible	Eligible
Non-eligible	21,362	15,459
Eligible	15,459	9,232

Notes: This table shows a confusion matrix that compares the scholarship assignment under two alternative scenarios: (A) original PMT targeting, which is the procedure that employs the PCA score to rank applicants using the original scholarship application; (B) ML-based targeting, which follows the ML procedure with the information from the original scholarship application. In each scenario, we keep constant the total number of scholarships assigned: $24,691$. The panels differ only based on the ML objective.

Table 8: Selection algorithm. PMT targeting vs ML targeting (dropout-based)

Variable	Mean			
	Always receivers	PMT beneficiaries	ML beneficiaries	Never receivers
(A) Locality characteristics				
Marginality index	-1.049	-1.082***	-1.316	-1.438
Rural†	0.364	0.306***	0.235	0.135
Hunger program†	0.644	0.651***	0.762	0.713
Violence program†	0.303	0.319***	0.504	0.512
(B) Household assets				
No washing machine†	0.929	0.933***	0.513	0.472
No vehicle†	0.948	0.956***	0.684	0.665
No boiler†	0.885	0.908***	0.552	0.577
No refrigerator†	0.261	0.270***	0.021	0.021
No mobile phone†	0.327	0.363***	0.126	0.134
No computer†	0.930	0.939***	0.745	0.731
No TV†	0.315	0.345***	0.073	0.083
No cable TV†	0.963	0.969***	0.755	0.752
No stove/grill†	0.213	0.245***	0.036	0.040
(C) Household services				
No sewage disposal†	0.276	0.303***	0.389	0.419
No electricity†	0.009	0.010***	0.000	0.000
No internet†	0.966	0.964***	0.689	0.661
No water supply†	0.384	0.425***	0.137	0.133
No gas†	0.252	0.284***	0.005	0.004
(D) Household characteristics				
Dirt floor†	0.089	0.105***	0.002	0.001
Low-quality roofing†	0.036	0.044***	0.007	0.005
Low-quality walls†	0.016	0.023***	0.004	0.005
Inhabitants per room	2.805	2.973***	2.150	2.214
Overcrowding†	0.417	0.453***	0.167	0.184
(E) Applicant's characteristics				
Age	16.659	16.166***	16.432	16.118
Female†	0.574	0.543***	0.569	0.524
Disability†	0.040	0.044***	0.034	0.035
Secondary GPA (MX scale)	8.005	8.145***	8.016	8.182
Income per capita	623.080	624.584***	830.582	922.968
Indigenous†	0.062	0.068***	0.003	0.002
Pregnancy†	0.008	0.007**	0.005	0.005
Distance to school (km)	3.901	6.433***	3.489	4.018
Dropout risk	3.989	4.002**	3.881	3.915
Last name match†	0.209	0.197	0.187	0.189
Number of applicants	9,898	14,793	14,793	22,028

Notes: This table shows descriptive statistics by type of applicant. CATE on dropout was predicted using a Causal Forest trained on simple RCT data. † indicates that the variable is a dummy. The p-values were calculated using standard errors clustered at the school level. Asterisks indicate significance at the ***1%, **5%, and *10% of the mean difference test between PMT and ML beneficiaries.

Table 9: Selection algorithm. PMT targeting vs ML targeting (GPA-based)

Variable	Mean			
	Always receivers	PMT beneficiaries	ML beneficiaries	Never receivers
(A) Locality characteristics				
Marginality index	-1.207	-0.986***	-1.440	-1.353
Rural†	0.257	0.372***	0.122	0.214
Hunger program†	0.690	0.624***	0.780	0.698
Violence program†	0.355	0.287***	0.545	0.483
(B) Household assets				
No washing machine†	0.937	0.928***	0.479	0.495
No vehicle†	0.960	0.948***	0.681	0.666
No boiler†	0.907	0.894***	0.556	0.574
No refrigerator†	0.272	0.263***	0.024	0.019
No mobile phone†	0.368	0.337***	0.137	0.126
No computer†	0.937	0.935***	0.730	0.742
No TV†	0.354	0.321***	0.081	0.077
No cable TV†	0.972	0.964***	0.751	0.754
No stove/grill†	0.232	0.232***	0.040	0.038
(C) Household services				
No sewage disposal†	0.308	0.283***	0.412	0.403
No electricity†	0.010	0.010***	0.000	0.000
No internet†	0.963	0.966***	0.659	0.682
No water supply†	0.404	0.411***	0.126	0.141
No gas†	0.252	0.283***	0.003	0.005
(D) Household characteristics				
Dirt floor†	0.096	0.101***	0.001	0.001
Low-quality roofing†	0.045	0.039***	0.006	0.005
Low-quality walls†	0.022	0.019***	0.005	0.005
Inhabitants per room	3.143	2.765***	2.249	2.144
Overcrowding†	0.527	0.386***	0.211	0.153
(E) Applicant's characteristics				
Age	16.241	16.438***	16.130	16.327
Female†	0.552	0.557***	0.508	0.567
Disability†	0.049	0.039*	0.034	0.035
Secondary GPA (MX scale)	8.148	8.053	8.101	8.126
Income per capita	652.529	606.933***	928.129	855.256
Indigenous†	0.046	0.077***	0.002	0.003
Pregnancy†	0.007	0.007***	0.004	0.005
Distance to school (km)	5.011	5.691***	3.819	3.726
Dropout risk	4.046	3.968	3.886	3.912
Last name match†	0.196	0.206***	0.186	0.189
Number of applicants	9,232	15,459	15,459	21,362

Notes: This table shows descriptive statistics by type of applicant. CATE on dropout was predicted using a Causal Forest trained on simple RCT data. † indicates that the variable is a dummy. The p-values were calculated using standard errors clustered at the school level. Asterisks indicate significance at the ***1%, **5%, and *10% of the mean difference test between PMT and ML beneficiaries.

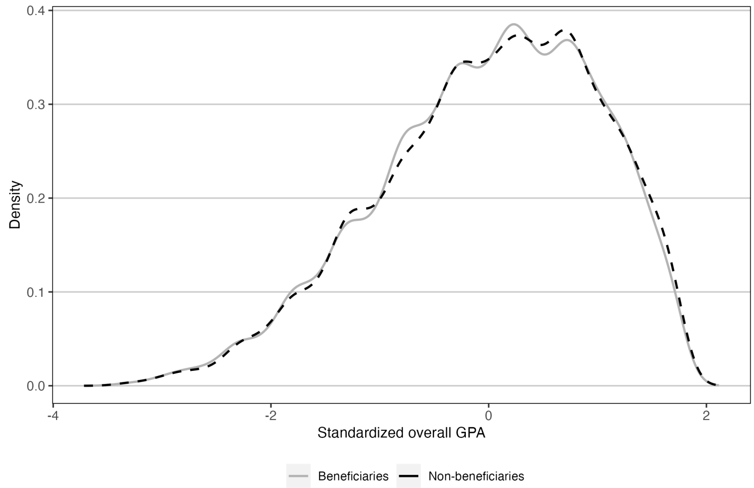
Table 10: Targeting RCT. Effect of cash transfers on educational outcomes

	(1) Dropout (*within*)	(2) Absent	(3) Conduct	(4) Overall GPA (z)	(5) Math GPA (z)	(6) Dropout (*across*)
(A) Effects on beneficiaries						
CBT	0.0139	0.8136	-0.0048	0.0199	0.0819	0.0040
	(0.010)	(1.955)	(0.013)	(0.062)	(0.070)	(0.012)
(B) Effects by type of applicant						
CBT × Always receiver	0.017	-0.012	-0.008	0.043	0.040	0.009
	(0.013)	(1.608)	(0.012)	(0.066)	(0.069)	(0.018)
CBT × CBT beneficiary	0.003	1.294	-0.008	0.072	0.188**	-0.002
	(0.013)	(2.064)	(0.015)	(0.071)	(0.083)	(0.017)
CBT × PMT beneficiary	0.011	1.688	0.002	-0.002	0.088	0.029
	(0.011)	(2.092)	(0.015)	(0.069)	(0.084)	(0.018)
CBT × Never receiver	0.017	1.134	-0.004	-0.005	0.059	-0.010
	(0.015)	(2.489)	(0.016)	(0.086)	(0.090)	(0.019)
(C) Effects by type of applicant (with controls selected by double-selection lasso†)						
CBT × Always receiver	0.016	-0.258	-0.008	0.039	0.054	0.010
	(0.012)	(1.470)	(0.013)	(0.039)	(0.049)	(0.017)
CBT × CBT beneficiary	0.006	1.387	-0.010	0.066	0.179**	0.003
	(0.012)	(1.743)	(0.015)	(0.053)	(0.072)	(0.016)
CBT × PMT beneficiary	0.011	1.834	0.001	0.017	0.100	0.023
	(0.010)	(1.807)	(0.015)	(0.052)	(0.075)	(0.018)
CBT × Never receiver	0.018	1.107	-0.005	0.002	0.061	-0.003
	(0.014)	(2.099)	(0.015)	(0.063)	(0.073)	(0.018)
Observations	36,094	32,666	32,666	32,666	32,666	10,074
$\bar{Y}$ | CBT = 0	0.071	13.681	0.053	-0.010	-0.040	0.125

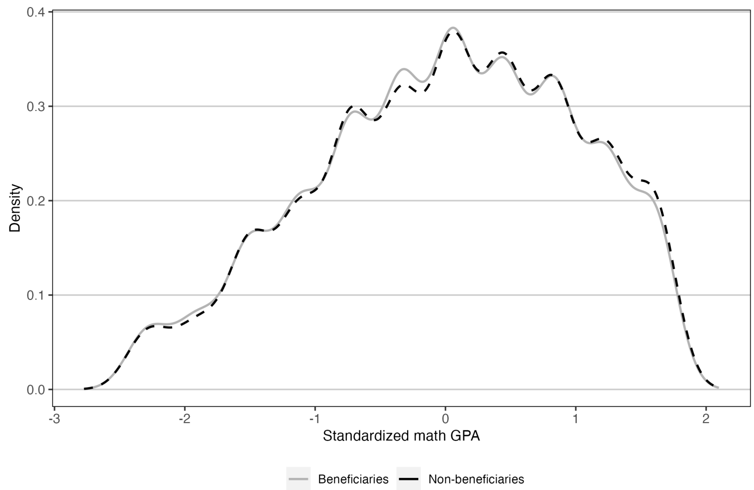
Notes: † Selected covariates listed in *Table* A.8. Standard errors clustered at the school level. Asterisks indicate significance at the ***1%, **5%, and *10%.

A Appendix

Figure A.1: Simple RCT. Overall GPA distributions



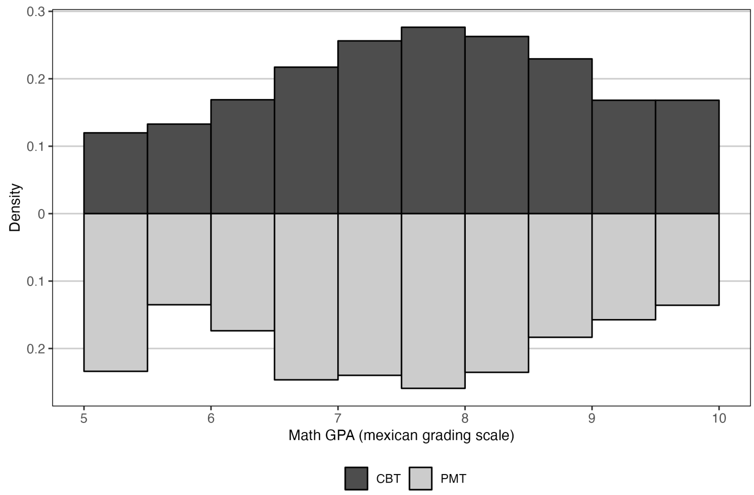
(a) Overall GPA



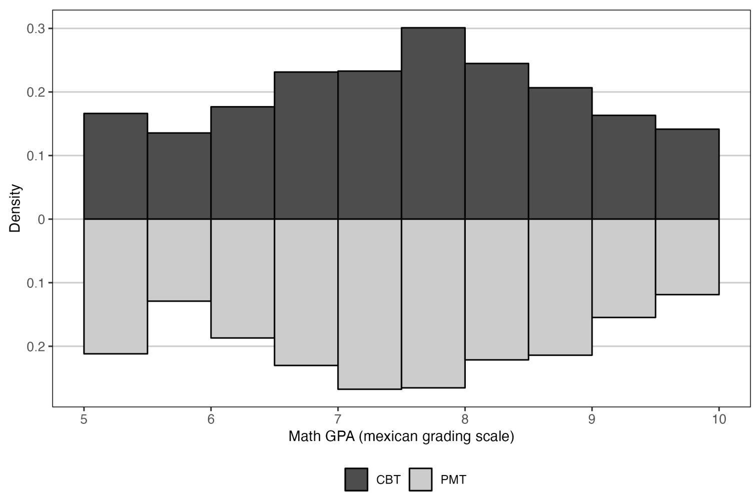
(b) Math GPA

Notes: This figure shows the distributions of standardized overall and math GPAs for beneficiary and non-beneficiary students in the cash transfers experiment.

Figure A.2: Targenting RCT. Math GPA distributions



(a) CBT beneficiaries



(b) PMT beneficiaries

Notes: This figure shows math GPA distributions for CBT and PMT beneficiaries using the Mexican grading scale (5 to 10). Students receiving a score below six are assumed to fail.

Table A.1: Simple RCT. Applicants with complete information on dropout

Variable	Mean			Difference	p-value
	Overall	No cash transfer	Cash transfer		
(A) Locality characteristics					
Marginality index	-1.044	-1.042	-1.047	-0.005	0.320
Rural†	0.304	0.304	0.305	0.001	0.818
Hunger program†	0.603	0.604	0.602	-0.002	0.663
Violence program†	0.241	0.245	0.237	-0.008	0.072
(B) Household assets					
No washing machine†	0.934	0.934	0.933	-0.001	0.802
No vehicle†	0.963	0.962	0.965	0.003	0.111
No boiler†	0.903	0.904	0.902	-0.002	0.579
No refrigerator†	0.393	0.395	0.391	-0.004	0.483
No mobile phone†	0.436	0.436	0.437	0.001	0.826
No computer†	0.930	0.929	0.931	0.002	0.434
No TV†	0.359	0.364	0.353	-0.010	0.056
No cable TV†	0.968	0.968	0.968	0.000	0.946
No stove/grill†	0.302	0.302	0.302	0.000	0.965
(C) Household services					
No sewage disposal†	0.000	0.000	0.000	0.000	–
No electricity†	0.021	0.021	0.021	0.000	0.995
No internet†	0.963	0.963	0.963	0.001	0.740
No water supply†	0.493	0.496	0.490	-0.007	0.188
No gas†	0.379	0.379	0.379	0.000	0.926
(D) Household characteristics					
Dirt floor†	0.167	0.167	0.168	0.001	0.774
Low-quality roofing†	0.090	0.088	0.092	0.003	0.250
Low-quality walls†	0.030	0.028	0.032	0.003	0.047
Inhabitants per room	3.191	3.195	3.188	-0.007	0.644
Overcrowding†	0.529	0.528	0.530	0.002	0.750
(E) Applicant's characteristics					
Age	16.088	16.087	16.090	0.004	0.837
Female†	0.560	0.565	0.554	-0.010	0.056
Disability†	0.028	0.027	0.029	0.003	0.117
Secondary GPA (MX scale)	8.348	8.355	8.341	-0.014	0.175
Income per capita	544.397	545.235	543.558	-1.677	0.711
Indigenous†	0.083	0.084	0.083	0.000	0.900
Pregnancy†	0.006	0.006	0.006	0.000	0.788
Distance to school (km)‡	4.562	4.598	4.526	-0.071	0.183
Number of applicants	35,472	17,732	17,740		
Number of schools	1,586	1,441	1,451		

This table shows descriptive statistics and balance of the random assignment considering only applicants with complete information on dropout. The unit of observation is the applicant. Data employed for this table comes from the socio-economic survey filled in the application. † Indicates that the variable is a dummy. F-stat= 1.030 (p= 0.420). ‡ Complete observations $n = 35,449$. The p-values were calculated using standard errors clustered at the school level.

Table A.2: Simple RCT. Attrition

Variable	Mean			Difference	p-value
	Overall	Non-attritors	Attritors		
(A) Locality characteristics					
Marginality index	-1.052	-1.044	-1.070	-0.026	0.229
Rural†	0.316	0.304	0.344	0.040	0.019
Hunger program†	0.614	0.603	0.639	0.036	0.306
Violence program†	0.268	0.241	0.335	0.094	0.016
(B) Household assets					
No washing machine†	0.936	0.934	0.942	0.009	0.347
No vehicle†	0.962	0.963	0.957	-0.006	0.043
No boiler†	0.899	0.903	0.891	-0.012	0.170
No refrigerator†	0.385	0.393	0.367	-0.026	0.087
No mobile phone†	0.441	0.436	0.451	0.015	0.242
No computer†	0.926	0.930	0.917	-0.013	0.001
No TV†	0.362	0.359	0.371	0.012	0.330
No cable TV†	0.967	0.968	0.964	-0.004	0.238
No stove/grill†	0.290	0.302	0.259	-0.043	0.003
(C) Household services					
No sewage disposal†	0.000	0.000	0.000	0.000	–
No electricity†	0.020	0.021	0.020	0.000	0.966
No internet†	0.960	0.963	0.954	-0.009	0.006
No water supply†	0.485	0.493	0.466	-0.027	0.045
No gas†	0.365	0.379	0.331	-0.048	0.008
(D) Household characteristics					
Dirt floor†	0.155	0.167	0.127	-0.040	0.000
Low-quality roofing†	0.079	0.090	0.052	-0.038	0.000
Low-quality walls†	0.029	0.030	0.026	-0.004	0.394
Inhabitants per room	3.165	3.191	3.102	-0.089	0.109
Overcrowding†	0.523	0.529	0.507	-0.022	0.135
(E) Applicant's characteristics					
Age	16.074	16.088	16.038	-0.050	0.158
Female†	0.564	0.560	0.573	0.014	0.060
Disability†	0.030	0.028	0.034	0.006	0.011
Secondary GPA (MX scale)	8.337	8.348	8.311	-0.037	0.133
Income per capita	559.396	544.397	596.016	51.619	0.000
Indigenous†	0.084	0.083	0.084	0.001	0.959
Pregnancy†	0.006	0.006	0.006	-0.001	0.443
Distance to school (km)‡	4.727	4.562	5.116	0.553	0.004
Number of applicants	50,001	35,472	14,529		
Number of schools	2,146	1,586	741		

This table shows the analysis of attrition of applicants for which administrative information is not available. The unit of observation is the applicant. Data employed for this table comes from the socio-economic survey filled in the application. F-stat= 3.370 (p= 0.000) † indicates that the variable is a dummy. ‡ Complete observations $n = 49,984$. The p-values were calculated using standard errors clustered at the school level.

Table A.3: Targeting RCT. Attrition

Variable	Mean			Difference	p-value
	Overall	Non-attritors	Attritors		
(A) Locality characteristics					
Marginality index	-1.261	-1.259	-1.262	-0.003	0.904
Rural†	0.237	0.258	0.207	-0.052	0.005
Hunger program†	0.699	0.654	0.762	0.109	0.004
Violence program†	0.430	0.360	0.529	0.169	0.002
(B) Household assets					
No washing machine†	0.666	0.672	0.657	-0.015	0.528
No vehicle†	0.785	0.778	0.795	0.018	0.262
No boiler†	0.700	0.695	0.707	0.011	0.641
No refrigerator†	0.119	0.118	0.122	0.004	0.634
No mobile phone†	0.218	0.227	0.205	-0.023	0.170
No computer†	0.817	0.830	0.798	-0.032	0.016
No TV†	0.181	0.182	0.179	-0.004	0.792
No cable TV†	0.839	0.842	0.834	-0.008	0.594
No stove/grill†	0.116	0.116	0.116	-0.001	0.958
(C) Household services					
No sewage disposal†	0.361	0.376	0.339	-0.037	0.175
No electricity†	0.004	0.005	0.003	-0.001	0.066
No internet†	0.790	0.808	0.763	-0.045	0.017
No water supply†	0.245	0.251	0.235	-0.016	0.339
No gas†	0.111	0.111	0.111	0.000	0.975
(D) Household characteristics					
Dirt floor†	0.040	0.042	0.039	-0.003	0.489
Low-quality roofing†	0.020	0.020	0.019	-0.001	0.643
Low-quality walls†	0.011	0.011	0.011	0.001	0.679
Inhabitants per room	2.476	2.475	2.479	0.004	0.895
Overcrowding†	0.282	0.278	0.288	0.010	0.414
(E) Applicant's characteristics					
Age	16.292	16.280	16.310	0.031	0.445
Female†	0.547	0.539	0.559	0.020	0.031
Disability†	0.038	0.038	0.037	-0.002	0.682
Secondary GPA (MX scale)	8.105	8.087	8.129	0.042	0.211
Income per capita	780.736	756.973	814.481	57.508	0.018
Indigenous†	0.028	0.029	0.026	-0.003	0.609
Pregnancy†	0.006	0.006	0.005	-0.001	0.107
Distance to school (km)‡	2.896	3.583	1.902	-1.681	0.000
Last name match† ‡	0.194	0.188	0.201	0.013	0.132
Dropout risk	3.940	3.919	3.969	0.050	0.636
Number of applicants	61,512	36,094	25,418		
Number of schools	1,001	633	627		

This table shows the analysis of attrition of applicants for which administrative information is not available. The unit of observation is the applicant. Data for this table comes from the socio-economic survey filled in the application. Information from all applicants in either PMT or CBT treatment schools is used. F-stat=3.660 (p=0.000). † Indicates that the variable is a dummy. ‡ Complete observations n =60,074. The p-values were calculated using standard errors clustered at the school level.

Table A.4: Misreporting. Descriptive statistics of misreported variables.

Variable	Mean		Difference	p-value
	Misreported	Non-misreported		
No washing machine†	0.653	0.175	0.478	0.000
No vehicle†	0.876	0.618	0.258	0.000
No refrigerator†	0.159	0.058	0.101	0.000
No TV†	0.122	0.017	0.105	0.000
No cable TV†	0.901	0.635	0.267	0.000
No internet†	0.656	0.360	0.296	0.000
Number of applicants	4,175			

Notes: This table shows the mean of variables for which we can observe misreported and non-misreported information. † indicates that the variable is a dummy. The p-values were calculated using standard errors clustered at the school level.

Table A.5: Misreporting. Oaxaca-Blinder decomposition using only washing machine and TV as misreported assets.

Overall difference	-0.966	
	(0.075)	
	Structural	Endowment
Decomposition	-0.504	-0.462
	(0.064)	(0.017)
Detailed decomposition		
Misreported variables	-0.320	-0.462
	(0.050)	(0.017)
No washing machine	-0.326	-0.397
	(0.047)	(0.016)
No TV	0.006	-0.065
	(0.006)	(0.006)
Other variables	-0.185	-0.000
	(0.015)	(0.000)

Notes: This table shows the results of the Oaxaca-Blinder decomposition of the mean difference between the original PCA score (PCA_s) and the simulated PCA score after we substitute the information on washing machine and TV, which correspond to the most and least misreported assets. Bootstrapped standard errors in parentheses (10,000 iterations).

Table A.6: Misreporting. Oaxaca-Blinder decomposition using only washing machine and refrigerator as misreported assets.

Overall difference	-1.028	
	(0.060)	
	Structural	Endowment
Decomposition	-0.547	-0.480
	(0.050)	(0.016)
Detailed decomposition		
Misreported variables	-0.349	-0.480
	(0.040)	(0.016)
No washing machine	-0.365	-0.397
	(0.039)	(0.016)
No refrigerator	0.016	-0.083
	(0.004)	(0.006
Other variables	-0.198	-0.000
	(0.014)	(0.006)

Notes: This table shows the results of the Oaxaca-Blinder decomposition of the mean difference between the original PCA score (PCA_s) and the simulated PCA score after we substitute the information on washing machine and refrigerator, with the latter being the household asset least correlated with washing machine (most misreported asset). Bootstrapped standard errors in parentheses (10,000 iterations).

Table A.7: Simple RCT. Variables selected by double-selection lasso.

Variable	Dropout (*within*)	Absent	Conduct	Overall GPA (z)	Math GPA (z)	Dropout (*across*)
(A) Locality characteristics						
Marginality index		$\times$				
Rural†						
Hunger program†		$\times$				
Violence program†		$\times$	$\times$	$\times$	$\times$	
(B) Household assets						
No washing machine†		$\times$				
No vehicle†		$\times$				
No boiler†		$\times$	$\times$			
No refrigerator†		$\times$	$\times$	$\times$		
No mobile phone†		$\times$	$\times$			
No computer†		$\times$				
No TV†						
No cable TV†		$\times$				
No stove/grill†						
(C) Household services						
No sewage disposal†						
No electricity†						
No internet†		$\times$				
No water supply†		$\times$	$\times$			
No gas†			$\times$			
(D) Household characteristics						
Dirt floor†					$\times$	
Low-quality roofing†						
Low-quality walls†						
Inhabitants per room		$\times$		$\times$		$\times$
(E) Applicant's characteristics						
Age		$\times$		$\times$	$\times$	
Female†		$\times$	$\times$	$\times$	$\times$	
Disability†						
Secondary GPA (z)	$\times$		$\times$	$\times$	$\times$	$\times$
Income per capita		$\times$				
Indigenous†						
Pregnancy†						
Distance to school (km)						

Notes: $\times$ indicates that the double-selection lasso algorithm selected the variable (Belloni et al., 2013). † indicates that the variable is a dummy.

Table A.8: Targeting RCT. Variables selected by double-selection lasso.

Variable	Dropout (*within*)	Absent	Conduct	Overall GPA (z)	Math GPA (z)	Dropout (*across*)
(A) Locality characteristics						
Marginality index	×	×	×	×	×	
Rural†	×					×
Hunger program†	×	×	×	×	×	×
Violence program†	×	×	×	×	×	
(B) Household assets						
No washing machine†	×	×	×	×	×	
No vehicle†	×	×	×			
No boiler†		×				×
No refrigerator†		×				
No mobile phone†		×		×		
No computer†				×	×	
No TV†	×	×	×	×	×	
No cable TV†	×	×	×	×	×	
No stove/grill†	×	×	×	×	×	
(C) Household services						
No sewage disposal†	×	×	×			×
No electricity†						
No internet†	×	×	×	×	×	×
No water supply†		×		×	×	
No gas†						
(D) Household characteristics						
Dirt floor†			×			
Low-quality roofing†						
Low-quality walls†						
Inhabitants per room						
(E) Applicant's characteristics						
Age	×			×	×	
Female†	×	×	×	×	×	
Disability†		×				
Secondary GPA (z)	×		×	×	×	×
Income per capita	×	×	×	×	×	
Indigenous†	×	×	×	×	×	×
Pregnancy†						
Dropout risk	×	×	×	×	×	
Distance to school (km)		×	×			
Distance to paypoint (km)		×	×			

Notes: × indicates that the double-selection lasso algorithm selected the variable (Belloni et al., 2013). † indicates that the variable is a dummy.