Selection Bias Among Participants in High School Financial Literacy Programs: Evidence from PISA *

LAURA HOSPIDO† Banco de España and IZA

ERNESTO VILLANUEVA‡ Banco de España

This draft: November 16, 2018

Abstract

Selection biases arise in the estimation of the impact of financial literacy programs when the set of schools that choose to participate is not random. This study identifies selection bias as the relative performance in the 2012 Financial PISA of students in Spanish high schools that selected into a financial literacy program launched immediately afterwards. We first document that schools that chose to participate were selective centers located in regions where students typically obtain below-average scores. Both sets of characteristics generate biases in conflicting directions. Once we compare students in schools that ended up participating in the program to students in similarly selective nonparticipant schools, we find that future participants underperform in Finance relative to Math. We study if reweighting and difference-in-difference estimators eliminate those biases at baseline. We also test each model's identifying assumptions.

Keywords: Financial Education, Impact Evaluation, Selection Bias

JEL: D14, I22

*We thank seminar participants at the 2016 ASSA Meetings, Banco de España and the Cherry Blossom Financial Education Institute for their comments and suggestions on a previous version of this study. Olympia Bover, Michael Collins, Annamaria Lusardi and Justin McCrary gave very useful comments. Madalen Castells, Lucía Sánchez, Ismael Sanz and Francisco Javier García Crespo provided invaluable help with the data used in this article and Maria Torrado superb research assistance. All opinions are our own. First draft: June 26, 2014.

†laura.hospido@bde.es

‡ernesto.villanueva@bde.es (corresponding author)

1 Introduction

Several countries around the world have introduced financial literacy courses in high school in an attempt to increase the financial competences of the population.[1] Whether or not the inclusion of those contents in high school have succeeded in raising financial knowledge is subject to debate. For example, in their review of the impact of financial literacy on knowledge and behavior across some 200 studies Fernandes et al (2014) note that the impact of financial literacy on behavior in observational studies differs from that in experimental ones. A possible explanation of that finding is that individuals receiving financial literacy training differ from those who do not in characteristics that are correlated with their level of financial literacy. This study characterizes possible selection biases among schools that deliver personal finance courses using data from the 2012 Financial PISA. Namely, we compare the baseline knowledge of students in schools that subsequently applied for a financial education program to the rest of students. As, according to school administrators, very few schools in Spain provided financial literacy training prior to 2012, schools choosing not to participate in the program are representative of the whole population.

We estimate the role of selection both in observable and in unobservable variables, by combining, on the one hand, the information available in the PISA 2012 financial literacy test for Spain with, on the other, the existence of a 10-hour financial literacy program offered between late 2012 and 2016. We identify in the PISA sample those schools that volunteered for the program and that had not offered a personal finance program before. In this sample, any difference in performance on PISA financial tests among students of the schools that volunteered for the program and the rest of the schools can be attributed solely to a selection bias. The wealth of PISA data allows us to (i) characterize the selection bias associated with volunteering to participate in a FE program and (ii) to study to what extent reweighting and difference-in-difference estimators mitigate the role of unobservable variables.

We find that schools selected into the program were located in regions where students typically obtain below-average scores in PISA measurements. However, these schools also had more selective student admissions and turnover policies. Both sets of characteristics generate biases in conflicting directions, so the unadjusted scores of students in participant schools are similar to those of nonparticipants. Once we compare students in schools that select into the program to students in similar locations and in similarly selective centers, we find important differences in competences in Finance, Math and Reading. Namely, at baseline, students in schools that participate into the program underperform in the Financial Competences assessment relative to the Math assessment. Broadly speaking, the schools that self-select into the program are those

[1] The lack of knowledge of basic financial concepts among various groups of the population of the US is discussed in Lusardi and Mitchell (2009, 2010), Mandell (2008) and Lusardi, Mitchell and Curto (2010). Fornero and Monticone (2011) discuss the case of Italy, Buecher-Koenen and Lusardi (2011) the case of Germany and Alessie et al (2011) the case of Netherlands. Van Rooij e tal (2011) discuss the consequences of financial literacy on household portfolios.

whose students had a lower comparative advantage in Finance -at least with respect to Math.

We study if two different estimators address those differences at baseline -or selection bias. The first estimator reweights the sample of controls so that the mean characteristics of the group are similar to those of students in "to-be-treated" schools. The second estimator is Differences-in-Differences, which compares the student-specific gap between the scores in Finance and Math (or between Reading and Math) for treated and controls. As in our setting the true impact on competences of applying for, but not yet delivering, Financial Literacy courses should be zero, we can test the underlying assumptions in both estimators, following Heckman et al (1998). Namely, reweighting estimators estimate the parameter of interest if, conditional on covariates, schools that select into the program do not do that on the basis of the associated gain in the financial knowledge of their students.[2] We test for that assumption analyzing if students in schools that eventually participate in the program differ from the rest in Math or Reading competences, conditional on covariates.[3] The underlying assumption in Differences-in-Differences estimators is that, in the absence of the intervention, treated and control students should have similar gaps between the scores in the Finance and Math assessments and between the Finance and Reading assessments. Even though the assumptions are not met in all specifications, we find that Differences-in-Differences estimators can eliminate the selection bias.

Our results highlight the relevance of having data on school's resources and admission policies as well as diverse measures of students' competences to properly evaluate the impact of Financial Literacy courses. In the case we analyze, information on student admissions and retention criteria plays a crucial role in identifying the right comparison group for students in regions that typically obtain low PISA scores. On the basis of basic information on school location and parental background alone, our models fail to detect important differences in knowledge that do arise once we compare schools with similar selection criteria. Having information in a wide set of competences in PISA (Mathematics, Reading and Literacy) is crucial in detecting the biases that arise because treated students happen to outperform similar controls in Math, but less so in Reading or Finance. Taking student-level pairwise differences between the scores in different domains allows us to detect and address selection biases.

Our results also contribute to the growing literature assessing the relevance of financial education courses for students under the age of 18 in two dimensions. The analysis of some interventions have shown that short financial literacy courses may increase financial knowledge of adolescents but that, unsurprisingly, the impact varies across studies.[4] That variation is to

[2] Formally, participating in a financial literacy program is independent of both the distribution of student's financial competences upon receiving the course and the the counterfactual distribution of financial knowledge had the school not received the course.

[3] For example, Imbens (2014) recommends "estimating the causal effect of the treatment on a pseudo outcome, a variable known to be unaffected by it, typically because its value is determined prior to the treatment itself. Such a variable can be time-invariant, but the most interesting case is in considering the treatment effect on a lagged outcome". Scores in Math and Reading parts of the PISA assessment are also determined prior to the treatment.

[4] Without being exhaustive, Walstad et al. (2010) show that a video on financial literacy presented to high

be expected, because teachers' financial knowledge, the intensity of implementation and the courses where the material is delivered differ across interventions. On top of those issues, our results suggest that selection bias can explain the diversity of results across studies, as different financial programs are directed to different groups of students.[5] Our results also offer insights for the evaluation of randomized experiments in Financial Literacy. In an experiment, each school that volunteers can be assigned to the treated group (schools that offer the course) or to the control group (schools that do not offer it), but a non-volunteer school cannot be forced to offer the course. If the treated schools differ from the rest along their student admission criteria, extrapolating the results of specific interventions require an assessment of the differences at baseline.

The rest of the paper is organized as follows. In Section 2 we describe the most important features of the FE program offered in Spanish schools. Section 3 discusses the approach to estimate selection biases. Section 4 presents the data, Section 5 the estimates of selection bias and Section 6 conducts robustness checks. Finally, Section 7 concludes.

2 The BdE-CNMV Program

We start by mentioning a few characteristics of the Program we analyze. Although none of the schools had even applied for participation at the time when we measure outcomes, some characteristics give background to understand selection. Participation in the program is voluntary, and is achieved after registering in a webpage. That registration gives access to a set of materials for the teacher and a different one to students. Those materials contain three main blocks on Personal Finance, covering consumption and saving, banking relationships and sustainable consumption. Consumption and saving include the definition and reasons for saving, as well as an introduction to basic budgeting clues, like classifying expenses into variable and fixed. Banking relationships covers the notion of credit and debit cards, explains what interest rates and fees are and provides examples. Finally, the sustainable consumption part gives clues to avoid conspicuous expenditures. In latter waves of the program included additional modules on insurance products and pension saving.

The program is intended to be delivered in a flexible manner: provides a set of materials that teachers could deliver in different classes, including Maths but also Social Sciences. Although

school students increased scores in financial knowledge exams substantially. Bruhn et al. (2013) conduct a randomized experiment that randomly assigns a 78-hours course across Brazilian states. The course increased financial (and economic) knowledge in an objective test by a third of a standard deviation. Lührmann et al. (2012) show that a 90 minutes financial literacy course increased self-assessed financial knowledge as well as the ability to assess risks correctly of a sample of German teenagers between 14 and 16. However, those students did not perform well in other topics related to the course. Bechetti et al. (2011) do not find any impact of financial literacy courses on financial knowledge in Italian high schools.

[5] For example, the sample used by Lürhmann et al. (2012) is composed of students of low stream high schools in Germany, leaving out other tracks. The samples in Bruhn et al. (2013) or Walstad et al. (2010) contain public high schools that volunteered for the program.

there were suggestions on a number of activities that could support the materials, participation involved no compulsory exams. Finally, the program was publicized through the webpage of both sponsoring institutions and through school federations. [6]

3 The definition of selection bias in our setting

This Section describes our use of the PISA data on financial literacy to characterize possible selection biases in a context in which we know that the true effect of a non-yet-delivered program should be zero. In principle, differences in financial knowledge between an untreated group yet untreated and controls reflects a host of factors, of which selection bias is only a particular component.

First of all, the score in Financial Competences may differ across students simply because students in "to-be-treated" schools have characteristics that no student in control schools shares. That term does not reflect a selection bias, because both groups are simply not comparable. A second component arises because students in "to-be-treated" schools have characteristics that differ from those in the control group. In our particular case, a disproportionate fraction of "to-be-treated" schools are located in the South of Spain, a set of regions where students obtain on average lower scores in the Mathematics and Reading part of the PISA assessment than students from other regions. As long as there are control schools located in the South, matching or regression estimators can mitigate that source of biases by giving more weight to Southern schools in the treated-control comparisons. Finally, the third component of the bias is a difference in knowledge within the same values of observable covariates, and is due to the influence of variables not observed in PISA that affect student's scores and lead schools to participate into the program. This is the selection bias term, that is well defined once we compare students with the same observable characteristics.

The definition of bias depends on the particular set distribution of characteristics of students we consider. The set of variables to include in the analysis becomes then a critical choice, and we discuss the issue in detail below.

3.1 Source of data

The source of data is the 2012 PISA, an international assessment of the Math, Reading and Financial Competences of 16-year old students. We make use of that particular assessment as it took place right before the introduction of "Finanzas para Todos", thus providing a representative sample of the financial competences of Spanish youths right before the introduction of the program.

PISA contains a comprehensive set of background characteristics both at the student and

[6] An federation of Catholic schools played an important role in advertising the program. For that reason, we construct an indicator of religious school in PISA and include it as a determinant of participation -basically all religious schools in Spain are Catholic.

school level. The school level part is mainly filled by the principal, whereas students answer questions about their family background. We conduct the analysis at the student level, including school characteristics as well. The main reason for the choice is that some of the estimators, like the differences-in-differences, exploit variation in grades within-student and across assessments.

The choice of variables also deserves discussion. We first consider a set of *student-level characteristics* that are commonly collected in other settings. These include the gender of the student, parental employment status, as well as an indicator of having repeated a grade. In addition, we include the geographical location of the school.

Secondly, we consider *school-level characteristics* as reported by the principal. These are meant to collect student admission policy, a likely source of selection bias. Schools that choose to participate in the program may have a different set of students because they are more selective. The PISA assessment asks principals to rate if they admit students based on their location -in principle, all publicly funded schools have a catchment area where students have priority to be admitted. Schools may also select students by encouraging low-performing students to move to other centers. We use then information on whether the school transfers students to other schools for bad behavior and whether they perceive competing with other schools in the local area.[7]

Finally, two student-level variables that play an important role in the analysis are the measurement of competences in the Mathematics and on the Reading parts of the assessment. Neither are affected directly by the contents of the financial literacy course, and they serve to measure general competences of the students, thus helping in assessing their general level of ability.

3.2 Methods considered

As it is often the case in interventions that examine the impact of financial literacy courses, our sample size does not allow comparisons between schools/students in the treated and control groups with *exactly* the same characteristics -i.e. we cannot perform exact matching. Hence, we match treated students to controls on the basis of the predicted probability of participating in the program.

3.2.1 Reweighting based on the propensity score

The first method eliminates differences between students in schools "to-be-treated" and controls by constructing a sample of controls with similar average characteristics X. Reweighting methods provide unbiased estimates of the average effect of treatment on the treated if for a set

[7] However, there are many other variables in PISA that may affect both participation into the program and financial competences. To give all variables a chance of being included as relevant determinants while not running into colinearity problems, we implemented LASSO methods where we test the influence of the "core" variables identified above against the rest of variables considered in PISA -school resources, management practices as well as parental possessions.

of characteristics X the propensity score is such that $0 < p(X) < 1$ for treated and controls and if, conditional on characteristics X, treatment is independent of the distribution of potential outcomes (*unconfoundedness*) -see Busso et al (2014) or Rosenbaum and Rubin (1983).[8] We test both assumptions below.

We implement reweighting estimators in three steps. The first step consists in estimating the propensity score $p(X)$ as the adjusted probability of participating in the program depending on characteristics, X, in which students in "to-be-treated" and "control" schools differ. The second step involves weighting the students in the control group in such a way that their observable weighted characteristics coincide with those of the treated group. The weights used are $\hat{\omega} = \frac{p(X)}{1-p(X)} \frac{1-\pi}{\pi}$,where $p(X)$ is the propensity score estimated in the first step and π is the unconditional proportion of treated students. In the third step, the (nonexistent) effect of participating in the BdE-CNMV program for the students in the "to-be-treated" schools or selection bias (B) – is constructed as the difference between the average outcomes of "treated" students Y_t (scores in the Financial Competences part of the PISA assessment) and weighted average outcomes of the "control" group – weighted with the $\hat{\omega}$ weights. The indicator of treatment is $D_{t+1} = 1$, where the subscript $t + 1$ indicates that treatment will be received in the near future

$$B(X) = E(Y_t|X_t, D_{t+1} = 1) - E(\hat{\omega}Y_t|X_t, D_{t+1} = 0) \tag{1}$$

$B(X)$ can be estimated via a linear regression model where the treated observations receive a weight of 1 and controls a weight $\hat{\omega}$. If the model is well specified, the estimate of $B(X)$ should not vary if the regression also includes the characteristics vector X as additional regressors. Selection bias $B(X)$ is zero if, for a given X, the average performance in the Financial Competences part of the PISA assessment of students in the "to-be-treated" group does not differ significantly from that of the reweighted average of the controls.[9]

We test for unconfoundedness by examining if students in "to-be-treated" schools differ from controls in Mathematics and Reading scores. The reason is that both measure students' competences distinct from those affected by a Financial Literacy course so one would not expect differences in that dimension. A nonzero estimate of $B(X)$ when we use Math or Reading scores as the dependent variable in (1) implies that similar X are not enough to obtain a group of

[8] We see two main advantages of using reweighting. The first is that compared to other matching estimators, reweighting can be implemented via weighted OLS, thus allowing simple comparisons to unweighted OLS -a popular estimator in this literature. In addition, reweighting provides a semiparametric method to estimate , so we do not have to search for the best functional form linking X and the score in the Financial part of the assessment. Busso et al (2014) show very good performance of reweighting vs other matching estimators, drawing on the asymptotic results of Hirano et al (2003)

[9] PISA gives five possible values of Financial literacy, Mathematics and Reading outcomes to each student that responds that part of the questionnaire. All Tables below show estimates for each possible value and combine the five estimates. Standard errors are adjusted for this multiple imputation procedure. The propensity score is estimated using a Logit model. All models below use population weights.

control students that is comparable enough to treated ones, pointing at the existence of selection bias.

3.2.2 Differences-in-differences

A common way to infer the impact of one intervention in schools that are domain-specific (i.e., affect a particular subject but not others) is to difference out a component common across subjects by examining if a course in financial literacy (say) increases financial competences more than it does other subjects (say, Mathematics or Reading). The implicit assumption in these models is that in the absence of treatment, the gap between Finance and Math scores would be similar among treated and control students. A further assumption is that in the absence of treatment, the gap between Finance and any other competence would be of similar size -see Lavy (2015), for a similar test.

To fix ideas, consider the following model, where Y_t^F is the score in the Financial competences part and Y_t^i, are the scores in the Mathematics ($i = M$) and Reading part of the assessment ($i = R$). Differences-in-differences estimators typically assume that finance-specific interventions affect only financial scores. That is,

$$E(Y_t^F|D_{t+1}) = \beta_0 + \beta_1 D_{t+1}$$

$$E(Y_t^i|D_{t+1}) = \beta_0 \qquad i = M, R$$

Differencing out both expressions for the same student one gets the Diff-in-Diffs estimate

$$E(Y_t^F - Y_t^i|D_{t+1} = 1) - E(Y_t^F - Y_t^i|D_{t+1} = 0) = \beta_1 \qquad i = M, R \tag{2}$$

Where $Y_t^F - Y_t^i$ is the within-student difference between the normalized score in Financial Competences and in assessment i (Math or Reading).[10] An advantage of differences-in-differences estimators is that they hold constant any individual fixed effect that is constant across assessments. Furthermore, Expression (2) suggest a test of the D-in-D model:

$$E(Y_t^F - Y_t^M|D_{t+1} = 1) - E(Y_t^F - Y_t^M|D_{t+1} = 0) = E(Y_t^F - Y_t^R|D_{t+1} = 1) - E(Y_t^F - Y_t^R|D_{t+1} = 0) = 0 \tag{3}$$

The hypotheses in (3) test if in the absence of a Financial Literacy intervention the gap between Finance and Math or Finance and Reading are equal to each other and both equal to zero. If the hypothesis of equal gaps were rejected in an observational setting, there would be multiple estimates of the impact of the program on financial literacy (as many as different assessment are used in the benchmark), suggesting that the true effect is not identified.

[10] We assume that β_0 is constant across assessments because researchers typically normalize scores. We work with mean zero variables throughout.

4 The sample

We use a sample of students who took the PISA financial literacy test in Spanish schools that had not offered Financial Education up until the academic year 2011-2012 -according to the principal. This sample contains 912 students from 150 schools in Spain. In each of these schools around 6 students per center took the PISA financial literacy test. In the subsequent academic years between 2012-2013 and 2015-2016, 30 schools requested to participate in the BdE-CNMV program, while the rest did not. As financial literacy courses were very atypical at the time in the Spanish educational system differences in financial competences between students in those 30 schools and the rest could only be due to underlying characteristics that led schools to participate in the program. The rest of schools either did not know about the program or were not interested in delivering such material. This sample, therefore, allows for characterizing the differences in financial literacy among students from schools that volunteered for the program and those in schools that did not. Furthermore, as PISA is representative of the universe of Spanish schools, we can understand the determinants of program participation in a well defined population -that would not be the case if financial literacy courses were widespread, and only a selected sample of schools had not participated yet in any personal finance program for students.

4.1 Are the schools that volunteered for the BdE-CNMV program different from the rest?

Table 1 compares the characteristics of the students. In comparing those schools that volunteered and not, we focus on three types of factors. First, we examine the geographical location as well as the institutional form of the school -public or private. Secondly, and given the possible differences in admissions criteria, we examine the student body selection practices, according to the school administration. Finally, we look at both the family environment and some performance measures - for example, if the student has been held back a year. The schools that showed an interest in participating in the program between 2012-2013 and 2015-2016 differ from the rest of institutions in all the mentioned factors.

The most striking difference corresponds to geographic distribution: 56% of students in schools that subsequently applied to offer the program are located in Andalusia, Ceuta, Melilla, Murcia and the Canary Islands (we refer to this group of regions as "South"). Among the schools that did not volunteer for the program, the percentage of students in the South is only 22%. This is important because students in the South tend to obtain lower scores in PISA assessments. Secondly, around 38% of the students in the schools that asked to offer the program were either non-public (i.e., either privately owned schools or privately managed), while among the institutions that did not request to participate in the program, the proportion of private or state-subsidized institutions was 22%. Finally, for the schools that participated in the program between 2012 and 2016, the proportion of students held back is 3 percentage points lower than among the rest of schools.

The information that the school administration provides illustrates that schools that eventually participate in the program are more selective than the rest. Firstly, applicant schools are exposed to greater competition with other schools in the district and use different admissions criteria: 33% never select their students based on the student's residence criterion, while in the rest of the schools this percentage is only 19%.[11] Secondly, 37% of the treated students attend schools that can transfer students with behavior problems to other schools, while the percentage in the rest of the institutions is 23%.

Students in volunteer schools obtained 7.3 percent of one standard deviation *lower* scores in financial competences than those in schools that did not apply for the program. The magnitude of the difference is not small, as typical estimates of the impact of financial literacy programs in financial knowledge are about 14-20% of one standard deviation. However, as we discuss below, it is also imprecise. The differences in math scores between students in volunteer and non-volunteer schools are negligible (about 3 percent of one standard deviation). Finally, students from the schools applying for the program obtained some 5 percent of one standard deviation higher scores in reading than those that never did. In other words, despite the strong differences in school type, location and school admission procedures, the degree of financial literacy of the student body is, at first glance, not statistically different from that of the other schools. However, selection bias is only well defined among groups that share similar characteristics, and clearly both types of schools differ (Heckman et al, 1998).

We also note that once we make treated-control comparisons by gender in Table A1, "treated" females perform 11 percent of one standard deviation worse in the financial competences test than girls in the control group (see Table A1). We come back to those differences in Section 6.

Summarizing, schools participating into the program appear to be more selective in their admission criteria than the rest of schools. A third of treated schools select students using criteria other than the place of residence, select their current student body (they are more likely than controls to transfer students behaving poorly to other schools) and basically all compete with other schools. On the other hand, they are located in regions with lower PISA scores in PISA. Each set of characteristics leads to different bias: more selective practices may result in higher scores, while location may result in lower scores. To estimate selection bias we need to hold those factors constant. The remainder of the paper examines which estimation procedures might eliminate the influence of selection bias in the relation between FE courses and financial literacy.

[11] Public schools and state-subsidized schools have an assigned area of influence in which residents are given preference for entering the school, while private schools do not have this preference. Still, some public or state-subsidized schools indicate having the capability of selecting their students regardless of their place of residence.

5 Estimates of the selection bias

The estimation of the selection bias involves three steps. The first consists of predicting the variable "request to teach Financial Education in the third year of secondary education" using characteristics observed for students within each type of school. Secondly, that prediction is used to construct a weighted average for students in the control group. Finally, the average financial literacy of "treated" students is compared with the weighted average of students from the "control group" using various estimates.

5.1 Determinants of participation in the program

We examine how successful are three sets of covariates in generating a sample of controls that is similar to treated students. First, a limited set of observed variables is used, usually available in other studies: geographical location of the school, student's gender, parents' employment status and whether the student has been held back (model 1). Given the evidence in Table A1, interactions of each of these variables with the student's gender are included. The second information set uses to a greater extent the information provided by school administrators (model 2). In this second model, we include variables that reflect the type of school: if it is a religious school, two variables that indicate the selection criteria of the student body, and a variable that normally is not observable such as "teacher morale" according to the schools' principals. The variable "religious school" is included because the BdE-CNMV program was publicized widely among Catholic schools. The rest of the variables are included for the purpose of comparing schools with similar admissions criteria.

In both cases (models 1 and 2), we exclude two important measures of student's competences: their mathematical and reading competences as measured in the PISA test. The exclusion serves as a validation check. Imagine that we find a sample of controls similar to treatment along the set of covariates in model 1 or 2. If balancing on those covariates we find that students in treatment groups have similar competences on average in mathematics or in reading, we can be confident that differences in financial competences of those similar groups truly reflects selection bias. Otherwise, students may differ in important competences, and the remaining differences may not reflect selection bias.[12]

Finally, the third set of covariates adds the score in the Mathematics part of the test. By choosing controls that are similar to treated students in their competences in Math, both groups become comparable. However, by choosing that variable to match students, we also lose the possibility of testing for unconfoundedness.

Using the set of covariates I, Table 2 documents that location is an important determinant of participation, but demographic variables -like employment status or grade repetition- play

[12] Heckman et al (1998) make the point that selection bias is only well defined when one compares observably similar individuals. In his practitioner's guide on matching. Imbens (2014) suggests testing the plausibility of the unconfoundedness assumption in matching by examining if treated and control students are similar in a variable closely related to the outcome of interest but unlikely affected by the intervention.

a lesser role. Actually, when we use the predicted score to reweight the sample of controls, the samples still look very different in terms of the school's admissions criteria -see Table 1, in particular the comparison between rows 6, 7 and 8 in column 1 (treated) and 3 (controls reweighted by location, parent employment status and grade repetition).

Model II in Table 2 shows how the school's admission criteria correlate with the probability of subsequently participating in the program. As the results in Table 1 already suggested, participant schools are more likely to compete with other schools in the area, select their students with criteria other than residence, and to transfer students with bad behavior. Interestingly, once we reweight the sample of controls so that the mean admission criteria are similar to treated students, controls have a score in maths 15 percent of a standard deviation below treated. That is, treated students obtain a higher Math scores in the PISA assessment than controls when we compare them to controls in schools that are similarly selective. Model II in Table 3 confirms the result: students in schools participating in the program have higher scores in the Math assessment than controls.

5.1.1 The distribution of the predicted probability of participation

We assess if treated and control students share similar characteristics by comparing the distribution of their probability of participation in the Finance program as predicted by the Set of covariates I, II and III. The propensity score of students in treated schools is shown in a solid line and that of controls as a dotted line. When we use the distribution of the propensity score in Figure 1, we note that both predicted probabilities are concentrated around values between 0 and 0.4, albeit with very different shapes, with scarce concentration in zero. As the set of discrete covariates, both propensity scores take only a few values. In principle, the Set of Covariates I achieves a distribution where "controls" can be found that are similar to "treated" in terms of their propensity score. However, as shown in Table 1, controlling only for location and parental background is clearly insufficient to achieve good balancing the admission criteria of schools.

Figure 2 shows the distribution of the probability of participating in the program when we include as covariates school admission and turnover characteristics. The support of the distribution ranges now between 0 and 0.8, being spread than that in Figure 1. Still, the support of the distribution of treated and control students overlaps, suggesting that we can find control students that are similar to the treated.[13]

5.2 Estimating selection bias: reweighting estimators.

We now turn to a comparison between the average financial performance of the "treated" group and that of the "control" sample but once the sample of controls is reweighted so that both

[13] Figure 2 shows that there is some mass of treated students with propensity scores between .6 and .8, while that mass is thin for controls. We experimented trimming the sample of treated above .6, without a noticeable impact on the results.

groups share similar average characteristics.

We start with the Set of Covariates I: grade repetition rate, geographic location and family environment. As it was shown in Table 1, the unconditional difference between the scores in financial competences in the schools that subsequently volunteered for the program and those that did not was minus 7.3 percent of one standard deviation (standard error: 8.7 percent, see Table 3, row 1 column 1). Once we reweight the sample of non-volunteer schools so that their sample means are similar to "treated" in terms of school location, student's parental labor status and grade repetition, the 7 percent difference falls to minus 3.2 percent, suggesting that location, grade retention and parental background play a role in explaining differences.

However, those basic covariates are not sufficient to make the samples comparable in terms of students' Math competences. Once we compare control students with location, retention rate and parental background similar to that of treated students, we observe that the latter are more proficient in Math, obtaining scores that are one-tenth of one standard deviation higher (see Table 3, Panel B, row 2, columns 2 and 3).

Controlling for location, parental labor status, school admission policy A second set of Covariates (the Set II), considers control students that, on top of location, grade repetition and parental background attend schools with admission policies similar to those schools that end up applying for the program. When we reweight the sample of controls to achieve similar samples along those dimensions, we observe that "treated" students obtain between 14 and 19 percent of one standard deviation higher Math scores than controls -see Table 3, Panel B, row 2, columns 4 and 5. The lack of comparability in the Math competences between treated and controls suggest that both samples remain too different even after controlling for school admission criteria.

Controlling for location, parental labor status, school admission policy and Math scores. To further balance the sample, we present the results of the third model of selection (Set of Covariates III). There, we reweight the sample of controls using a propensity score that also includes the score in the Mathematics part of the assessment -model III in Table 2. Once we compare treated students to controls with similar Math grades, we find that their financial competences are 12% of one standard deviation *below* those of controls (standard error: 7%) -see Table 3, Panel A, row 1, column 7. That is, schools that end up applying for the BdE-CNMV program have students that obtain lower scores in the financial competences part of the assessment than similar students in control schools. One possibility is that somehow those schools are aware of the comparatively lower competences in the financial domain. Alternatively, it can be the case that students in schools applying for the program perform worse in competences other than Math than similar students in the rest of schools. The availability of the reading score allows us to test for that possibility. "Treated" students obtain 8% of a standard deviation lower scores in reading than students in similar schools and equivalent math competences -see Table

3, Panel B, row 3, column 7. The estimate is not very precise -the standard error is 7.3 percent of one standard deviation. In any case, it is somewhat smaller in magnitude to the 12% gap in financial knowledge documented for financial competences. Thus, we interpret that students in volunteer schools perform worse than controls not only in financial competences, but also in reading.

As a way of summary, reweighting or matching methods implicitly assume that if treated and controls are similar in their observable characteristics, they will also be similar in their unobservables. In this sample and with the set of covariates we tried, the assumption is not fulfilled for two reasons. Firstly, in our setting, there is no exposition to any course whatsoever, so if we compare treated students to similar controls, we should find no differences in either financial or in other type of competences. However, for the various measures of covariates we try, we find statistically significant differences in Mathematics. When we make treated and control students similar in their Math scores, we find that treated students' financial competences or their Reading scores are systematically lower than those of controls.[14]

5.3 Estimating selection bias: Differences-in-differences (D-in-D).

Table 4 shows Differences-in-differences estimates of selection bias. Applied to cross-sectional data, the implicit assumption behind D-in-D estimator is that selection bias is constant across different competences (say, Reading, Math and Finance). For example, assume that we are evaluating an intervention that seeks to boost financial knowledge in an observational setting. Even if treated students differ from controls in their Math competences, as long as math is not affected by the intervention, one can difference out the bias by examining if the intervention affects the Finance-Math gap differently among treated and controls. With three different measures, the assumption can be further tested by examining if the intervention affects similarly the gap between Finance and Math and that between Finance and Reading. Those are the tests we implement in Table 4.

Firstly, we compute for each student the difference in the normalized finance score and math score, on one hand, and between the normalized finance and reading scores, on the other. The normalized finance-math difference measures if the student performed comparatively better in the Finance than in the Math part of the assessment. The first column in Table 4 compares measures of relative performance between students in schools that applied for the BdE-CNMV program and those that did not. Those differences are unadjusted by differences in relevant covariates and imply that students in treated schools performed 7.6 percent of one standard deviation worse in the Financial part of the test than in the Math part -see Table 4, row 1, column 1. Interestingly, students in "to-be-treated" schools also performed 12 percent of one standard deviation worse in the Financial part of the assessment than in the Reading part -see Table 4, row 2, column 1. Both are large differences in terms of magnitude, even though the

[14] Of course, it can also be the case that we are not making full use of the information in PISA, and failing to find the right covariates. Section 6 discusses the issue.

precision is not large enough to distinguish them from zero at usual confidence levels.

Controlling for location, parental labor status In columns (2)-(4) of Table 4 we control for the differences between students and schools detected in Table 1. We introduce covariates in two different ways. Column (2) includes the set of covariates I in the Diffs-in-diffs part (column 2) while in Column 3 we reweight the sample using the propensity score shown in column 1 of Table 2. Students in "to-be-treated" schools have a gap between Finance and Math scores that 9 percent lower than among controls (Table 4, row 1, column 2). If instead of controlling for differences via regressions, we reweight the sample of controls so that they have similar location, parental employment status and repetition rates, the deficit in Finance vs Math scores increases to 13.6 percent of one standard deviation (Table 4, row 1, column 4). Secondly, among students in treated schools, the gap between the Financial and the Reading part of the assessment is 7 percent of one standard deviation. Both are negative estimates and large in absolute magnitude.

Controlling for location, parental labor status and school admission policy Columns (5)-(7) in Table 4 compare students in "to-be-treated" schools with controls in schools with similar admission criteria. Column (5) does this via regression methods and introducing the Set of Covariates I, while Columns (6) and (7) reweight the sample of controls using the propensity score estimated in Table 2, column 2. Using the set II of covariates, the gap between the normalized Finance and Math assessments is 16.9 percent lower among students in "to-be-treated" schools than among those in control schools (standard error: 8 percent) -see Table 4, row 1, column 7. The corresponding gap between normalized Finance and Reading scores is 4.3 percent, four times smaller.

The results in columns (6) and (7) of Table 4 are consistent with those in Table 3. There, we showed that, once we control for differences in schools admission criteria, students in "to-be-treated" schools score relatively worse in the Finance than in the Maths part of the PISA assessment. Table 4 shows that the worse performance in the Finance part than in the Math part of the assessment holds even when we remove an individual fixed-effect. Underperformance in the Financial part relative to the Reading part is less obvious once we adjust for school admission criteria. Hence, one solution to the relative underperformance in Finance (vs Math) among students in "to-be-treated" schools is to compare them to students in control schools with a similar performance in Math -on top of the set of covariates I and II.

Controlling for location, parental labor status, school admission policy and Math scores Columns 8-10 in Table 4 conducts a Difference-in-Difference analysis where we compare the relative scores in Finance vs Reading between students in "to-be-treated" schools and controls with similar Math competences. Interestingly, the estimates are close to zero, although not very precisely estimated. When we estimate a simple Differences-in-Differences model between Finance score and the Reading score controlling for performance in the Math part and the covariates in Set I and II, the estimate in Table 4, row 2, column 8 is 0.7 percent of one standard deviation, a very small number. When we estimate instead the Difference-in-Difference model in a sample that reweights the sample of control students so that their average location, parental background, grade retention, school admissions and performance in Math is similar to that of "to-be-treated", the estimate is -3.7 percent of one standard deviation, very close to zero.

As a way of summary, Difference-in-difference estimates of the impact of an intervention assume that the gap between the competences directly affected and any other competence would be the same in the absence of the program. We use scores in Finance, Math and Reading in the PISA assessment to test if students in schools applying for a financial literacy course (but not delivering it yet) have similar gaps between Finance and Math and Finance and Reading. The estimates in Table 4 show that shortly pre-intervention, students in "to-be-treated" schools scored worse in Finance than in Math, while that did not happen among comparable students in control schools. The corresponding gap between the competences in Finance and in Reading was between 10% and 25% as large, however. Both numbers should be similar to support a D-in-D specification. Furthermore, those differences happen *prior* to delivering any Financial Literacy course that may have affected Mathematics and Reading differently. Having said that, one advantage of D-in-D is that it is very easy to detect (and possibly correct) selection biases by testing if treated and controls have similar gaps in Math vs Finance than in Reading vs Finance.

5.3.1 Summary of the results.

Schools that apply to deliver the BdE-CNMV program are more selective centers in regions that perform worse in PISA. Both characteristics generate complex differences in their student's scores in Finance, Mathematics Reading parts of the assessment. While unconditional differences in different competences are imprecise, once we compare students in "to-be-treated" schools to students in control schools with similar location and admission policies, a consistent pattern emerges. Students in those "to-be-treated" schools outperform control students in Math but less so in Finance or Reading. That finding arises both when we make comparisons across students (Table 3) or when we compare the performance of the same student across assessment (Table 4). The differential performance in Math cast doubt on whether both sets of students are really comparable. A combination of (1) holding constant a student fixed-effect across Finance and Reading while comparing to controls students with similar Math scores makes the students ex-ante comparable. The next Section explores the robustness of those results.

6 Heterogeneity analysis.

Surveys around the world indicate that females are less likely to answer correctly basic questions on financial literacy than males -see Lusardi and Mitchell, 2014. That evidence motivates analysis of gender-specific measures of the effectiveness of financial literacy programs. It seems natural then to test if the differences in observed and unobservable variables that Tables 3 and 4 uncover are specially salient for a particular gender. Were that the case, gender-specific impacts of financial literacy program could be the result of different selection criteria.

Tables 5 and 6 re-estimate the models in Tables 3 and 4, now distinguishing by gender. The overall patterns in Tables 3 and 4 can be observed for females mostly. In Tables 3 and 4 we found that students in "to be treated schools" underperformed controls in schools with similar admission policy and location in the Finance part of the assessment -relative to Math. Two pieces of evidence support that interpretation. Firstly, holding Math scores constant, females in "to-be-treated" schools attain 15 percent of a standard deviation lower scores in Finance than females in similar control schools (standard error: 9.4% of one standard deviation). That result is shown in Table 5a, row 1, column 5. Secondly, the within-student gap between the score in Finance and that in Math among female students in "to-be-treated" schools is 19 percent of one standard deviation lower than in control schools -see Table 6a, row 1, column 4.

The results among males are less precisely estimated and we do not comment them in detail.

6.1 Are we controlling for the right determinants of participation?

The choice of the variables included in the propensity score is crucial to properly reweight the sample of controls so that they are comparable to treated students. In Table 2, we chose the covariates to be included based on the observation of differences in school location and stated student admission criteria. However, these may not be the right controls. To see the robustness of the results, we experimented adding to Model 2 a broad set of variables in PISA, including parental wealth (as estimated by PISA data producers) and a wide measure of school resources. To decide among the relevant variables, we estimated a (linear) LASSO model to detect the main determinants of participation and reestimated a Logit model with the selected covariates a model of the probability of participating in the program. The variables favored by the model were, apart from the school admission criteria, those indicating school resources. The results are shown in Tables 7 and 8. We summarize two key results. The first is that students in "to-be-treated" schools underperform in Finance or Math relative to controls in schools with similar admission policy and resources by 10 percent of one standard deviation (Table 7, column 3, rows 1 and 2). However, they underperform controls in Reading by 27 percent of one standard deviation (Table 7, column 3, row 3). As a result, the new set of controls does not makes the students comparable either.

However, there are a number of issues with this model. Firstly, the overlap of the model is not very good, with treated students being too different from controls. Secondly, the results in the reweighting method are very sensitive to the inclusion of covariates, indicating misspecification.

While we prefer the results in Tables 3 and 4, we show this results for the sake of completeness. At any rate, the new results confirm that students in "to-be-treated" schools underperform in Finance once compared to controls in similarly selective schools.

7 Conclusions

A number of studies have analyzed if financial literacy training in high school has long lasting effects on financial decisions over the life cycle. Bernheim (2001) documented that individuals who studied in US states that had in the curriculum financial education had accumulated higher wealth levels than the rest. Urban et al. (2018) and Brown et al (2013) find that individuals who underwent financial education are less likely to hold debt or hold lower debt-income ratios. Both studies suggest that the financial content of high school curriculum shapes portfolio decisions, presumably by increasing financial knowledge of the students.[15] However, empirical studies on how specific financial literacy programs increase students' financial knowledge delivers mixed results (see Becchetti et al. 2011, or Lührmann et al. 2012). Those studies measure the evolution of financial knowledge by delivering the same questions before and after the course. Evidence from a large program in Brazil in Bruhn et al. (2013) yield stronger evidence, as the financial literacy course studied there increased financial literacy scores by 20% of one standard deviation, measured by different sets of questions delivered before and after the course.[16] There are many explanations behind the heterogeneity in results. This paper stresses that the schools that select into participation have different characteristics at baseline, and that the reasons that lead schools to participate in those programs may correlate with their students' financial knowledge.

Our study provides an assessment of the extent of (and possible solutions to) the problem of selection bias. We do this using the 2012 PISA test on financial competence, where we can identify the set of schools that took part in a program to promote Financial Literacy between the academic years of 2012-2013 and 2015-2016 and compare the students' performance in Financial Competence to that in the rest of the schools. As the financial literacy course only took place at least one year after the PISA exam, any difference in performance must be due to a selection bias.

We document that participant schools adopt more selective admission policies than the rest of schools but also that they are located in areas where students obtain lower PISA scores on average. Both factors compensate each other, so unadjusted differences in scores in Financial Competence assessments are small. However, once we adjust for school admission criteria and school location we find that students in schools that eventually participate in the program underperform in the Finance part of the test, relative to Math and, more tentatively, in the Reading part of the assessment. Those patterns are most precisely estimated for females. We eliminate the bias using Differences-in-Difference estimators exploiting the availability of three measures of student's competences..

[15] Alan and Ertac (forthcoming), Lührmann et al (2018) and Bover et al (2018) evaluate of programs that deliver personal finance contents in primary and high school and achieve changes in preferences and decisions.

[16] However, that course was unusually intensive: 78 hours spread over one year and a half. The change in the curriculum following such intervention is by no means trivial. Walstad et tal (2010) report positive impacts of a much lighter intervention.

For future studies, our results highlight that data on school admission policies and student retention are very important to understand to what extent the results can be extrapolated to other settings. Furthermore, having information on a wide set of student competences proves useful in detecting selection biases. Finally, subgroup analysis of the impact of an intervention must be interpreted with caution, as selection patterns may vary along gender lines. In sum, the analysis underscores the relevance of complementing information on interventions with information on school practices.

References

[1] Alan S, and S. Ertac (forthcoming) Fostering Patience in the Classroom: Results from a Randomized Educational Intervention" Journal of Political Economy.

[2] Alessie, R., M. Van Rooij, and A. Lusardi. 2011. Financial Literacy, Retirement Preparation and Pension Expectations in the Netherlands. Journal of Pension Economics and Finance, 10 (4): 527-45.

[3] Becchetti, L., S. Caiazza, and D. Coviello. 2011. Financial education and investment attitudes in high schools: evidence from a randomized experiment.CEIS Working Paper No. 210.

[4] Bernheim, D., D. Garrett, and D. Maki. 2001. Education and Saving: The Long-term Effects of High School Financial Curriculum Mandates. Journal of Public Economics 80: 435-565.

[5] Bover, O., L. Hospido and E. Villanueva (2018) The Impact of High School Financial Education on Financial Knowledge and Choices: Evidence from a Randomized Trial in Spain CEPR DP12632

[6] Brown, M, W. van der Klaauw, J. Wen, and B. Zafar. 2013. Financial Education and the Debt Behavior of the Young. Federal Reserve Bank of New York, Staff report, Number 634.

[7] Bruhn, M., L. de Souza Leao, A. Legovini, R. Marchetti, and B. Zia. 2013. The Impact of High School Financial Education Experimental Evidence from Brazil. The World Bank Development Research Group & Latin America and Caribbean Region. Policy Research Working Paper 6723.

[8] Bucher-Koenen, T., and A. Lusardi. 2011. Financial Literacy and Retirement Planning in Germany. Journal of Pension Economics and Finance, 10 (4): 565-84.

[9] Busso, M., J. DiNardo, and J. McCrary. 2014. New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators. Review of Economics and Statistics, December, 96(5): 885–89

[10] Fernandes, D., J. Lynch and R. Netemeyer (2014) Financial Literacy, Financial Education and Downstream Financial Behaviors Management Science Vol 60 No. 8

[11] Fornero, E. and C. Monticone. 2011. Financial Literacy and Pension Plan Participation in Italy. Journal of Pension Economics and Finance, 10 (4): 547-564.

[12] Heckman, J., H. Ichimura, J. Smith and P. Todd. 1998. Characterizing Selection Bias Using Experimental Data. Econometrica, 66(5): 1017-1098.

[13] Hirano, K., G. Imbens, and G. Ridder. 2003. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. Econometrica, 71(4): 1161-1189.

[14] Imbens, G. W. 2014. Matching Methods in Practice: Three Examples. IZA DP No. 8049.

[15] Lavy, V. (2015) Do Differences in School's Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries Economic Journal 125: F397-F424.

[16] Lührmann, M., M. Serra-Garcia, and J. Winter. 2012. The effects of financial literacy training: Evidence from a field experiment with German high-school children. Discussion Papers in Economics 14101, University of Munich, Department of Economics.

[17] Lührmann, M., M. Serra-Garcia, and J. Winter. 2018. The Impact of Financial Education on Adolescents' Intertemporal Choices American Economic Journal: Economic Policy 10(3), 309-332

[18] Lusardi, A., and O. S. Mitchell. 2009. How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy and Retirement Readiness. NBER Working Paper 15350.

[19] Lusardi, A. and O. S. Mitchell. 2011. Financial Literacy and Retirement Planning in the United States. Journal of Pension Economics and Finance 10(4): 509-525.

[20] Lusardi, A. and O. S. Mitchell. 2014. The Economic Importance of Financial Literacy: Theory and Evidence. Journal of Economic Literature, 52 (1): 5-44.

[21] Lusardi, A., O. S. Mitchell, and V. Curto. 2010. Financial Literacy Among the Young. Journal of Consumer Affairs, 44 (2): 358-380.

[22] Mandell, L. 2008. Financial Education in High School. In Overcoming the Saving Slump: How to Increase the Effectiveness of Financial Education and Saving Programs, Annamaria Lusardi Editor, Chicago and London: University of Chicago Press, 257-279.

[23] Rosenbaum, P. and D. B. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70 (1): 4155.

[24] Urban, C., Schmeiser, M., Collins, J.M., and Brown, A. (2018) The effects of high school personal financial education policies on financial behavior Economics of Education Review (forthcoming)

[25] Van Rooij, M., A. Lusardi, and R. Alessie. 2011. Financial Literacy and Stock Market Participation. Journal of Financial Economics 101(2): 449-472.

[26] Walstad, W., K. Rebeck, and R. MacDonald. 2010. The Effects of Financial Education on the Financial Knowledge of High School Students. Journal of Consumer Affairs 44(2): 336-357.

8 Topics covered in the program

The financial literacy course taught in 2012-2013 was structured in nine chapters. There was a tenth one that reviewed the previous material.

Financial security Learning to manage money. Establishing savings goals and how to reach them

Intelligent consumption How to improve quality of life by expenditure choices. The definition of healthy, responsible and sustainable consumption. The role of advertising.

Saving Definition of saving. Reasons for saving and the role of uncertainty. Differences between saving and investing

Personal budgeting Elaboration of a personal budget. Identifying fixed and variable expenses. Identifying adjustable expenses.

Money and monetary transactions Advantages and disadvantages of using cash. Identification of fake currency

Bank accounts Differences between checking and savings accounts. Opening a bank account. Consequences of an overdraft. Understanding checking account statements. Managing bank accounts. Internet and phone banking.

Credit and Debit Cards The use of credit and debit cards: advantages and disadvantages. The use of tellers and cash stations.

Data protection Problems using Internet banking.

Banking relationships Opening and closing bank accounts. Different forms of bank fees.

Table 1: Differences between participants in the Financial Education program and the rest of schools

	(1)	(2)	(3)	(4)	(5)
	Unadjusted means*		Reweighted means of controls		
	"Treated"	Controls	Set I **	Set II***	Set III****
SCHOOL CHARACTERISTICS					
1. South	0.56	0.22	**0.56**	**0.57**	**0.57**
2. Public school	0.62	0.68	0.68	0.62	0.61
3. State-subsidized school	0.18	0.13	0.13	**0.17**	**0.17**
4. Private school	0.20	0.19	0.19	**0.22**	**0.22**
5. Religious school	0.28	0.20	0.21	**0.26**	**0.27**
6. Does not admit students according to residence	0.33	0.19	0.18	**0.31**	**0.31**
7. Transfers students to other schools if bad behaviour	0.37	0.23	0.26	**0.36**	**0.35**
8. Competes with at least one school in the area	0.93	0.81	0.75	**0.93**	**0.93**
9. High teacher morale	0.19	0.13	0.13	**0.16**	**0.15**
10. Percent parents who discuss with teacher child behavior at own initiative	44.70	28.71	31.03	34.47	34.09
11. Principal refers to the school's academic goals when making curricular decisions with teachers (from 1 to 6, 6 most frequent)	3.08	2.59	2.66	2.68	2.66
12. Quality of school educational resources	0.42	-0.07	-0.16	-0.11	-0.07
STUDENT CHARACTERISTICS					
13. Normalized score, Financial Competences	-0.06	0.01	-0.03	-0.08	0.03
14. Normalized score, Maths	0.02	0.02	-0.08	-0.17	**-0.01**
15. Normalized score, Reading	0.03	-0.02	-0.01	-0.02	0.08
Girl	0.47	0.46	**0.47**	**0.47**	**0.48**
Has repeated grade	0.31	0.34	**0.31**	**0.33**	**0.32**
Father works	0.79	0.79	**0.79**	**0.80**	**0.80**
Father with college degree	0.32	0.35	0.30	0.34	0.35
Mother works	0.61	0.66	0.58	**0.58**	**0.58**
Mother with college degree	0.36	0.36	0.33	0.32	0.33
South x Girl	0.27	0.10	**0.27**	**0.28**	**0.28**
South x Has repeated grade	0.19	0.10	**0.19**	**0.22**	**0.21**
South x Girl x Has repeated grade	0.07	0.03	**0.07**	**0.07**	**0.07**
Family Structure: Single parent	0.13	0.11	0.10	0.13	0.14
Immigrant	0.10	0.12	0.08	0.09	0.09
Cultural Possessions (family owns works of art, poetry)	0.13	0.09	0.07	0.02	0.04
Wealth (indicator of having own room, desktop, etc)	0.08	-0.08	-0.06	-0.12	-0.11
ICT Availability at Home	0.12	0.14	0.10	0.08	0.09
Time of computer use (mins)	248	589	504	806	675
Number of students	210		702		
(schools)	30		120		

*Averages weighted by the population weights. Cells in bold in Columns 3-5 indicate the variables are included in each propensity score. Grades combine the 5 different scores provided.

** The column Set I show propensity-score reweighted means of controls when the propensity score includes school location, the parents' employment status, and the student's gender, repetition status, all interacted with female.

***The Column "Set II" shows mean characteristics of the propensity-score reweighted control group when the propensity score includes variables in Set I and, additionally, the school admissions criteria (whether admits based on residence, whether transfers students for bad behavior) and teachers´ high morale.

****The Column "Set III" shows propensity-score reweighted means of controls when the propensity score includes covariates of Set II plus student's grade in mathematics.

Table 2: Determinants of participation into the FE program

	Covariate Set I	Covariate Set II	Covariate Set III
	(1)	(2)	(3)
	SCHOOL CHARACTERISTICS		
South	1.727	2.029	2.061
	(0.318)	(.353)	(.360)
State-subsidized school		-.278	-.335
		(.396)	(.400)
Private school		-.785	-.823
		(.376)	(.378)
Religious school		.840	.836
		(.369)	(.370)
Does not admit students according to residence		2.007	2.034
		(.276)	(.277)
South x Does not admit students by residence		-.906	-.908
		(.500)	(.499)
Transfers students to other schools		.719	.733
		(.226)	(.225)
Competes with at least one school in the area		1.971	2.009
		(.388)	(.387)
High teacher morale		.708	.697
		(.342)	(.338)
	STUDENT CHARACTERISTICS		
Girl	-.340	-.085	-.089
	(.514)	(.587)	(.593)
Has repeated grade	-0.219	-0.168	.000
	(.358)	(.386)	(.405)
Girl x Has repeated grade	-.069	-.378	-.342
	(.573)	(.600)	(.601)
Father works	-.051	-.120	-.122
	(.362)	(.403)	(.403)
Missing Father works	1.208	1.084	1.090
	(0.669)	(0.735)	(0.726)
Girl x Father works	.405	.407	.465
	(.516)	(.584)	(.591)
Girl x Missing Father works	-.610	-.659	-.608
	(.886)	(.981)	(.971)
South x Girl	-.048	.223	.265
	(.436)	(.474)	(.481)
South x Has repeated grade	-.388	-.34	-.293
	(.523)	(.598)	(.595)
South x Girl x Has repeated grade	.419	.453	.399
	(.829)	(.912)	(.905)
Mother works		.098	.067
		(.219)	(.220)
Missing Mother works		.352	.413
		(.675)	(.674)
Girl x High teacher morale		-1.307	-1.333
		(.528)	(.526)

(continues)

Table 2: Determinants of participation into the FE program (cont.)

	Covariate Set I	Covariate Set II	Covariate Set III
	(1)	(2)	(3)
Normalized score on Maths			.174
			(.109)
Constant	-1.916	-4.609	-4.721
	(.360)	(.590)	(.590)

Sample of 912 students in 150 schools that had not offered financial schools as of the second quarter of 2012. 702 are controls of 120 schools and 210 belong to the treatment group of 30 schools.

The dependent variable takes value 1 if the school participated in the program at some point between the academic years 2012-2013 and 2015-2016.

The estimation method is a Logit, weighted by population weights. Coefficients shown are those from the latent index model. Standard errors in parentheses.

Estimates in Model III are the averages of the results for the 5 different PISA grades, with standard errors corrected for multiple imputation

Table 3: Estimates of selection bias in PISA 2012, reweighted estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Unconditional difference	Covariate Set I		Covariate Set II		Covariate Set III	
Covariates in outcome regression?	No	No	Yes	No	Yes	No	Yes
Panel A: Main outcome variable: Score in financial competences test.							
Normalized score in financial competences							
1. Financial course in 2012-2016	-.073	-.032	-.037	.017	-.026	-.090	-.123
	(.087)	(.100)	(.084)	(.109)	(.091)	(.106)	(.071)*
Panel B: Proximate variables that test for unconfoundedness							
Normalized score in mathematical competences							
2. Financial course in 2012-2016	.003	.104	.098	.185	.143	--	--
	(.076)	(.089)	(.072)	(.109)*	(.085)*		
Normalized score in reading competences							
3. Financial course in 2012-2016	.051	.036	.033	.054	.019	-.052	-.080
	(.081)	(.092)	(.079)	(.109)	(.097)	(.107)	(.073)

Sample of 912 students in 150 schools that have not offered financial schools as of the second quarter of 2012. 702 are students in 120 control schools and 210 belong to the treatment group of 30 schools.

"Treated" schools are those that participated in the program at some point between the academic years 2012-2013 and 2015-2016.

Estimation method: reweighting based on the estimated propensity score. Models (2) and (3) use Set I of covariates, Models (4) and (5) use Set II and Models (6) and (7) use Set III. All estimates combine the 5 different PISA grades and are weighted by population weights. Standard errors are corrected for multiple imputation.

Covariate Set I. Covariates: female, grade repeater, parents work, school located in the South and second-order interactions

Covariate Set II. Set I plus school admissions criteria (whether admits based on residence, whether transfers students for bad behavior) and high morale

Covariate Set III. Set II plus student's grade in mathematics.

***1, **5, and *10 percent significance level, respectively.

Table 4: Estimates of selection bias in PISA 2012, differences-in-differences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Unadjusted	Covariate Set I			Covariate Set II			Covariate Set III		
		Unw.	Reweighted		Unw.	Reweighted		Unw.	Reweighted	
Covariates included?		Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
Dependent variable I: Normalized score in Finance - Math										
1. Financial course in 2012-2016	-.076	-.088	-.136	-.136	-.095	-.169	-.169	--	--	--
	(.061)	(.064)	(.066)**	(.067)**	(.070)	(.079)**	(.079)**			
Dependent variable II: Normalized score in Finance - Reading										
2. Financial course in 2012-2016	-.124	-.051	-.068	-.070	.007	-.037	-.043	.007	-.037	-.043
	(.076)	(.080)	(.086)	(.084)	(.087)	(.102)	(.099)	(.086)	(.105)	(.099)

Sample of 912 students in 150 schools that have not offered financial schools as of the second quarter of 2012. 702 are students in 120 control schools and 210 belong to the treatment group of 30 schools.

The dummy "Financial course in 2012-2016" takes value 1 for schools that participate in the program at some point between the academic years 2012-2013 and 2015-2016.

Estimation method: Differences-in-differences, computed by regressing the student-specific difference in scores on the dummy of "Financial course".

Columns 2, 4, 5, 7, 8 and 10 include the covariates included in each Covariate Set as additional controls in the regression.

Columns 3, 4, 6, 7, 9 and 10 reweight the sample based on the estimated propensity score.

Models (2) and (3) use Set I, Models (4) and (5) use Set II and Models (6) and (7) use Set III. All estimates combine the 5 different PISA grades.

Covariate Set I. Covariates: female, grade repeater, parents work, school located in the South and second-order interactions

Covariate Set II. Set I plus school admissions criteria (whether admits based on residence, whether transfers students for bad behavior) and high morale

Covariate Set III. Set II plus student's grade in mathematics.

***1, **5, and *10 percent significance level, respectively.

Table 5a: Estimates of selection bias in PISA 2012 among females, reweighted estimates

	(1)	(2)	(3)	(4)	(5)
	Unconditional difference	Covariate Set II		Covariate Set III	
Covariates in outcome regression?	No	No	Yes	No	Yes
Panel A: Main outcome variable.					
Normalized score in Financial Competences					
1. Financial course in 2012-2016	-.130	-.070	-.082	-.163	-.155
	(.120)	(.143)	(.117)	(.139)	(.094)*
Panel B: Proximate variables that test for unconfoundedness					
Normalized score in Mathematics					
2. Financial course in 2012-2016	-.013	.115	.108	--	--
	(.110)	(.145)	(.114)		
Normalized score in Reading					
3. Financial course in 2012-2016	.071	.008	-.002	-.078	-.070
	(.125)	(.153)	(.131)	(.154)	(.108)

Sample of 424 females in 150 schools that have not offered financial schools as of the second quarter of 2012. 327 students are controls and 97 belong to the treatment group of schools that participate in the program at some point between the academic years 2012-2013 and 2015-2016.

Estimation method: reweighting based on the estimated propensity score. Models (2) and (3) use Set I, Models (4) and (5) use Set II and Models (6) and (7) use Set III. All estimates are weighted by population weights and combine the 5 different PISA grades.

Covariate Set I. Covariates: female, grade repeater, parents work, school located in the South and second-order interactions

Covariate Set II. Set I plus school admissions criteria (whether admits based on residence, whether transfers students for bad behavior) and high morale

Covariate Set III. Set II plus student's grade in mathematics.

***1, **5, and *10 percent significance level, respectively.

Table 5b: Estimates of selection bias in PISA 2012 among males, reweighted estimates

Panel A: Main outcome variable.					
Normalized score in Financial Competences					
1. Financial course in 2012-2016	-.022	.094	-.024	-.025	-.095
	(.128)	(.158)	(.131)	(.155)	(.097)
Panel B: Proximate variables that test for unconfoundedness					
Normalized score in Mathematics					
2. Financial course in 2012-2016	.019	.247	.174	--	--
	(.106)	(.163)	(.123)		
Normalized score in Reading					
3. Financial course in 2012-2016	.031	.097	.037	-.025	-.089
	(.108)	(.156)	(.138)	(.148)	(.096)

Sample of 487 males in 150 schools that have not offered financial schools as of the second quarter of 2012. 374 students are controls and 113 belong to the treatment group of schools that participate in the program at some point between the academic years 2012-2013 and 2015-2016.
Estimation method: reweighting based on the estimated propensity score. Models (2) and (3) use Set I, Models (4) and (5) use Set II and Models (6) and (7) use Set III. All estimates are weighted by population weights and combine the 5 different PISA grades.
Covariate Set I. Covariates: female, grade repeater, parents work, school located in the South and second-order interactions
Covariate Set II. Set I plus school admissions criteria (whether admits based on residence, whether transfers students for bad behavior) and high morale
Covariate Set III. Set II plus student's grade in mathematics.
***1, **5, and *10 percent significance level, respectively.

Table 6a: Estimates of selection bias in PISA 2012 among females, differences-in-differences

	(1)	(5)	(6)	(7)	(8)	(9)	(10)
	Unadjusted		Covariate Set II			Covariate Set III	
		Unweighted	Reweighted		Unweighted	Reweighted	
Covariates included?		Yes	No	Yes	Yes	No	Yes
Dependent variable I: Normalized score in Finance - Math							
1. Financial course in 2012-201(	-.117	-.132	-.185	-.189	--	--	--
	(.086)	(.092)	(.111)*	(.105)*			
Dependent variable II: Normalized score in Finance - Reading							
2. Financial course in 2012-201(	-.200	-.058	-.078	-.080	-0.058	-.085	-.085
	(.107)*	(.116)	(.147)	(.142)	(0.116)	(.148)	(.141)

Table 6b: Estimates of selection bias in PISA 2012 among males, differences-in-differences

	(1)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable I: Normalized score in Finance - Math							
Financial course in 2012-2016	-0.041	-.063	-.153	-.150	--	--	--
	(.085)	(.092)	(.104)	(.102)			
Dependent variable II: Normalized score in Finance - Reading							
Financial course in 2012-2016	-0.053	.063	-.003	-.014	0.063	.000	.006
	(.107)	(.111)	(.131)	(.128)	(0.111)	(.133)	(.128)

Sample of 912 students in 150 schools that have not offered financial schools as of the second quarter of 2012. 702 are students in 120 control schools and 210 belong to the treatment group of 30 schools.

Estimation method: Differences-in-differences, computed by regressing the student-specific difference in scores on the dummy of "Financial course in 2012-2016".

Columns 2, 4, 5, 7, 8 and 10 include the covariates included in each Covariate Set.

Columns 3, 4, 6, 7, 9 and 10 reweight the sample based on the estimated propensity score.

Models (2) and (3) use Set I, Models (4) and (5) use Set II and Models (6) and (7) use Set III. All estimates make use of the 5 different PISA grades.

Covariate Set I. Covariates: female, grade repeater, parents work, school located in the South and second-order interactions

Covariate Set II. Set I plus school admissions criteria (whether admits based on residence, whether transfers students for bad behavior) and high morale

Covariate Set III. Set II plus student's grade in mathematics.

***1, **5, and *10 percent significance level, respectively.

Table 7: Estimates of selection bias in PISA 2012, reweighted estimates (Covariate sets IV and V)

	(1)	(2)	(3)	(4)	(5)
	Unconditional difference	Covariate set IV		Covariate set V	
Covariates in outcome regression?	No	No	Yes	No	Yes
Panel A: Main outcome variable: Score in financial competences test.					
Normalized score in financial competences					
1. Financial course in 2012-2016	-.073	-.054	-.104	.043	.028
	(.087)	(.117)	(.102)	(.123)	(.100)
Panel B: Proximate variables that test for unconfoundedness					
Normalized score in mathematical competences					
2. Financial course in 2012-2016	.003	-.028	-.105	.106	.070
	(.076)	(.112)	(.085)	(.117)	(.066)
Normalized score in reading competences					
3. Financial course in 2012-2016	.051	-.179	-.271	--	--
	(.081)	(.106)*	(.091)***		

Sample of 912 students in 150 schools that have not offered financial schools as of the second quarter of 2012. 702 are students in 120 control schools and 210 belong to the treatment group of 30 schools.

"Treated" schools are those that participated in the program at some point between the academic years 2012-2013 and 2015-2016.

Estimation method: reweighting based on the estimated propensity score. Models (2) and (3) expand Set II of covariates to include an index of parental wealth, an index of the quality of parental resources and indicators of whether schools have a goal-oriented curriculum. Models (4) and (5) add scores in the Reading part of the assessment. All estimates combine the 5 different PISA grades and are weighted by population weights.

Standard errors are corrected for multiple imputation.

Covariate Set IV. Set II plus parental wealth, an index of the quality of parental resources and indicators of whether schools have a goal-oriented curriculum

Covariate Set V. Set IV plus student's grade in reading.

***1, **5, and *10 percent significance level, respectively.

***1, **5, and *10 percent significance level, respectively.

Table 8: Estimates of selection bias in PISA 2012, differences-in-differences (Covariate set IV and V)

	Unadjusted	Covariate set IV			Covariate set V		
		Unweighted	Reweighted		Unweighted	Reweighted	
Covariates included?		Yes	No	Yes	Yes	No	Yes
Dependent variable I: Normalized score in finance - math							
1. Financial course in 2012-2016	-.076	-.083	-.026	.001	-.101	-.063	-.042
	(.061)	(.077)	(.087)	(.081)	(.077)	(.089)	(.083)
Dependent variable II: Normalized score in finance - read							
2. Financial course in 2012-2016	-.124	.074	.124	.168	--	--	--
	(.076)	(.097)	(.107)	(.106)			

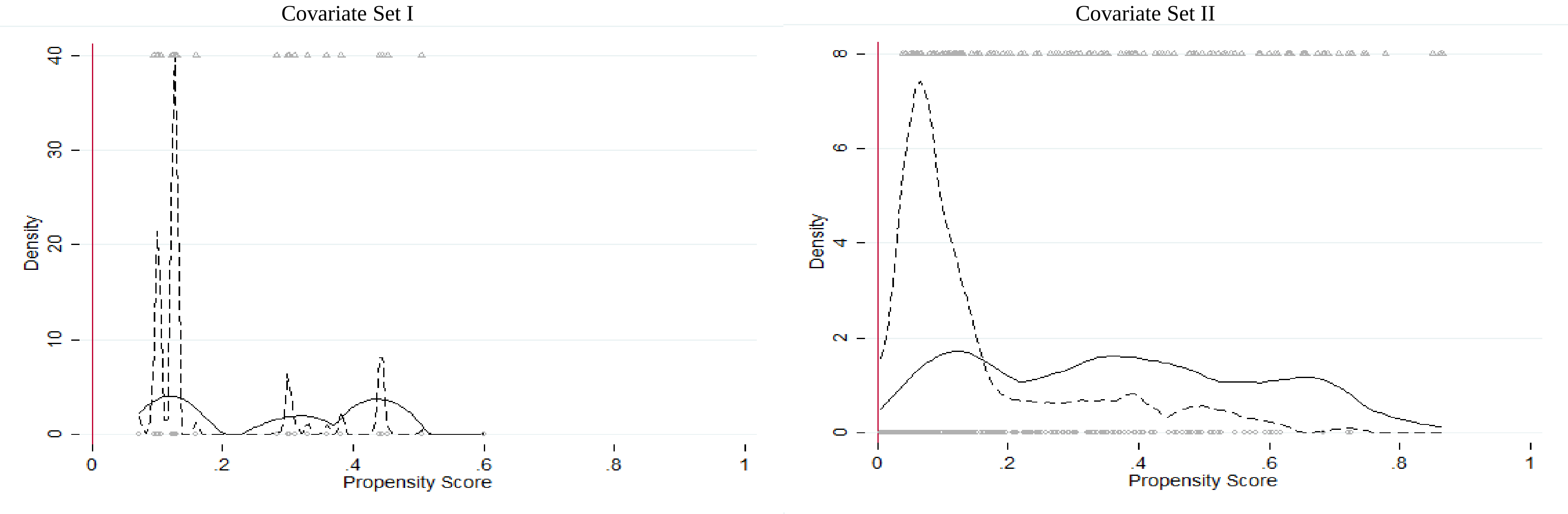
Sample of 912 students in 150 schools that have not offered financial schools as of the second quarter of 2012. 702 are students in 120 control schools and 210 belong to the treatment group of 30 schools.
"Treated" schools are those that participated in the program at some point between the academic years 2012-2013 and 2015-2016.
Estimation method: Differences-in-differences, computed by regressing the student-specific difference in scores on the dummy of "Financial course in 2012-2016". All estimates combine the 5 different PISA grades and are weighted by population weights.
Standard errors are corrected for multiple imputation.
Covariate Set IV. Set II plus parental wealth, an index of the quality of parental resources and indicators of whether schools have a goal-oriented curriculum
Covariate Set V. Set IV plus student's grade in reading.
***1, **5, and *10 percent significance level, respectively.
***1, **5, and *10 percent significance level, respectively.

Table A1: Differences between participants and non-participants in the FE program, by gender.

	Males		Females	
	Weighted averages		Weighted averages	
	Treated	Controls	Treated	Controls
	STUDENT CHARACTERISTICS			
Normalized score, Financial Competences	0.02	-0.01	-0.09	0.02
Normalized score, Maths	0.03	-0.01	-0.02	0.01
Normalized score, Reading	0.03	-0.01	0.03	-0.01
Has repeated a grade	0.36	0.39	0.24	0.28
Father works	0.80	0.82	0.79	0.76
Missing Father works	0.07	0.03	0.08	0.07
Father with college degree	0.31	0.34	0.29	0.29
Mother works	0.62	0.67	0.61	0.65
Mother with college degree	0.05	0.03	0.30	0.38
Family Structure: Single parent	0.12	0.10	0.14	0.12
Immigration status	0.11	0.12	0.09	0.12
Cultural Possessions	0.11	0.08	0.16	0.10
Wealth	0.21	-0.04	-0.06	-0.13
ICT Availability at Home	0.32	0.24	-0.10	0.03
Time of computer use (mins)	409.19	660.06	66.33	506.72
Number of students	113	374	97	328

Notes: Sample of 912 students in 150 schools that have not offered financial schools as of the second quarter of 2012. 702 students are in 120 control schools and 210 belong to the treatment group of 30 schools. All averages weighted by the sample weights.

Figure 1: Distribution of the predicted probability of participating in the program using Covariate Sets I and II



Notes: Density of probability of participation (propensity score) for treated (solid line) and control groups (dotted line). The circles on the X-axis show the probability of participation values for each student in the control group. The triangles represent the probability of participation for each student of the treated group. The graph in the left hand side corresponds to Covariate Set I. It contains the location of the school, the parents' work status, and the student's gender, rate of repetition and second-order interactions. The graph in the right hand side refers to Covariate Set II. It contains all the variables included in Set I and, additionally, school admissions criteria (whether admits based on residence, whether transfers students for bad behavior) and teachers' high morale.