Student literacy one year later. On school value added estimation using PISA–OECD*

Massimiliano Bratti Daniele Checchi

May 20, 2012

Abstract

Thanks to an 'experiment' realized in the Trento province (North-Eastern Italy) in 2010, we investigate the potential advantages of re-testing the same students originally sampled in PISA 2009 one year later. While in principle this represents the proper strategy to measure the schools' 'value added' (VA) in student competence formation, our preliminary exploration of the data suggests that cross sectional estimates of school VA provide rather similar results. We put forward that this may be due to the rich information provided by the PISA survey, which may be used to proxy for students' past performances. We also discuss the issue of panel attrition, which is typical of any longitudinal study, and estimate models in which a school's value added may differ across students' characteristics [*JEL*. I21 J24]

Keywords. Italy, PISA OECD, school effectiveness, value added

1 Introduction

Measuring schools' or teachers' value added using longitudinal data is increasingly considered as the most appropriate methodology to assess their effectiveness (Chetty et al.

*This paper represents work in progress. We wish to thank for their collaboration INVALSI (Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione), and in particular Piero Cipollone, Patrizia Falsetti, Sabrina Greco e Laura Palmerio, and IPRASE (Istituto Provinciale per la Ricerca e la Sperimentazione Educativa), in particular Francesco Rubino. We also thank Anna DePaoli for excellent research assistantship.

Massimiliano Bratti. Department of Economics, Business and Statistics, Università degli Studi di Milano and IZA, Via Conservatorio 7, 20122 Milano, Italy. E-mail: massimiliano.bratti[at]unimi.it

Daniele Checchi. Department of Economics, Business and Statistics, Università degli Studi di Milano and IZA, Via Conservatorio 7, 20122 Milano, Italy. E-mail: daniele.checchi[at]unimi.it

2012, Rothstein 2010, Ray 2006). However this claim does not go uncontested, due to the implict assumptions underlying sample selectivity and functional forms (Figlio 1999). In this paper we exploit one experiment conducted in the province of Trento (Noth-Eastern Italy) where the same students who were tested in the PISA (OECD) survey of 2009 were retested using a random assignment of the test booklets. Given the sampling strategy of PISA, which draws a random selection of approximately 35 students per school (enrolled in different classes), we can only evaluate school effectiveness, while we cannot compute measures of class effectiveness (Dearden et al. 2011).

There are several methodological issues associated to properly measuring schools' value added (VA, hereafter) using PISA test scores. First of all, OECD claims to measure 'knowledge for life' and not simply curricular competences: it is not clear how these competences evolve over-time and the contribution that schools can give to increasing this specific type of knowledge.[1] Second, attrition (of both schools and students) plagues any longitudinal dataset, and this may be relevant in measuring schools' value added. Third, in countries where students are tracked according to ability (in the German tradition), measuring the value added has to take students' sorting into account. Last but not least, many models estimate the value added through *intercept* effects, but it may clearly be the case that schools are effective in equalizing opportunities.

In the present paper we explore all these dimensions, showing that whenever the information available on students are rich enough (as in the PISA dataset), longitudinal datasets are less crucial in properly assessing school effectiveness.

The paper is structured as follows. Section 2 briefly introduces the geographical context of our analysis. Section 3 describes in detail the main characteristics of the 2010 PISA re-test. In section 4 we describe the main empirical models that we will use to compute school value added, distinguishing between cross-sectional and longitudinal models. Panel attrition, either originating from schools' or individual students' non-participation in the exercise, is discussed in section 5. The results of school value added estimation are reported in section 6. In section 7 we relax the assumption of the homogeneity of school VA across students (i.e. *intercept* effect), and section 8 concludes.

2 Geographic context

The Autonomous Province of Trento is a small region of half million of inhabitants, located in the North-East of Italy, close to the Austrian border. As other bordering regions (like Valle d'Aosta), due to political reasons related to the difficult process of country unification it enjoys grater autonomy in administration (like school design) and revenue

[1]Indeed, most studies focusing on longitudinal data for the computation of school VA use measures of 'curricular' knowledge.

collection (not participating to the cross-region subsidisation). It follows the Italian scheme of a tracked secondary school system, even if the vocational tracks enjoys better standards, in the German tradition. When looking at student achievements through the PISA lens (see table 1) we observe that students from Trento schools obtain results that are in line with the bordering regions, at a higher level than the Centre-Southern regions. This is mostly attributable to the relative performance of vocational schools, which score almost half of a standard deviation above than their counterpart in the rest of the country.

Before going into the analysis of school quality, the differences across regions and across school tracks raise questions about the student allocation across tracks. When we look at the variance decomposition (table table 2), we notice that the within-track variance is higher in Trento compared to the rest of the country, while it is not the case when we consider the within-school variance. This suggests that school tracks play in the Trento province a significant role in sorting students, but then school quality seems rather homogenous within tracks.

The literature points to family background as one of the main factors driving students into different tracks. Without resorting to multivariate analysis, simple descriptive statistics (table 3) suggest that sorting by social background is less pronouced in Trento *vis a vis* the rest of the country. While in the rest of Italy students from better backgrounds (higher social prestige associated to parental occupation, better parental education as measured by years of education and better ESCS scores) are gathered by high schools, then by technical schools and eventually by vocational schools, in Trento this process is less pronounced, and students in vocational schools seem better endowed with parental resources compared to the rest of the country.

3 Description of the PISA re-test

During the winter 2010 all secondary schools in the Trento province (located in the North-Eastern region Trentino Alto Adige) whose pupils had taken the test PISA 2009 were contacted and asked to resubmit an equivalent test to the same students. It was decided to reuse the same PISA booklets as in 2009 for two reasons. First, the questions are not related to academic curricula set by the Ministry of Education, but are intended to verify that students have accumulated the competences needed to 'play a conscious and active role in society and to continue learning throughout life' [our translation] (INVALSI, OECD, 2009). Consequently, administering the test a second time does not limit the ability to check for improvements in pupils' literacy levels. Second, the test is available in 13 different versions of an equivalent level of difficulty (booklets) and has, therefore, been possible for each participant to minimize the number of identical questions between the two test sessions.

3.1 Structure of the PISA test in 2009 and 2010

Each student has been assigned a booklet in 2009 and one in 2010, according to the scheme described in Table 4. Each booklet consists of four sections and each section is identified by a type (M = maths, S = sciences and R = reading) and a numerical value (Table 4). There are, therefore, three different sections for mathematics (M1, M2 and M3), three different sections for sciences (S1, S2, S3) and seven different sections for reading (R1, ..., R7). It has to be noted that the redistribution of the booklets was such that each student answered in 2010 to some questions that had already responded in 2009. This overlap occurs for all students, but only for a quarter of the questions, which is a single section of the test (Table 4 shows the section for which there is overlap). For example, all students who had received the booklet 1 in 2009 were assigned the booklet 7 in 2010, and the section M3 is repeated in both years.

The students then had to answer several questions, three quarters different and one quarter equal between sessions, and as shown in the diagram the test is heavily biased towards the assessment of reading skills. In fact, the test PISA 2009 (and therefore also 2010 re-test) is aimed at testing students' understanding and analysis of written texts, although there are sections that test the competences learned in science and mathematics.

3.2 Participating schools and test administration

The upper secondary schools in the Trento province whose students had taken the PISA test in 2009 are 50 of which only 35 agreed to a second administration of the test.[2] We will discuss in section 5 any potential selection bias due to school's or student's non participation to the 2010 re-test.

For each school, and whenever it was possible, the reference person who had been responsible for testing in 2009 was contacted. They are school teachers, but are not necessarily the pupils' teachers since students in the PISA school samples are drawn from different classes. On April 16, 2010 a meeting was held with all the schools' reference persons to illustrate the mode of administration of the test, the objectives of our 'experiment' and clarify any doubt from the school side.

In May, the bookets containing the test questions were sent to IPRASE (*Istituto Provinciale per la Ricerca e la Sperimentazione Educativa*) by INVALSI (*Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione*) together with the lists containing the identification codes of the students and of the booket attributed

[2]As our main focus is on VA estimation for upper secondary schools, we dropped from the analysis one lower secondary school which was sampled in PISA 2009.

in 2010 according to the criteria described above. Schools were given a window of two weeks for conducting the test. The schools were required to return to IPRASE the booklets compiled within 48 hours from completion of the re-test. In June, the tests were sent from IPRASE to INVALSI which carried out the correction and scoring.

In the meeting of April 2010 some doubts have emerged about absent pupils and the conditions of administration. In the case of pupils who were absent on the day fixed by the school referent, they were not given the opportunity to be tested in a second session, so all the students who were not present on the appointed day were excluded from the re-test. However school referents were asked to specify the reasons of the student absence from school: transfer to another school, authorisation denied by parents, school drop-out, or a 'ordinary' absence.

School referents were instead asked to administer the tests to the students who were sampled for PISA 2009 but were absent in the test day, and who were present in 2010. Last, with the people in charge of the test administration was largely discussed the importance of trying to replicate for the 2010 excercise the same testing conditions as in 2009 (length, rules, rooms' characteristics, time, etc.). The main goal was to make the two testing exercises as similar as possible in order to avoid different testing conditions to have a bearing on the test results.

INVALSI was constantly present in each stage of the re-test exercise, providing us with the technical assistance for the test administration, the correction of the open questions, and the estimations of student scores (including the plausible values).

4 Econometric models for school value added estimation

There exist in the literature various definitions of school value added (VA). In general, school VA can be defined as the contribution that schools give to students' competences over and above contextual factors. A good school VA model should take into account the chacteristics of school's student intake, which are likely to affect their competences irrespective of the schools they are enrolled in. Educational value added can be evaluated at different levels of aggregation. One can be interested in the school's VA or in teachers' VA, as in Rothstein (2009). In this last case in order to distinguish the effect of teachers from that of the peer group it is necessary to have information on different classes within the same school, which is not the case for PISA-OECD.

In the background of the exercise we make in this paper there are some important assumptions. The first one is that PISA tests, which measure 'competences for life' and not curricular competences, can be a useful means to evaluate schools. To put in other words, although performance in PISA tests may depend on a variety of factors, some internal and some others external to the schools (e.g., family and geographical contextual factors), the contribution given by schools to a student's 'competences for life' is an important dimension of a school's social role, and as such can be a matter for school evaluation. The second assumption is that the administration to 16 years old students of tests which are built to test the knowledge of 15 years old students can give useful information on the former's 'competences for life'.

We assume that data generating process for student literacy can be expressed through an educational production function (EFP)

$$T_{ijt} = f(\mathbf{X}_{ijt}, T_{ijt-1}, VA_{jt}, \mu_{ij0}, \epsilon_{ijt}) \tag{1}$$

where i, j and t are subscripts for individual, school and time respectively. Current student performance in a test measuring literacy T_{ijt} depends on a vector $\mathbf{X}_{ijt}$ of current inputs external to the school (e.g., family inputs), on past educational performance T_{ijt-1}, on school VA (VA_{jt}) which depends on the quantity and quality of school inputs (such as schooling infrastructures, teachers, etc.), on inherited child ability (μ_{ij0}), which is often unobservable by the researcher, and on other unobservable factors entering the residual ϵ_{ijt}. In this specification we have assumed that past inputs only affect current performance through past performance. Todd and Wolpin (2003) define (1) as the 'value-added specification'. This specification recognizes the cumulative nature of the learning process, hence past cognitive achievement contributes to current cognitive achievement.[3] Assuming a linear specification, the EPF becomes

$$T_{ijt} = a_0 + \alpha\mathbf{X}_{ij} + \gamma T_{ijt-1} + \sum_j VA_{jt}S_{jt} + \beta\mu_{ij0} + \epsilon_{ijt}. \tag{2}$$

where S_{jt} is a school indicator which equals one if student i is enrolled in school j at time t and zero otherwise. In this formulation, one assumes that the contribution given by school j to its students' literacy levels (VA) is the *same for all students*, i.e. the school has an intercept effect only. This assumption could be relaxed by including interaction terms between the school indicator S_{jt} and student characteristics ($\mathbf{X}_{ij}$) but this is feasible only if a large number of students is sampled for each school, which will not be the case for the data used in this paper (as PISA–OECD sampled 35 students per school).

Although equation (2) it is the one that the researcher would like to estimate, very often there are no repeated observations on student performance, and he is forced to

[3] We depart from Todd and Wolpin (2003), who define as the true model the 'cumulative' model in which all (current and) past inputs have an effect on current student achievement which is not necessarily mediated by past achievement. This is equivalent to saying that inputs provided by parents during childhood may have long-term effects on children's achievement levels over and above past achievements. We prefer to adopt a 'value-added specification' as early parental inputs are almost never available and in this case if the 'cumulative' specification were the true model it will be never possible to estimate it consistently.

omit past performance T_{ijt-1} from the controls. Using the language of Todd and Wolpin (2003), the researcher is forced to estimate a 'contemporaneous specification'

$$T_{ij} = a_0 + \alpha \mathbf{X}_{ij} + u_{ij} \tag{3}$$

where the new residual is $u_{ij} = \gamma T_{ijt-1} + \sum_j VA_{jt}S_{jt} + \beta\mu_{ij0} + \epsilon_{ijt}$. This specification is often estimated using ordinary least squares, and used to obtain a measure of school value added. Indeed, with an estimate for α it is possible to also obtain an estimate of students' residuals u_{ij}, which are averaged at the school level to build a measure of the school VA, $VA_j = (\sum_j \hat{u}_{ij})/N_j$, where N_j is the number of students in school j. We call this procedure as the Contextual Value Added (CVA) model. School VA is calculated as the mean difference by school from the regression line, that is as the mean difference between students' observed performance and the performance predicted according to the average behavior of the students in the sample. It is immediate to note that such procedure will generally lead to biased estimates of current school VA. First, consistent estimation of α would require it to be orthogonal to all components entering ther rediual u_{ij}, that is past performance, school VA (and therefore also the choice of school inputs), student unobserved ability and other current student unobservables. Second, even with a consistent estimate of α the residual will include not only school VA but also past student performance — a function of past educational inputs, past school VA, other past unobservable student characteristics, and unobserved ability — , student ability and current unobservable inputs. This implies, among other things, that any factor omitted from the regression but common to all students of a given school will enter ϵ_{ijt} and contribute to the determination of its 'estimated' value added, even if it has nothing to do with the educational process taking place inside the school. These factors may include for instance the effect of contextual factors such as residential peer groups' effects related to the schools' catchment areas.[4]

In the absence of repeated observations for each student, the researcher may want to obtain the school VA using fixed effects. We label this model as the School Fixed Effect (SFE) model.

$$T_{ijt} = a_0 + \alpha \mathbf{X}_{ij} + \sum_j \phi_j S_{jt} + v_{ij} \tag{4}$$

where $v_{ij} = \gamma y_{ijt-1} + \beta\mu_{ij0} + \epsilon_{ijt}$. Estimating this model with OLS one can obtain the estimates for the fixed effects (ϕ_j), which could be considered as a measure of school VA. However, let us consider under which conditions estimation of this model leads to

[4]This is a subtle issue. Indeed, residential peer-group effects may appear as VA for parents, which enrolling their children in a given school will also benefit from these additional positive returns of living in a high-quality neighborhood. However, residential peer effects should not be considered as part of the school VA by the Government, if school VA has to be used to evaluate the 'performance' of teachers and schools and to reward them with the allocation of additional financial resources.

consistent estimates of school VA. This happens only when the regressors are orthogonal to the error term, and implies that (i) external inputs X_{ij} are uncorrelated with past achievements; (ii) external inputs are orthogonal to student unobserved ability; (iii) external inputs are orthogonal to all the omitted inputs entering the error term or no relevant input has been omitted from the model; (iv) all these conditions must hold also for school VA. Assumption (i) is likely to fail as past achievements are a function of past inputs, which are likely to be correlated with current inputs (think of family inputs). Also assumption (ii) is not very credible as parents may decide educational inputs according to children's ability, applying reinforcing or compensatory policies. Last but not least, assumption (iii) requires either very rich data or strong confidence on the part of the researcher that he knows the process of cognitive achievement, which is rarely the case; The same arguments hold for assumption (iv). If these assumptions are not met, then the estimate of school VA may reflect differences in past student performance, student ability or other current unobservable factors across schools. Two advantages of model (4) over model (3) however are that it does not require school VA to be orthogonal to current external inputs ($\mathbf{X}_{ij}$), and the measure of school VA does not include other factors in addition to school VA, even if it may capture part of them in case the conditions (i)-(iv) above do not hold. We consider the SFE model as an improvement over the CVA model but it is still far from being the ideal method for computing school VA.

Let us consider now the lucky case in which the researchers has multiple time observations for the same student's competences and can estimate the following model

$$T_{ijt} = a_0 + \alpha \mathbf{X}_{ij} + \gamma T_{ijt-1} + \sum_j \phi_j S_{jt} + \xi_{ijt} \tag{5}$$

where $\xi_{ijt} = \beta \mu_{ij0} + \epsilon_{ijt}$ and the fixed effects ϕ_j provide the estimates of school VA. Unfortunately, even in this case the researcher is not so lucky. Indeed, OLS estimation of model (5), which we label the Longitudinal School Fixed Effect (LSFE) model, does not generally provide consistent estimates of school VA. This happens for a variety of reasons: (i) observed external inputs may be correlated with student unobserved ability and with other unobserved external omitted inputs, (ii) past student performance is correlated with student unobserved ability and may be correlated with current unobserved external inputs, (iii) school VA may be correlated with student ability or with the omitted inputs. Consistent estimation of model (5) could be achieved including in the regression a measure of student inherited ability, and by instrumenting $\mathbf{X}_{ij}$ and T_{ijt-1} to make them orthogonal from ϵ_{ijt}. Note that using instrumental variables (IVs) for X_{ij} and T_{ijt-1} in the absence of a control for μ_{ij0} will generally lead to inconsistent estimates of ϕ_j's if school VA is correlated with student unobserved ability (i.e., students' allocation to schools must be random to have consistent estimates). It is however true that on the grounds that the unobservables (ϵ_{ijt}) are highly correlated over time, T_{ijt-1} will partly capture the effect of student ability and other current unobservable inputs, purging the error term from potential sources of correlation with both $\mathbf{X}_{ij}$ and school VA. To put in other words, although consistency may be difficult to achieve, the LSFE model may often represent the best available option to compute school VA.

5 Panel attrition

As we have anticipated in Section 3, participation of schools to the Trento's PISA 2010 re-test took place on a voluntary basis. For this reason, the results of the re-test exercise cannot be considered as representative of the 16-year-olds population of Trento. What kind of bias can be expected from the schools' self-selection into the re-test? It is plausible to think that the relatively better peforming schools in PISA 2009 may have accepted to participate (*positive selection*) since they were expecting better results also in the 2010 re-test, that is our sample may overestimate the competences of Trento's students. However, such *positive selection* is less likely to have taken place in terms of VA — the specific contribution given by schools to the improvement of student competences, as schools may have only a very vagous idea of their VA.[5]

Among the 50 Trento's upper secondary schools sampled in PISA 2009, 15 (30%) refused to participate to the 2010 follow-up. It is important to stress that at the time the schools were asked to participate in the re-test (May 2010), PISA's results were not officially released yet (it happened in December 2010). Therefore, schools may have had some idea of their expected performance in PISA, but they did not know their results with certainty. This may be relevant as far as the schools' self-selection into the follow-up test is concerned. As a matter of fact, many schools did not participate advocating 'confidentiality' reasons. Italy has a tracked secondary school system. School types are: high school (*liceo*), the 'academic' track; technical schools (*Istituto tecnico industriale statale*, ITIS); vocational schools (*Istituti di Formazione Professionale*, IFP); training courses (*Corsi di Formazione Professionale*, CFP). The last two tracks are generally chosen by least able students, who do not plan to continue in Higher Education. Among the non-participating schools, 4 were high schools, 4 technical schools, 2 vocational schools, and 5 training schools. In Table 5 we present the estimates of the marginal effects from a probit for a school's probability to have participated to the re-test. Given the low number of observations, we have specified very parsimonious models. The first model (column 1) only includes the average of the 2009 school's PISA score, the idea is that schools with low scores may be less likely to participate (being able to predict their not yet released 2009 score). The second model (in column 2) also includes school type fixed

[5]In fact, for the US Rothstein (2009) shows evidence of non-random assignment of teachers to classes also in terms of potential competences' improvements, i.e. of VA.

effects, the idea is that although schools did not know their 2009 PISA score at the time they were asked to participate, schools in each type had an idea of their relative ranking (e.g., that vocational schools and training courses would have worse performances). In column (1) the marginal effect of past PISA score turns out be positive, but statistically insignificant: a one-standard deviation (100 PISA points) increase in the average school PISA 2009 score is associated with a statistically insignificant 13.1 percent points lower probability of participating. Results are the same in model (2) in which we control for school types. Overall, none of the regressions in Table 5 show clear evidence of a positive and significant school's self-selection into the re-test exercise according to their 'expected' PISA 2009 peformance, as neither past PISA performance nor school types turn out to be significant predictors of the probability to participate.

However, our data may be subject to a second source of selection. Indeed, the schools which decided to participate may adopt strategic behaviors by encoraging participation of abler students and discouraging that of less able students, so as to artificially inflate their measured performance.[6] In any case, voluntary absences are likely to be higher among low-peforming students, which will bias downward the PISA scores.[7] There is also another source of *panel attrition* which may potentially bias our estimates: some students may have dropped-out from education or transferred to another school, and this is unlikely to be random with respect to their past performance. In particular, we expect least able students to have dropped-out or transferred to other schools, which may introduce an upward bias in the measured competences. Also in this case, like for schools' participation, is less likely that the students' self-selection takes place on school VA. The percentage of absences from the re-test is 18.8%: 10.43% represent standard truancy, 0.19% refer to children whose parents denied permission, 0.10% to children with 'special needs' —e.g., disability—, 6.6% to children transferred to another school, whose name is known, and 1.44% to those who dropped out or tranferred to another school, whose name is however unknown. As it is clear most part of absences are 'ordinary' absences, but as we said randomness is unlikely also for these absences.

In Table 6 we report the marginal effects from the probit estimates for a student's probability to participate to the re-test exercise, conditional on her school having decided to participate. In this model, which is estimated at the student level, we can include a wider set of controls. We estimated three models, one only including past PISA score, another also including gender, household's demographic structure, student immigrant status, parental education and parental occupation (HISEI, i.e. the highest International Socio-Economic Index between the two parents), and the last one controlling also for school type. It must be noted that most of the additional controls may have a direct

[6] See Bratti et al. (2004) for a discussion on this in the context of Higher Education.

[7] Although the direction of the bias on school VA is less clear.

effect on absenteism, e.g., highly educated parents may put more value to education and induce their children to reduced absenteism, and an indirect effect through their impact on past (and expected future) performance. Models in columns (2) and (3) also include the day of the week in which the test was administered, as student absenteism may be concentred especially in certain days of the week. This variable will be also useful in our later attempt to address student self-selection in the estimation of the educational peformance equation. The results in column (1) show that there is indeed a statistically significant positive association between PISA past peformance and the likelihood of having participated in the re-test exercise, although the marginal effect is not very large: a one-standard deviation increase in the PISA score raises the probability of participating by 5.8 percent points. Column (2) shows that such positive association is not reduced when other contextual factors potentially affecting both past performance and student absenteism are included in the regression. Curiously enough, absenteism does not appear to be related to family background (such as parents' education and socio-economic status), except for the negative marginal effect of living in a single-parent household (-7.8 percent points), which is only significant at the 10% statistical level. By contrast, absences turn out to be significantly more frequent especially on Friday and Saturday (-7.8 percent points), i.e. at the end of the week with respect to the beginning of the week (Monday-Tuesday). We posit that this may be capturing voluntary absences, as we are not aware of any evidence that students tend to be sick in those particular days of the week. Column (3) confirms the main findings in the previous column. In particular, we observe a slight reduction in the marginal effect of the past PISA score and of its significance. Conditional on the covariates included, absenteism does not appear to be lower in other types of school compared to high school. Thus, the positive association between past student performance and attendance to the PISA re-rest does not seem to be capturing the higher absenteism in the worst performing schools (typically vocational schools and training courses) in terms of PISA unconditional outcomes, as it remains significant even after conditioning on school types. This may signal strategic behavior by all schools, which may induce weaker students not to participate in the test, or the simple fact that worse performing students in PISA are also more likely to be more absent from school irrespective of the test, and this does not depend on school type.

Hence, as for panel attrition, it is especially the self-selection operating at the student level which might increase the average ability levels of the students who took part to the re-test exercise. This should be kept in mind when we will present the results of school VA estimation in the following sections of this paper.

6 School value added estimation

Table 7 reports performance equations estimated with the CVA, the SFE and the LSFE models described in section 4. Although we present the estimation of the student performance equation also for the CVA model, school VA will be presented for the SFE and LSFE only, as the assumptions underlying the first model seem too strong. In all regressions we used the 5 plausible values (PVs), the 80 balanced repeated replications (BRR), and the Fay's adjustment (0.5) to compute the standard errors. In column (1) we observe a negative association between male gender and performance in reading, of about 26 PISA points, and large negative and significant coefficient on first-generation immigrants, of about -78 PISA points. By contrast, highest parents' socio-economic status (hisei) is positively associated with student peformance. As for the parents' education only children of parents with isced 5b have significantly (at 10%) lower performance. In model (2) we control for school fixed effects. This makes the coefficients on male gender and hisei lower and statistically insignificant. The coefficient on first-generation immigrants falls at about -53 points, but remains statistically significant at the 1% level. In model (3) we also include the 2009 PISA score in reading (namely the mean of the 5 reading literacy's PVs in 2009), which turns out to be highly significant with a coefficient of 0.275, and the main consequence is that of further reducing the magnitude and significance of the coefficient on first-generation immigrants, and no other variable remains statistically significant.[8]

Thus, controlling for the past PISA score does not seem to make a huge difference in terms of the coefficients estimated on the contextual variables included in the performance equation. However, we have a specific interest in the changes that it produces on school VA estimation. For this reason school VA point estimates, along with their confidence intervals, are reported in figures 1 and 2. Visual ispection of figure 1 suggests that after controlling for several contextual factors, which may explain inviduals's choices of specific schools, there are significant differences in student peformance across school types. Indeed, high schools and technical school's students generally perform better than students enrolled in vocational schools and training courses. However, in most cases it is not possible to rank schools within type: their students' performances do not differ statistically (confidence intervals are always overlapping). This of course casts serious doubts on the usefulness of these ranking exercises, when the final objective is to rank single schools and just a small group of students is sampled in each school (see Goldstein and Spiegelhalter 1996). The overall picture remains unchanged when the LSFE model is used to rank the schools. From figure 2 two things are worth noting. First, the performance gap between the low-performing and the high-peforming school types tend

[8]A simple regression of the reading literacy in 2010 on reading literacy in 2010, without controls, returns a coefficient of 0.60.

to decrease after controlling for past student performance. Second, the ranking obtained is very smiliar as the one in figure 1: even after controlling for past PISA score, students enrolled in high schools and technical schools tend to perform better. The mean of the fixed effects is -46.32 for high schools, -76.60 for technical schools, -104.86 for vocational schools and -147.97 for training courses. This can be interpreted as evidence of true differences in VA by school type: there is an effect of schools on literacy levels over and above the level of students' past academic readiness as controlled by the PISA 2009 score.

As neither figure 1 nor figure 2 are helpful to evaluate the change in the ranking of specific schools produced by using the two different methods, figure 3 shows a cross-plot of the school rankings obtained applying the SFE and the LSFE methods. In the absence of changes in the school ranking all schools should lie on the 45-degree line. The schools above the 45-degree line improve their position when using the LSFE vs. the SFE model, those below the 45-degree line loose positions. The figure shows that most 'movers' are located in the middle of the ranking, while bottom and top peformers remain the same. The Spearman rank correlation index (Spearman's ρ) for the two sets of school fixed effects (obtained with SFE and LFSE) is 0.974 confirming that rankings obtained from the two methods are very close.

6.1 Correcting the estimates for panel attrition

In Section 5 we reported evidence of non-random panel attrition. In particular, student absenteeism during the PISA 2010 re-test appear not to be random with respect to PISA 2009 performance. For this reason, not taking account of panel attrition may bias the estimates of school VA, and the ranking produced using these estimates.

In this section we make an attempt to control for panel attrition. This is done by using the Heckman's selection model (Heckman 1979), i.e. by simultaneously estimating the selection equation and the EPF using Maximum Likelihood (ML). The Heckman's selection model (Heckit, hereafter) is formally identified without exclusion restrictions, although they are often useful as the inverse Mill's ratio is approximately linear in most of its domain (Puhani 2000). For this reason, we use as the exclusion restriction the day of the week in which students were tested. The identifying assumption is that the day of the week affects only absenteeism but not performance in the test.[9] Being tested on Friday or Saturday turned out to be a highly significant predictor of student absence. The Wald test for the exclusion of the day of the week from the selection equation returns a p-

[9]It is possible to find reasons why this assumption may fail. For instance, if students are more tired the last days of the week because of their weekly study workload. In an exactly identified model, the validity of the instrument cannot be tested, but when simply including the day of the week's indicators in the educational performance equations they turned out to be statistically insignificant.

value of 0.0055 and a $\chi^2(2) = 10.42$).[10] Curiosly enough, the Heckit estimator reveals some evidence of significant *negative selection* in terms of students' unoservables, which appear to positively affect participation to the PISA re-test but negatively affect performance in the 2010 re-test.

Table 8 shows the estimates for the selection and the educational performance equations using the Heckit estimator only for the LSFE model. The estimates in column (1) should be compared with those in column (3) of Table 7. Correcting for student attrition slightly decreases the coefficient on past performance and first-generation immigrants, but has no other relevant effect. Figure 4 presents a cross-plot of the LSFE rankings not adjusted and the one adjusted for attrition. Also in this case, 'movers' are mainly located in the middle of the ranking. Some technical schools appear to improve their positions when applying the correction for panel attrition, unlike high schools which loose ground, but most schools lie on or close to the 45-degree line. Indeed, the Spearman's ρ between the school fixed effects estimated with and without correction for sample selectivity is 0.96.

7 Heterogeneity in school value added

TBW

In this section we relax the 'intercept' approach and allow for the school's VA to differ across students, e.g., according to their socio-economic status (hisei, ecsc).

8 Concluding remarks

This paper represents a first attempt to build schools' longitudinal value added indicators using the OECD–PISA survey, and measures of student competences which are not merely curricular but that may be particularly relevant for individuals' welfare ('knowledge for life'). Although value added indicators using PISA data are generally based on contemporaneous specifications using cross-section data, as the survey does not follow individuals overtime, thanks to an 'experiment' implemented in the autonomous province of Trento (North-Eastern Italy) a second administration of the PISA test took place in 2010 to the same students originally tested in 2009.

We first analyze and discuss potential non-random panel attrition, due either to schools' or students' non-participation to the 2010 re-test, which had a voluntary nature. Our analysis shows that although schools did not appear to be positively selected

[10] Here, unlike for instrumental variables' estimation, there are no papers suggesting threshold values of the statistics for detecting 'weak' idenfication. In any case, even in the absence of very strong excluded 'instruments', in the Heckit model non linearity helps the model identification.

into the re-test — perhaps also because at the time schools were asked to participate PISA 2009 results were not public yet — the same cannot be said for individual students within the schools who accepted to participate. Indeed, students who were present the day of the re-test were significantly abler than those who were absent. This may signal strategic behavior by schools in eliciting participation of weaker students, or the fact that ordinary truancy is higher among the low-performing students.

We then use the longitudinal structure of the data to build school VA indicators which control for past student performance and report the relative schools' performance indicators and school rankings, by comparing 'cross-sectional' vs. 'longitudinal' methods. Our results show that irrespective of the particular method used, students in high schools and technical schools perform significantly better than those in vocational schools and training courses, and that this happens also in 'longitudinal' measures of school VA. The use of the latter measures, though, tend to reduce the performance gap between top and bottom rankers. The correlation between 'cross-sectional' and 'longitudinal' rankings turns out to be very high (the Spearman's ρ is 0.97) confirming that in the case of Trento the schools' positions in the overall ranking are not sensitive to the particular method used. Our intrepretation of this finding is that in very rich cross-sectional data, such as the one provided by PISA, the many contemporaneous contextual variables that can be included in the analysis are good predictors of past performance, and therefore the use of longitudinal data does not give particular advantages for the computation of school value added. Of course, it would be interesting to repeat our exercise to the whole country to assess whether the same conclusions hold when very heterogenous geographical contexts (such as North, Centre and Southern Italy) are included in the analysis.

References

Bratti, M., McKnight, A., Naylor, R., Smith, J., 2004. Higher education outcomes, graduate employment and university performance indicators. Journal of the Royal Statistical Society Series A 167 (3), 475–496.

Chetty, R., Friedman, J. N., . Rockoff, J. E. ., 2012. The long-term impacts of teachers: teacher value-added and student outcomes in adulthood. NBER Working Paper 17699.

Dearden, L., Micklewright, J., Vignoles, A., 2011. The effectiveness of english secondary schools for pupils of different ability levels. DoQSS Working Paper No. 11-06.

Figlio, D. ., 1999. Functional form and the estimated effects of school resources. Economics of Education Review 18, 241–252.

Goldstein, H., Spiegelhalter, D. J., 1996. League tables and their limitations: statistical issues in comparisons of institutional performance (with discussion). Journal of the Royal Stastistical Society A 159 (3), 385–443.

Heckman, J. J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161.

Puhani, P. A., 2000. The Heckman correction for sample selection and its critique. Journal of Economic Surveys 14 (1), 53–68.

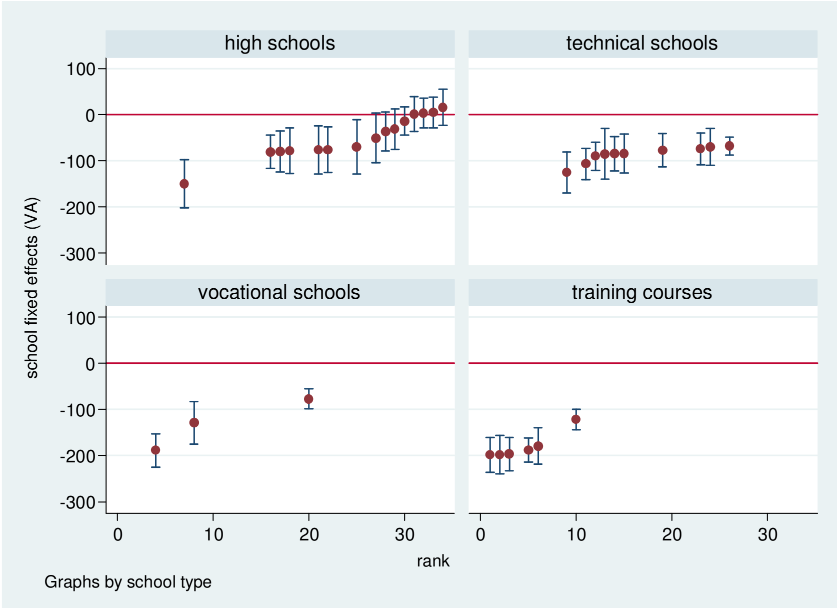
Ray, A. ., 2006. A paper for the oecd project on the development of value-added models in education systems. mimeo. OECD.

Rothstein, J., 2009. Student sorting and bias in value-added estimation: Selection on observables and unobservables. Education Finance and Policy 4 (4), 537–571.

Rothstein, J., 2010. Teacher quality in educational production: tracking, decay, and student achievement. Quartely Journal of Economics, 175–214.

Todd, P., Wolpin, K. I., 2003. On the specification and estimation of the production function for cognitive achievement. Economic Journal 113 (485), F3–33.

Figure 1: School fixed effects estimated using the SFE model

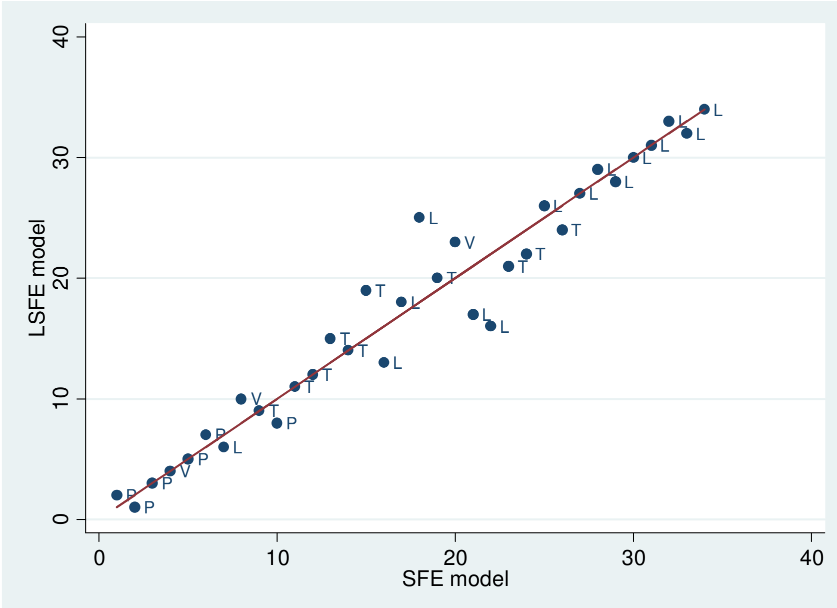


Notes. This figure shows the point estimates and the confidence intervals for the school fixed effects obtained using the School Fixed Effects (SFE) model.

Figure 2: School fixed effects estimated using the LSFE model

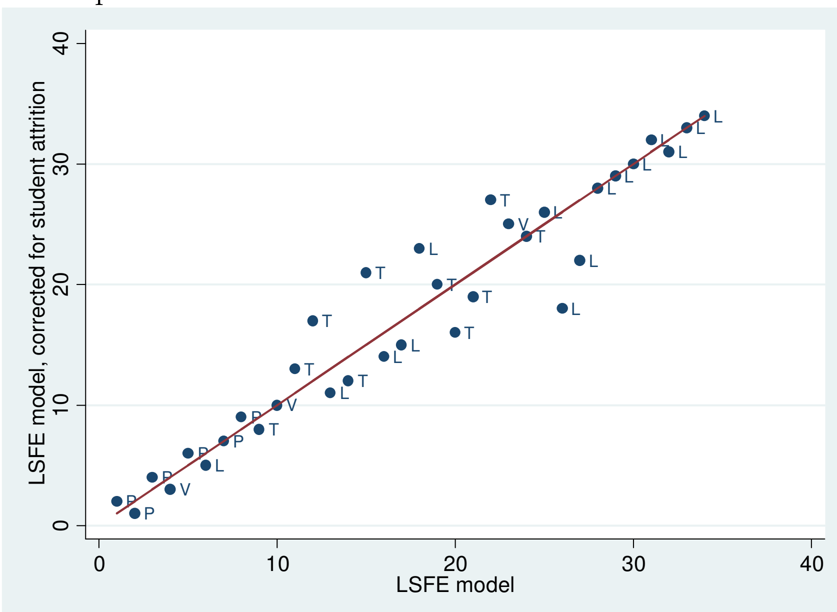
Notes. This figure shows the point estimates and the confidence intervals for the school fixed effects obtained using the Longitudinal School Fixed Effects (LSFE) model.

Figure 3: Cross-plot of LSFE and SFE rankings



Notes. This figure reports a cross-plot of the school rankings obtained using the LSFE and the SFE models. L=*liceo*, T=technical school, V=vocational school, P=professional school.

Figure 4: Cross-plot of LSFE and LSFE corrected for student attrition rankings



Notes. This figure reports a cross-plot of the school rankings obtained using the LSFE with and without correction for student attrition. L=*liceo*, T=technical school, V=vocational school, P=professional school.

Table 1: Reading literacy levels (means and standard deviations) by geographic areas

	statistics	high school	technical school	vocational school	training course	all
Trento	mean	563.69	510.6	472.73	414.57	507.5
	S.D.	62.69	61.77	70.35	71.98	86.49
	N. obs.	575	425	136	311	1447
Valle d'Aosta	mean	560.37	513.58	464.01	424.71	517.11
	S.D.	65.33	61.37	71.78	55.63	81.34
	N. obs.	432	121	283	35	871
Rest of North East	mean	562.38	508.98	459.41	414.16	507.27
	S.D.	63.38	62.61	78.08	75.55	85.97
	N. obs.	2019	1558	947	737	5261
Rest of North West	mean	562.28	503.58	435.11	401.96	512.36
	S.D.	64.35	67.95	82.89	73.58	88.19
	N. obs.	2031	1343	824	235	4433
Centre and South	mean	532.8	457.98	399.31	365.6	480.19
	S.D.	66.7	72.69	75.35	67.55	88.85
	N. obs.	8819	5934	3796	219	18768
Italy	mean	543.56	476.08	418.47	405.7	491.78
	S.D.	67.19	73.77	81.02	74.91	89.16
	N. obs.	13876	9381	5986	1537	30780

Table 2: Family background by school track and geographic areas

	highest parents'	high school	technical school	vocational school	training course	all
Trento	HISEI	51.1	43.83	44.3	37.12	45.39
	year of education	13.99	13.01	13.64	12.28	13.31
	ESCS	0.2	-0.22	-0.12	-0.61	-0.13
Valle d'Aosta	HISEI	52.95	45.02	41.94	40.66	47.81
	year of education	13.86	12.7	12.23	11.2	13.07
	ESCS	0.21	-0.23	-0.46	-0.73	-0.11
Rest of North East	HISEI	53.92	45.25	43.33	38.42	47.33
	year of education	14.14	12.71	12.51	12.23	13.16
	ESCS	0.33	-0.22	-0.36	-0.59	-0.08
Rest of North West	HISEI	55.96	45.42	41.96	36.42	49.18
	year of education	14.39	12.88	12.42	11.58	13.42
	ESCS	0.43	-0.18	-0.47	-0.79	0.01
Centre and South	HISEI	52.55	42.66	38.65	35.77	46.48
	year of education	13.86	12.32	11.72	11.29	12.92
	ESCS	0.28	-0.34	-0.65	-0.87	-0.12
Italy	HISEI	53.2	43.58	40.15	37.53	47
	year of education	13.98	12.5	12.01	11.98	13.05
	ESCS	0.3	-0.29	-0.56	-0.67	-0.09

Table 3: Variance decomposition of reading literacy levels by geographic area

	variance explained by	
	school type FE (R^2)	school FE (R^2)
Trento	0.43	0.55
Valle d'Aosta	0.33	0.50
Rest of North East	0.38	0.54
Rest of North West	0.37	0.57
Centre and South	0.37	0.57
Italy	0.34	0.58

Table 4: Scheme for booklet distribution in the PISA 2010 re-test

Booklet	Pisa 2009				Booklet	Pisa 2010			
1	M1	R1	R3	**M3**	7	R6	**M3**	S3	R4
2	R1	**S1**	R4	R7	3	**S1**	R3	M2	S3
3	S1	R3	**M2**	S3	5	R4	**M2**	R5	M1
4	R3A	R4	**S2**	R2	9	M2	**S2**	R6	R1
5	R4A	**M2**	R5	M1	11	M3	R7	R2	**M2**
6	**R5**	R6	R7	R3	13	S3	R2	R1	**R5**
7	R6	**M3**	S3	R4	1	M1	R1	R3	**M3**
8	R2	M1	**S1**	R6	2	R1	**S1**	R4	R7
9	M2	**S2**	R6	R1	4	R3	R4	**S2**	R2
10	S2	R5	M3	**S1**	8	R2	M1	**S1**	R6
11	**M3**	R7	R2	M2	10	S2	R5	**M3**	S1
12	**R7**	S3	M1	S2	6	R5	R6	**R7**	R3
13	**S3**	R2	R1	R5	12	R7	**S3**	M1	S2

Notes. Green cells correspond to the sections commmon to both the 2009 and the 2010 tests. Each booklet section is identified by a letter indicating the typology (M for maths, S for sciences and R for reading) and a numeric value.

Table 5: Schools' probability to participate to the Trento's 2010 Re-test

	(1)	(2)
PISA 2009 score (SD's)	0.131	-0.034
	(0.095)	(0.191)
School type: high school		
technical school		-0.105
		(0.181)
vocational school		-0.226
		(0.288)
training course		-0.289
		(0.311)
N. observations	50	50

* ** *** significant at 10, 5 and 1 percent, respectively.

Notes. The table reports marginal effects for schools' probability of participating to the Trento's 2010 Re-test estimated with a probit model. Standard errors in parantheses. The PISA 2009 score (mean of the 5 PVs) is expressed in standard deviations (100 PISA points).

Table 6: Students' probability to participate to the Trento's 2010 Re-test

	(1)	(2)	(3)
PISA 2009 score (SD's)	0.058***	0.064***	0.056***
	(0.011)	(0.013)	(0.016)
age in months		-0.061	-0.063
		(0.046)	(0.046)
Gender: female			
male		-0.052	-0.048
		(0.027)	(0.031)
Household structure: nuclear			
single parent family		-0.080*	-0.078*
		(0.036)	(0.036)
mixed family		-0.029	-0.036
		(0.104)	(0.105)
Immigrant status: native			
second-generation immigrant		-0.102	-0.097
		(0.084)	(0.085)
first generation immigrant		-0.004	-0.008
		(0.049)	(0.049)
HISCED: isced 5a, 6			
isced 2		0.051	0.052
		(0.052)	(0.051)
isced 3b, c		-0.017	-0.018
		(0.053)	(0.054)
isced 3a, 4		0.002	0.000
		(0.031)	(0.032)
isced 5b		0.003	0.008
		(0.077)	(0.078)
Day of the week: monday-tuesday			
wednesday-thursday		-0.063*	-0.063*
		(0.029)	(0.031)
friday-saturday		-0.078**	-0.085***
		(0.025)	(0.026)
hisei		-0.000	-0.001
		(0.001)	(0.001)
School type: liceo			
technical schools			-0.016
			(0.039)
vocational schools			-0.062
			(0.047)
training courses			-0.036
			(0.041)
N. observations	942	942	942

* ** *** significant at 10, 5 and 1 percent, respectively.

Notes. The table reports marginal effects for students' probability of participating to the Trento's 2010 re-test estimated with a probit model, conditional on their schools having accepted to participate. Reference categories in parentheses. Standard errors in parantheses are clustered by school. The PISA 2009 score (mean of the 5 PVs) is expressed in standard deviations (100 PISA points).

Table 7: Student performance equations

	CVA model (1)	SFE model (2)	LSFE model (3)
PISA 2009 score (mean 5 PVs)			0.275***
			(0.061)
age in months	26.355*	6.624	1.654
	(12.956)	(11.904)	(11.611)
Gender: female			
male	-16.685	-0.188	0.723
	(9.531)	(7.201)	(7.143)
Household structure: nuclear			
single parent family	-4.013	2.834	0.029
	(12.620)	(11.751)	(11.477)
mixed family	-36.552	-28.012	-23.586
	(47.782)	(44.355)	(45.069)
Immigrant status: native			
second-generation immigrant	-30.784	9.285	10.423
	(28.083)	(35.346)	(33.451)
first generation immigrant	-77.752***	-52.992***	-32.646*
	(17.438)	(13.025)	(14.764)
HISCED: isced 5a, 6			
isced 2	-25.190	12.256	10.906
	(14.529)	(15.867)	(15.412)
isced 3b, c	-11.680	14.573	17.812
	(16.996)	(15.477)	(14.837)
isced 3a, 4	7.509	16.933	15.029
	(11.119)	(11.173)	(10.785)
isced 5b	-44.582*	2.767	7.576
	(20.180)	(18.524)	(18.777)
hisei	1.025**	-0.098	-0.156
	(0.336)	(0.282)	(0.296)
school fixed effects	no	yes	yes
N. observations	756	756	756

* ** *** significant at 10, 5 and 1 percent, respectively.

Notes. All regressions are estimated using 5 plausible values (PVs) and 80 balanced repeated replications (BRR). CVA, SFE and LSFE stand for ‘Contextual Value Added’, ‘School Fixed Effects’ and ‘Longitudinal School Fixed Effects’, respectively.

Table 8: Student performance equations, correcting for panel attrition

	(1) Performance equation	(2) selection equation
PISA 2009 score (mean PVs)	0.241***	0.002**
	(0.055)	(0.001)
age in months	6.229	-0.201
	(9.658)	(0.156)
Gender: female		
male	1.597	-0.183
	(5.903)	(0.096)
Household structure: nuclear		
single parent family	10.242	-0.242
	(11.249)	(0.144)
mixed family	-1.464	0.094
	(42.026)	(0.452)
Immigrant status: native		
second-generation immigrant	-1.006	-0.430
	(24.987)	(0.295)
first generation immigrant	-30.192**	-0.043
	(11.520)	(0.185)
HISCED: isced 5a, 6		
isced 2	-0.259	0.185
	(8.699)	(0.170)
isced 3b, c	19.323*	-0.042
	(8.757)	(0.199)
isced 3a, 4	10.450	-0.029
	(5.478)	(0.111)
isced 5b	0.838	-0.042
	(14.848)	(0.256)
hisei	0.124	0.000
	(0.253)	(0.003)
Day of the week: monday-tuesday		
wednesday-thursday		-0.191
		(0.113)
friday-saturday		-0.347***
		(0.087)
school fixed effects	yes	no
ρ	-0.793***	
	(0.106)	
Wald test instruments	10.42[0.0055]	
N. observations	942	

* ** *** significant at 10, 5 and 1 percent, respectively.

Note. For both the prformance equation and the selection equation (probit), coefficients' estimates are reported in column (1) and (2).