ISSN 1503-299X

WORKING PAPER SERIES

No. 7/2011

School Facilities And Student Achievements: Evidence From The Timss

Arnt Ove Hopland

Department of Economics

Norwegian University of Science and Technology
N-7491 Trondheim, Norway
www.svt.ntnu.no/iso/wp/wp.htm

SCHOOL FACILITIES AND STUDENT ACHIEVEMENTS: EVIDENCE FROM THE TIMSS*

Arnt Ove Hopland
Department of Economics
Norwegian University of Science and Technology
NO-7491 Trondheim
Norway
e-mail: arnt.hopland@svt.ntnu.no

Abstract
This paper studies the link between school facilities and student achievements in eight countries using data from the TIMSS 2003 database. OLS and propensity score matching is used to control for observable characteristics. Both methods indicate that poor school facilities may be negatively associated with student achievements, but the estimated coefficients are mainly insignificant. Significantly negative estimates are found in only three out of eight countries when using OLS. When using matching on propensity scores I only find significant coefficients in one of the countries.

JEL classification: I20, I21, I22
Keywords: Educational production, school facilities, matching

Version of
June 13, 2011

---

* I would like to thank participants at the 33rd Annual Meeting of the Norwegian Economic Association and seminar participants at the Norwegian University of Science and Technology (NTNU) for useful suggestions. I'm also grateful for comments on earlier drafts of the paper from Manudeep Singh Bhuller, Lars-Erik Borge Bjarne Strøm and John Marius Ørke. This paper is part of a project funded by the Norwegian Ministry of local government and regional development. The author bears the full responsibility for the analysis and the conclusions that are drawn.

1. Introduction

All modern societies spend vast resources on educating their young. It is widely acknowledged that spending on education is an important investment in future human capital, which is increasingly important for economic growth in increasingly advanced societies. However, there is not much consensus regarding which kind of spending that most effectively will improve student achievements (and thus increase future human capital) or even if simply increasing resources in itself will have desirable effects.[1] Given scarcity of resources it is important to perform policy evaluation in order to study how to spend in the most effective way. The activities in schools may be interpreted as a "production" of future human capital. Given this interpretation, a popular approach within the economist literature has been to formulate and estimate school production functions seeking to identify causal effects from different capital factors. This is a very useful approach because, if successfully implemented, it may give policy makers guidance as to which policies yield most benefits in form of improved student achievements.

A good example of evaluation of resource use is the very rich literature studying effects from reducing the student/teacher ratio (i.e. reducing class size). In the production function analogy this will in essence be to increase the human capital intensity in the school production. A few of the contributions to this literature are seminal works by Krueger (1999) and Angrist and Lavy (1999) and a more recent study by Leuven et al. (2008). These are important studies because if reducing class size in fact boosts student achievements, it will be a fairly easy policy to implement (at least in the long run)[2] for policy makers. However, there are also obviously high costs associated with such a policy, since the need for more teachers and instructional space will lead to a cost increase roughly proportional to the reduction of class size. Notably, the literature is not conclusive with respect to the magnitude of effects from class size reductions. Some studies identifies significantly positive effects by reducing class size (e.g. Krueger, 1999 and Angrist and Lavy, 1999), whereas other studies reject even small

[1] The debate between Hanushek and Card and Krueger on the effects of resources gives a good overview of the debate in the general school spending literature. Hanushek (1996) reviews more than 90 studies and concludes that simply increasing resources leaves little hope for improved student achievements. However, Hanushek's interpretation of the literature is disputed by Card & Krueger (1996).

[2] In the short run one may be faced with a restriction due to shortage of teacher personnel. However, in the long-run more teachers can be educated and this constraint will be less of a problem.

effects (e.g. Leuven et al, 2008). Thus one should also consider if there are other ways to spend resources that may improve student achievements.

Interestingly, effects from the real capital components of the school production function have not been subject to much empirical study by economists. Similarly to reduction of the student/teacher ratio, improving infrastructure in a large scale may involve substantial financial costs. If schools are in poor condition initially, the need for major construction work gives that one will have large investments costs in the short-run. Where school building conditions are not critically poor initially however, it is possible that school owners to some extent may be able to improve long-run infrastructure conditions by increasing their maintenance expenditures slightly. This gives that a "better school facilities policy" does not necessarily imply very large expenses (except for some initial investments where this is necessary). Thus, the long-run cost increase from this policy is not necessarily very high.[3] Furthermore, improving the infrastructure will also be fairly easy to organize for policy makers if it turns out to be an effective policy.

Intuitively one would assume that in highly developed countries with no direct shortage of school infrastructure, human capital in form of teaching staff will have a greater impact than the condition of school facilities. However, the condition of school facilities has been a widely debated topic in the popular debate in several countries, indicating that both voters and policy makers consider this to be an important issue. An often observed argument is that poor school facilities can have negative consequences for the students' learning environment. The present paper is inspired by this debate and aims to test whether poor quality of school facilities are associated with poor student achievements in eight OECD countries. Some studies suggest that improving environmental conditions may gain student achievements by reducing distractions and missed school days (literature reviewed by Earthman (2002) and Mendell and Heath (2005)). This may also benefit teachers by improving their morale and reducing absenteeism and turnover, indirectly affecting student achievements (Buckley et al. 2005).

[3] In fact, it is often claimed that increasing maintenance expenditures today, will actually reduce the long-run costs due to a slower deterioration rate of the infrastructure. See Borge and Hopland (2011), for a discussion about maintenance and building conditions in Norwegian local governments.

The closest relative to this paper is the recent study by Cellini et al. (2010). They study broad effects from investments in school facilities in Californian school districts using a regression discontinuity approach. Their findings suggest that investment in school facilities may have some positive long-run effects on student achievements, but that there are no effects in the short run. Note that their estimates are imprecise and their evidence in favor of long-run effects is not unambiguous. Thus they conclude that there is, at best, weak evidence in favor of the hypothesis that increased investments in school facilities will boost student achievements even in the long run.

Even though their results are not unambiguous, it will still be reasonable to expect that the effects from physical work conditions (and investments in such) will be stronger in the long run than in the short run. It is unlikely that performances will have a sudden boost when school facilities are improved, since a student's performances in earlier years obviously will affect his performances in the years ahead. Thus it will take some time before effects from the improved facilities are observed. We should also keep in mind that those students who are enrolled after the investment period will only have been exposed to the good facilities. If school facilities matter, these new students should then, all other things equal, perform better than students in earlier cohorts. Thus, the finding that effects from investment in school facilities on student achievements (if any at all) are stronger in the long run than in the short run should not come as a major surprise. In this paper I will study the link between school facility conditions and fourth grade students' achievements, using data from the TIMSS. The TIMSS also includes a test performed in eight grade, but in most countries these students have recently progressed to secondary school. Thus, studying these would to a large extent be to measure effects from the facilities in their primary school, which we do not observe, rather than their present school.

My main contribution to the literature is to look directly at the condition of school facilities rather than investments in such. Investments will obviously be correlated with the condition of the facilities, but far from perfectly, since the daily maintenance expenditures will also be an important determinant for the condition of building facilities. Investments may also capture other factors, such as relative price differences or taste regarding the trade off between maintenance and investments. Thus, where investments in school facilities may serve as a proxy for the quality of the school facilities, this paper utilizes a direct measure of the quality.

I will combine OLS regressions and matching based on propensity score in order to control for observable characteristics.

Studies based on data from international tests, such as the TIMSS, have an advantage compared to studies using solely national data because of the possibility to compare results for several countries. One drawback of using international tests compared to national data is that national data are typically more detailed than the survey data connected with international tests. However this drawback need not be all that important since the TIMSS questionnaire provides useful information on students, parents, schools and school districts. A potentially more critical drawback for the analysis in this paper is that the TIMSS does not offer any credible instrumental variables to battle potential bias due to unobservable characteristics that may be correlated with both student achievements and the quality of the school facilities. This is potentially critical since both OLS and matching will be sensitive to unobservable characteristics.

However, it is not given in which direction such may bias the estimates, as illustrated by the following examples. It is unlikely that all characteristics of a good teacher are observable in the data. Thus, if good teachers have a positive effect on student achievements and sort themselves into schools with good buildings, OLS will tend to overestimate negative effects from poor facilities. A similar effect will occur if resourceful parents sort their children into schools with good building conditions, since it is unlikely that the controls capture all relevant characteristics of the family background and peer effects. Compensatory or regressive policies are other potential causes of bias. If policy makers believe that school facilities are important for student achievements, this may lead them to upgrade school facilities where achievements are low. This draws towards underestimation of negative effects from poor school facilities. Regressive policies could occur if politicians observe that voters in school districts with poor student achievements are less likely to vote and will tend to bias the estimates in the opposite direction. Unfortunately, the TIMSS database does not provide any credible instrumental variables that could solve these problems. The best I can offer is an alternative approach based on matching on propensity scores. This method offers an alternative way to control for observable characteristics and differ from OLS regressions. However, like OLS it is sensitive to unobservable characteristics and does therefore not solve the omitted variable bias that may potentially be plaguing the OLS estimates. Thus, one must keep in mind that some caution is required when interpreting the results of this study. The estimated coefficients in this paper are more a reflection of a potential association, rather than robustly identified causal effects.

The remainder of the paper is organized as follows: In Section 2 we take a first look at the data. The empirical strategies, OLS and propensity score matching are discussed in Section 3 before the results are presented in Section 4. The findings of the paper are summarized in Section 5.

2. A first look at the data

The dataset consists of fourth grade students[4] from four Western European, ((Flemish) Belgium, Great Britain (GB in the tables), Italy and The Netherlands) and four non-European countries (Australia, Japan, New Zealand (NZL in the tables), and the USA) representing a broad specter of highly developed 'western-like' countries. In Great Britain, schools from England and Scotland are reported separately to the TIMSS database. However, due to missing variables, we are left with rather few observations in the multivariate analysis when treating them separately (roughly 60 schools in each). Thus, I choose to analyze the British data pooled, rather than separating the Scotsmen from the English. Fourth grade students from Norway also participated in the TIMSS 2003. However, the Norwegian data is missing some important socio-economic control variables and I therefore have chosen to not include the Norwegian schools in the analysis.

The non-European countries in the analysis have similar attributes (rich and democratic market economies) as the Western European countries. In order to have a ground for comparison across countries one should have countries that are fairly similar at least with respect to economic conditions. This is because principals in very poor countries may have very different perceptions of what is to consider as poor or good facilities than their colleagues in richer countries.

[4] The countries report results from the cohorts with the largest share of nine and thirteen year old students. For most countries this is the grades four and eight respectively and even though there is some variation across countries, the TIMSS simply refers to fourth and eight grade.

Each country draws the participating schools from a stratified sample, to ensure a representative sample of schools. Further, within each of the schools, generally one class from the fourth grade is randomly chosen to get a representative student sample[5]. The database, in addition to test results, also includes rich information from questionnaires answered by students, teachers and principals. From these I obtain the key explanatory variables and control variables used in the analysis. The key explanatory variables are based on the following question from the principals' questionnaire: *Is your school's capacity to provide instruction affected by a shortage or inadequacy of school buildings and grounds?* (Question 23c). The following answers are possible: 1: none, 2: a little, 3: some or 4: a lot.

In the econometric specification I will use two different formulations based on this index. Firstly, I will apply a flexible formulation where I include dummies for each of the categories, using the best buildings as reference category. Secondly, I will introduce a poor facilities dummy (pbuild) which equals one if the school is reported to be in category 2, 3 or 4. We then simply compare all schools with good facilities to those which have unsatisfying facilities, asking: Are poor school facilities associated with poor student achievements?

Table 1 About here

Table 1 summarizes the percentage of schools reported to be in each of the categories and descriptive statistics for the poor facilities dummy, which is simply one minus the share of schools in the best category. We observe that there is sufficiently spread along the categories to give meaningful variation, even though all countries have a majority of their schools reported in one of the two best categories. Japan and New Zealand have the highest proportion of schools in one of the two best categories, with 88 % and 86 % of the schools respectively. In the other end we find Italy (52 %) and the Netherlands (56 %).

The observation that most schools are in fairly good condition is not very surprising given that the study targets wealthy countries exclusively. Since it in the analysis is only the very best category that is used as reference we also note that New Zealand has by far most schools in this category (62 %) in front of the USA (54 %) and Japan (51 %). Italy is dead last with only 26 % of the schools reported to be in the best category.

[5] Further details on the practical sampling can be found in Beaton et al. (1996) and Martin and Kelly (1997).

In TIMSS, student performance in mathematics and science is tested separately using international achievement scores with an international mean of 500 and an international standard deviation of 100. Summary statistics for the test scores are reported in the upper part of Table 2. I will transform the results into ‘school contributions’. The transformation takes place by running regressions where school fixed effects are included together with variables describing the student’s individual and family characteristics as presented in Equation (1).[6] Similar approaches are used by Borge and Naper (2006), Fiva and Rønning (2008) and Naper (2010).

$$y_{ij} = \mathbf{Z}_{ij}\boldsymbol{\beta} + \alpha_j + u_{ij} \tag{1}$$

y_{ij} is the test score for student *i* in school *j*, the vector $\mathbf{Z}_{ij}$ includes observable student characteristics and family background, α_j represents school fixed effects and u_{ij} is an error term. The student characteristics include a dummy which equals one if the student is a girl and a dummy indicating whether the student is native to the country. The family background variables include a dummy indicating whether the student’s father is a foreigner and an index indicating the approximate number of books in the family’s home. The latter of these are to be considered as a proxy for the educational level in the home and the socio-economic status of the family. Descriptive statistics and a closer description of the variables are given in the Appendix.

The school fixed effects (the α’s) may be interpreted as the average test score in the school, adjusted for student characteristics and family background. In this paper the expressions *adjusted test scores* and *school contribution* are synonymous, and both refer to the school fixed effects. Since the observable individual and family characteristics are accounted for in the adjusted test score they will play no role in the further analyses. Note that using estimated variables in the regressions will in general introduce measurement errors. Measurement errors in the dependent variable will increase the variance of the OLS estimates, but not introduce additional bias. Descriptive statistics for the adjusted test scores are presented in the lower parts of Table 2.

[6] The estimations of the school contributions are reported in the Appendix.

The averages are zero by definition since the average school is the benchmark, while poorer and better schools have negative and positive effects on student achievements respectively. Note however that the standard deviation varies significantly between the countries and also to some extent between the tests within the countries. While Belgium, Japan and the Netherlands have standard deviations below or just over 20 for both adjusted test scores, the standard deviations are in the 30s in Australia, Great Britain and the USA. Italy and New Zealand have the highest standard deviations on both adjusted test scores, with standard deviations well above 40.

In Australia, Belgium, Great Britain, Japan and the Netherlands we observe a higher standard deviation in adjusted test scores in mathematics than in science while the opposite is observed in the USA. In Italy and New Zealand we observe almost identical standard deviations on the two adjusted test scores. Importantly, these figures illustrate that there is still a considerable amount of variation to analyze, even though we are now only considering the school contributions rather than the raw test scores.

Table 2 About here

3. Empirical strategies

I start out by estimating a standard school production function using OLS

$$\alpha_j = \gamma_0 + \boldsymbol{b}_j \gamma_b + \mathbf{X}_j \boldsymbol{\gamma}_x + u_j \tag{2}$$

where α_j is the adjusted test scores, the vector $\boldsymbol{b}_j$ includes the measure(s) of school facilities and the vector $\boldsymbol{\gamma}_b$ includes the coefficient(s) of interest. Further, $\mathbf{X}_j$ is a vector of school specific controls. The vector includes variables describing teacher and school district characteristics. The information about the school and school district includes a variable indicating the number of inhabitants and the size of the cohort. To capture the general resource situation in the school a dummy indicating whether the school suffers from a shortage of teachers is included. This is constructed the same way as the building condition dummy from a similar index.

As measures of teacher characteristics I use the length of the teacher's education, a dummy indicating whether or not the teacher has a licence to teach and the teacher's tenure. To account for peer group effects I include the share of economically disadvantaged and economically affluent families with children in the school. Finally, I include the number of students in the mathematics or science class, depending on which test is analyzed. Descriptive statistics and closer definitions of the controls are presented in the Appendix.

Further I will use matching based on propensity score, as suggested by Rosenbaum and Rubin (1983) to check the robustness of the OLS estimates. This method will only be applied on the one-dummy formulation, following the algorithm programmed by Becker and Ichino (2002). Following the terminology from the evaluation literature, the outcome studied is the school contribution, α_j for school *j*. Having poor facilities is considered as a treatment. School *j* either has $(\text{pbuild}_j = 1)$ or $(\text{pbuild}_j = 0)$. The school contribution for school *j* is denoted $\alpha_j(1)$ if the school has poor facilities, and $\alpha_j(0)$ if the school has good facilities.

My primary interest is whether poor facilities affect the school contribution, i.e. the difference $\alpha_j(1) - \alpha_j(0)$. The fundamental problem is that we do not observe both $\alpha_j(1)$ and $\alpha_j(0)$ for the same school. The statistical challenge is the possible sample selection bias since schools with good facilities may not be representative of those with poor facilities in the counterfactual situation with good facilities. Decomposition of the raw comparison of average/expected school contributions for schools with poor and good school facilities, clarifies the selection bias:

$$\begin{aligned} &E\left[\alpha_j(1) \middle| \text{pbuild}_j = 1\right] - E\left[\alpha_j(0) \middle| \text{pbuild}_j = 0\right] = E\left[\alpha_j(1) - \alpha_j(0) \middle| \text{pbuild}_j = 1\right] \\ &+ \left\{E\left[\alpha_j(0) \middle| \text{pbuild}_j = 1\right] - E\left[\alpha_j(0) \middle| \text{pbuild}_j = 0\right]\right\} \end{aligned} \tag{3}$$

The first term on the right-hand side defines the treatment effect of interest, i.e. the average treatment effect of poor school facilities in the schools with poor facilities. This is referred to as the average treatment effect on the treated (ATT). The second term reflects the bias that occurs if the school contribution of schools with good facilities is not representative for the school contribution of the schools with poor facilities if their facilities were good.

If the assignment of schools into treatment is random, the second term equals zero and there is no such bias. However, the economic priorities which determine the quality of school facilities are not random, and there may be systematic differences between schools with poor and good facilities. Propensity score matching and OLS are two different solutions to this problem, since they allow us to control for observable characteristics.

As OLS regressions, matching assumes selection on observables. However, Angrist (1998) compares the methods regression and matching and shows that the methods yield different results. The differences occur because the observations are weighted differently, hence the methods will generate different results, even when controlling for the same characteristics.

Technically the difference origins from that while the estimated coefficients from regressions reflect variance-weighted averages, the matching estimator generates weights that are proportional to the probability of the treatment (here: poor school facilities) given the observed characteristics. Thus there may be some value added in applying matching in addition to OLS, even though it does not solve potential problems due to unobservable characteristics.

Matching is frequently used to evaluate policy programs, and is based on a comparison of treated and non-treated observations. As opposed to most policy evaluation I have no pre-treatment observations, since I only observe school facilities at a fixed point in time. However, this is not a critical shortcoming for the analysis. Persson (2001), Persson and Tabellini (2002) and Borge and Rattsø (2008) also use matching in situations with no pre-treatment observations.

The main idea in matching is to approach the evaluation of causal effects the same way as in a controlled experiment. The data is therefore split into two groups; 'treated' $(\text{pbuild}=1)$ and 'non-treated' or 'control' $(\text{pbuild}=0)$ observations. The unobservable counterfactual outcome for a given treated observation is then estimated from the outcome among otherwise similar, but non-treated observations.

When the schools that are compared are similar, the selection with respect to the treatment should to a large extent be random, as in an experiment. Successful matching thus removes bias due to systematic selection. The key assumption for the matching analysis is that selection into treatment depends only on the vector of observables, $\mathbf{X}_j$. Alternatively, selection into poor school facilities is random conditioned on the observables. If this is fulfilled we obtain

$$E\left[\alpha_j(0)\middle|\text{pbuild}_j=0,\mathbf{X}_j\right]=E\left[\alpha_j(0)\middle|\text{pbuild}_j=1,\mathbf{X}_j\right] \tag{4}$$

where α_j is the school contribution in school *j*. Given (3) we can define the average treatment effect on the treated (ATT) as

$$ATT=E\left\{E\left[\alpha_j(1)\middle|\text{pbuild}_j=1,\mathbf{X}_j\right]-E\left[\alpha_j(0)\middle|\text{pbuild}_j=0,\mathbf{X}_j\right]\middle|\text{pbuild}_j=1\right\} \tag{5}$$

The outer expectation in Equation (5) is over the distribution of the characteristics of the schools with poor facilities. The counterfactual school contribution for a specific school with poor facilities can be estimated from the outcome for schools with good facilities, but similar characteristics.

The remaining problem is that $\mathbf{X}_j$ may contain several continuous controls and that this problem with dimensionality is likely to make the matching strategy impossible to implement in practice. However Rosenbaum and Rubin (1983) show that if conditioning on $\mathbf{X}_j$ eliminates selection bias, then conditioning on $p\left(\mathbf{X}_j\right)$, where *p* is the probability of having a school with poor facilities achieves the same:

$$ATT=E\left\{E\left[\alpha_j(1)\middle|\text{pbuild}_j=1,p\left(\mathbf{X}_j\right)\right]-E\left[\alpha_j(0)\middle|\text{pbuild}_j=0,p\left(\mathbf{X}_j\right)\right]\middle|\text{pbuild}_j=1\right\} \tag{6}$$

Schools with the same probability of having poor facilities will have the same distribution of the full vector of control variables. This probability is called the propensity score. The propensity score can be estimated using any standard model for estimation of probabilities (here: probit). In order to avoid a biased estimate for the propensity score, the vector of explanatory variables in the probit should consist of those variables that are expected to affect both the outcome (adjusted test scores) and the probability of treatment (poor school facilities). Hence, I include all the control variables from the OLS analysis in the propensity score equation (the same vector $\mathbf{X}_j$ is used in both analyses).

The results from the estimation of the propensity score are reported in the Appendix. We observe that the vector $\mathbf{X}_j$ do not have much explanatory power and that most variables included are insignificant. Importantly, we have that the balancing property is satisfied for all countries. The balancing property of the probit specification is essential for the comparison of test results. The test checks whether the explanatory variables are different for schools with buildings in good condition compared to those with poor buildings given that they have approximately the same propensity score.

The first step is to stratify all schools into blocks such that the propensity score does not differ significantly between schools with poor and good buildings within each block. The second step is to test whether the mean of the explanatory variables differ significantly between treated and non-treated observations within each block. If they do not, the balancing property is satisfied.

4. Results

Table 3 About here

Table 4 About here

Tables 3 and 4 present results from OLS regressions. In the regressions in the upper part, I have included dummies for each of the three least favorable categories, leaving the schools with the best facilities as reference. In the lower parts of the table I include the poor facilities dummy as the key explanatory variable. We observe from Tables 3 and 4 that the dummies for each of the non-optimal categories are in general negative, but mostly insignificant. Interestingly, it does not appear to be a linear development, where the adjusted test scores become gradually lower as the facilities deteriorate. Thus, including the facilities index linearly in the regressions will not be a valid simplification. However, it does seem that the one-dummy formulation is a reasonable simplification of the model.

Given the intuitive interpretation, and the fact that it seems to be a reasonable simplification, I will base the remainder of the discussion on the results from the one-dummy formulation. Interestingly, we observe that the coefficients for the poor facilities dummy do not change dramatically when moving from the simple regressions (Table 3) to the regressions where the rich vector of controls are included (Table 4). That the coefficient for the poor facilities dummy is not heavily biased by exclusion of observable characteristics of the controls is interesting. This may also indicate that it is not heavily biased by unobservable characteristics of these. However this observation is, of course, not to be interpreted as evidence that OLS provides unbiased estimates.

When using the poor facilities dummy in the multiple regressions I find that for 5 out of 8 countries there are no significant coefficients The results in these countries differ quite a lot and are also very imprecisely estimated, typically with t-values well below one in absolute value. Thus, it is for these countries difficult to identify any clear patterns.

When estimating the Japanese test scores I only find significant effects from the poor facilities dummy on the adjusted test scores in science. The coefficient value indicates a negative treatment effect from poor facilities of roughly 32 % of a standard deviation. However, even though the effect seems strong in terms of standard deviations when estimating adjusted science test scores, we also note that the estimated effect on the adjusted test scores in mathematics is not significant. Thus, the results for Japan are also only vaguely supporting the hypothesis that school facilities matter for student achievements.

In the multiple regressions for Australia, I find a significantly negative coefficient for the poor facilities dummy when estimating both adjusted test scores. However, the estimates are not very precise and only significant at the 10 % level. On face value, the coefficient values indicate a treatment effect from poor facilities of about 31 % and 29 % of a standard deviation on adjusted scores in mathematics and science respectively.

When estimating the school contributions in the Dutch sample, I also find significantly negative coefficients for the dummy on both test scores in the multiple regressions. In terms of standard deviation, the effects are roughly 33 % and 34 % for adjusted test scores in mathematics and science respectively. Thus, the Dutch and Australian estimations are the ones that provide the strongest support for the hypothesis, since we observe significantly negative effects from poor facilities on both tests. For these countries there seems to be clearly negative effects from poor school facilities. Interestingly, the effect seems not to be especially sensitive to exclusion of the large set of controls.

Table 5 About here

Finally, Table 5 presents the results from the propensity score matching to test the robustness of the OLS analysis. There are several methods available for testing whether there are significantly different test results between treated and non-treated schools. I have applied the nearest neighbor method as programmed by Becker and Ichino (2002). This method is intuitively appealing since it matches each treated unit with the control unit that has the closest propensity score and all treated units will find a match. The main picture still remains, all estimates are negative or insignificantly positive. The only counter-intuitive positive estimates are found in Great Britain and the USA. We note that these are very imprecise estimates with t-values well below one on all four estimates, and actually less than a half for three out of the four.

Due to the low correlation between the vector $\mathbf{X}_j$ and the facility indicators the matching estimates are in general less precise than the OLS estimates. Thus, the only significant coefficient obtained from this method is when estimating the adjusted test scores in science in the Dutch sample. On face value the negative treatment effect from poor school buildings on adjusted test score correspond to roughly 45 % of a standard deviation. Note however that the effect is not very precisely estimated and is only significant on the 10 % level.

The finding that the vector $\mathbf{X}_j$ is only weakly correlated with the building conditions is also consistent with the findings in the regression analyses. We remember from the regressions that the estimated coefficients for the poor facilities dummy did not change much from the simple to the multiple regressions where the rich vector of controls ($\mathbf{X}_j$) was included. The finding that the OLS estimates are not severely biased when excluding observable characteristics of teachers and school districts may indicate that OLS will also be fairly robust to the exclusion of any unobservable features of these important factors. Unfortunately, I have no credible instruments available and it is therefore not possible to formally test whether this is actually the case.

However, if bias due to unobservable characteristics is not a major issue, both OLS and propensity score matching will provide estimates that are not heavily biased. Further, since the vector $\mathbf{X}_j$ is only weakly correlated with the condition of the school facilities, the OLS estimates will be the most precise. Thus, these should be considered as the main results of the analysis.

Anyway, neither of the methods provides evidence in favor of the hypothesis that poor school facilities are significantly associated with poor student achievements for most of the countries. The exceptions are Australia and Japan where I obtain significant coefficients for the poor facilities dummy when using OLS, and the Netherlands where I find significantly negative estimates when using both OLS and matching. The low significance in most of the countries may come from that the difference between the schools in the different building condition categories is simply too small for it to matter for student achievements.

5. Concluding remarks

This paper attempts to study the importance of the school facilities within a school production function framework. To my knowledge this is the first paper which studies this issue directly and the paper is inspired by the ongoing popular debate about school building conditions in several countries. The analysis focuses on the link between the quality of school facilities and student achievements in eight countries and is based on data from the TIMSS database.

The empirical strategies used are OLS and matching on propensity scores. The methods offer two alternative ways to control for observable characteristics but are both sensitive to sorting on unobservable characteristics. Unfortunately, the TIMSS database does not offer any credible instruments for the quality of school facilities, making the results sensitive to sorting on unobservable characteristics. Thus one should interpret the coefficients in this study with some caution, since I estimate associations rather than strictly identified causations.

The OLS estimates indicate that there may exist a negative link between poor school facilities and student achievements. However, the estimated coefficients are mostly insignificant and the results provide, at best, very weak support for the hypothesis that school facilities affect student achievements. Further a procedure with matching based on propensity score is introduced as an alternative procedure to control for observable characteristics. Similar to OLS, the estimates are mostly negative. However, the precision of the estimates are lower when using matching, leaving only one significantly negative estimate, and thus even less support for the hypothesis than the OLS estimates. The lack of significance may origin from that the difference between the school facilities within the different categories is simply not sufficiently large for them to affect student achievements in a rich country like the ones in this study.

References

Angrist, J. (1998): "Estimating the labor market impact of voluntary military service using social security data on military applicants" *Econometrica* 66, 249 – 288

Angrist, J. D., Lavy V. (1999): "Using Maimonides' rule to estimate the effect of class size on scholastic achievement" *Quarterly Journal of Economics 114*, 533-75

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzales, E.J., Kelly, D.L. Smith, T.A. (1996): *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*, Chestnut Hill, MA: Boston College

Becker, S., Ichino, A. (2002): "Estimation of average treatment effects based on propensity scores" *The Stata Journal* 2, 358 – 377

Borge, L.-E., Hopland, A.O., 2011. "Maintenance and Building Conditions in Norwegian Local Governments: Economic and Political Determinants" Unpublished draft

Borge, L.-E., Naper, L.R., 2006. "Efficiency Potential and Efficiency Variation in Norwegian Lower Secondary Schools", *FinanzArchiv* 62, 221-249

Borge, L.-E, Rattsø, J. (2008): "Property taxation as incentive for cost control: Empirical evidence for utility services in Norway" *European Economic Review* 52, 1035 – 1054

Buckley, J., Schneider, M., Shang, Y. 2005. "Fix It and They Might Stay: School Facility Quality and Teacher Retention in Washington D.C". *The Teachers College Record*. 107, 1107-1123

Card, D., Krueger, A.,B., 1996. "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina". *Journal of Economic Perspectives*, 10, 31-50

Cellini, S.R, Ferreira, F., Rothstein, J., (2010). "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design" *Quarterly Journal of Economics*, 125, 215-261

Earthman, G.I., 2002. "School Facility Conditions and Student Academic Achievement". UCLA's Institute for Democracy, Education and Acces (IDEA) Paper no. wws-rr008-1002

Fiva, J.H, Rønning, M., 2008. "The incentive effects of property taxation: Evidence from Norwegian school districts" *Regional Science and Urban Economics* 38, 49-62

Hanushek, E.A., 1996. "School Resources and Student Performance", in *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Gary Burtless, ed- Washingtion D.C: Brookings Institution

Krueger, A., (1999): "Experimental estimates of education production functions" *Quarterly Journal of Economics* 114, 497 – 532

Leuven, E., Oosterbeek, H., Rønning, M. (2008): Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway" *Scandinavian Journal of Economics* 110(4), 663 – 693

Martin, M.O., Kelly, D.L. (eds) (1997): *Third International Mathematics and Science Study Technical Report, Volume II: Implementation and Analysis – Primary and Middle School Years,* Chestnut Hill, MA: Boston College

Mendell, M. J. and Heath, G. A. (2005), Do indoor pollutants and thermal conditions in schools influence student performance? A critical review of the literature. *Indoor Air*, 15: 27–52

Naper, L.R. 2010. "Teacher hiring and educational efficiency" *Economics of Education Review* 29, 658-668

Persson, T. (2001): "Currency unions and trade: How large is the treatment effect" *Economic Policy* 33, 435 – 448

Persson, T., Tabellini, G. (2002): “Do constitutions cause large governments? Quasi-experimental evidence” *European Economic Review* 46, 908 – 918

Rosenbaum, P., Rubin, D. (1983): “The central role of the propensity score in observational studies for causal effects” *Biometrica* 70, 41 – 55

Tables in main text

Table 1. Descriptive statistics for the key explanatory

Distribution of answers	Australia	Belgium	GB	Italy	Japan	Netherlands	NZL	USA
None (Cat 1)	44 %	49 %	37 %	26 %	51 %	36 %	62 %	54 %
A little (Cat 2)	26 %	23 %	37 %	26 %	37 %	20 %	24 %	23 %
Some (Cat 3)	20 %	16 %	18 %	33 %	8 %	29 %	10 %	17 %
A lot (Cat 4)	10 %	12 %	8 %	15 %	4 %	15 %	4 %	6 %
Poor buildings (pbuild) average	0.56	0.51	0.63	0.74	0.49	0.64	0.38	0.46
(St. dev.)	(0.50)	(0.50)	(0.48)	(0.44)	(0.50)	(0.48)	(0.49)	(0.50)
Observations (schools)	200	146	206	171	150	118	213	221

Table 2. Summary statistics for test scores

	Australia		Belgium		GB		Italy		Japan		Netherlands		NZL		USA	
	Math	Science	Math	Science	Math	Science	Math	Science	Math	Science	Math	Science	Math	Science	Math	Science
Average score	499.55	523.89	551.95	520.35	509.47	519.42	501.32	514.22	565.60	544.70	541.92	526.05	494.33	520.03	512.49	528.74
(Standard deviation)	(75.54)	(75.31)	(54.83)	(49.65)	(82.16)	(79.16)	(78.85)	(80.81)	(70.27)	(68.87)	50.62	(47.69)	(81.60)	(82.03)	(73.81)	(78.13)
Observations	5219	5219	10067	10067	8757	8757	4282	4282	5322	5322	2937	2937	8502	8502	18448	18448
Number of schools	204	204	149	149	248	248	171	171	150	150	130	130	220	220	248	248
	Adjusted math	Adjusted science	Adjusted math	Adjusted science	Adjusted math	Adjusted science	Adjusted math	Adjusted science	Adjusted math	Adjusted science	Adjusted math	Adjusted science	Adjusted math	Adjusted science	Adjusted math	Adjusted science
Average score	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(Standard deviation)	(36.58)	(32.48)	(19.85)	(17.03)	(39.07)	(37.31)	(45.37)	(46.68)	(20.11)	(18.05)	(18.20)	(16.74)	(43.07)	(43.34)	(34.34)	(37.10)
Observations	4480	4480	9592	9592	8285	8285	3852	3852	5014	5014	2675	2675	6473	6473	17510	17510
Number of schools	203	203	149	149	246	246	171	171	150	150	130	130	218	218	248	248

Table 3. Estimation of test results. OLS without controls.

Variables	Australia (A) Adjusted math	Australia (B) Adjusted science	Belgium (C) Adjusted math	Belgium (D) Adjusted science	GB (E) Adjusted math	GB (F) Adjusted science	Italy (G) Adjusted math	Italy (H) Adjusted science	Japan (I) Adjusted math	Japan (J) Adjusted science	Netherlands (K) Adjusted math	Netherlands (L) Adjusted science	NZL (M) Adjusted math	NZL (N) Adjusted science	USA (O) Adjusted math	USA (P) Adjusted science
Cat2 (second best)	-3.097	-3.584	1.633	2.812	-1.673	1.071	-18.24*	-16.15	-1.061	-6.236	-4.174	-4.348	5.271	5.050	-12.62**	-4.046
	(6.521)	(6.094)	(3.076)	(2.680)	(6.781)	(6.296)	(10.23)	(10.86)	(4.758)	(3.920)	(3.362)	(3.707)	(7.875)	(7.598)	(6.305)	(6.853)
Cat3	-19.01**	-16.37**	4.929	3.485	3.626	6.868	3.227	6.983	-2.484	-8.513*	-8.923*	-6.091	0.756	-0.219	2.553	-1.535
	(8.191)	(6.695)	(4.260)	(3.711)	(7.524)	(7.792)	(10.82)	(11.56)	(4.315)	(4.831)	(4.969)	(4.180)	(8.460)	(8.960)	(5.717)	(6.195)
Cat4 (worst)	-14.37**	-7.239	-15.22**	-8.495	-8.264	-9.498	-6.422	-0.136	-9.903	-10.67	-6.710	-5.536	7.103	3.521	-10.16	-10.92
	(6.369)	(6.663)	(6.870)	(5.988)	(22.53)	(22.89)	(12.93)	(13.63)	(7.324)	(6.823)	(5.129)	(4.423)	(12.32)	(12.93)	(12.55)	(12.97)
R-squared	0.043	0.034	0.085	0.044	0.006	0.012	0.037	0.037	0.011	0.042	0.039	0.025	0.003	0.003	0.027	0.006
Pbuild	-10.60*	-8.471*	-1.259	0.326	-1.310	0.930	-6.561	-2.803	-2.197	-7.086**	-6.782**	-5.365*	4.072	3.287	-6.272	-4.076
	(5.477)	(4.924)	(3.229)	(2.751)	(6.983)	(6.780)	(9.709)	(10.43)	(3.884)	(3.294)	(3.392)	(2.997)	(6.226)	(6.173)	(4.927)	(5.254)
R-squared	0.020	0.017	0.001	0.000	0.000	0.000	0.004	0.001	0.003	0.038	0.031	0.023	0.002	0.001	0.008	0.003
No. of schools	199	199	146	146	204	204	171	171	150	150	118	118	212	212	221	221
No. of students	4,405	4,405	9,412	9,412	6,889	6,889	3,852	3,852	5,014	5,014	2,447	2,447	6,348	6,348	15,579	15,579

Note: Robust standard errors (adjusted for school level clustering) in parentheses
*** p<0.01, ** p<0.05, * p<0.1.
Constant term (not reported) included.

Table 4. Estimation of test results. OLS.

Variables	Australia		Belgium		GB		Italy		Japan		Netherlands		NZL		USA	
	(A) Adjusted math	(B) Adjusted science	(C) Adjusted math	(D) Adjusted science	(E) Adjusted math	(F) Adjusted science	(G) Adjusted math	(H) Adjusted science	(I) Adjusted math	(J) Adjusted science	(K) Adjusted math	(L) Adjusted science	(M) Adjusted math	(N) Adjusted science	(O) Adjusted math	(P) Adjusted science
Cat2 (second best)	-8.843	-8.860	0.774	2.006	1.132	1.171	-17.80*	-14.75	1.337	-5.270*	-6.654**	-7.594*	-1.243	7.617	-8.809*	1.311
	(7.154)	(6.037)	(3.118)	(2.861)	(6.824)	(6.240)	(10.19)	(10.95)	(4.413)	(3.172)	(3.321)	(3.892)	(6.627)	(4.760)	(4.489)	(4.584)
Cat3	-18.06**	-14.13*	4.943	4.285	6.490	5.690	3.012	6.822	-1.375	-10.58*	-2.683	-1.368	-1.239	-4.096	1.019	-0.110
	(7.785)	(7.335)	(4.236)	(3.642)	(9.697)	(7.858)	(10.79)	(11.39)	(6.495)	(5.689)	(3.947)	(3.921)	(7.265)	(7.073)	(4.967)	(5.089)
Cat4 (worst)	-5.447	-2.446	-2.774	4.622	8.183	0.448	-3.009	3.629	-0.905	-2.324	-9.111*	-8.819*	0.867	-7.209	-13.81	-17.80**
	(7.391)	(6.406)	(5.681)	(4.417)	(9.657)	(9.443)	(14.07)	(14.75)	(5.761)	(6.099)	(5.223)	(4.666)	(11.13)	(9.430)	(8.912)	(9.011)
R-squared	0.308	0.305	0.285	0.270	0.316	0.273	0.116	0.135	0.097	0.220	0.359	0.265	0.445	0.452	0.536	0.584
Pbuild	-11.25*	-9.303*	1.409	3.369	3.029	2.165	-5.622	-1.457	0.695	-5.856**	-5.951**	-5.723*	-1.020	2.131	-6.160	-2.798
	(5.730)	(5.068)	(2.791)	(2.435)	(6.725)	(5.795)	(9.921)	(10.63)	(3.800)	(2.888)	(2.961)	(3.044)	(4.994)	(4.678)	(3.811)	(3.814)
R-squared	0.298	0.297	0.276	0.267	0.312	0.270	0.085	0.102	0.096	0.213	0.347	0.244	0.445	0.441	0.524	0.567
School district controls	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Teacher controls	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Social controls	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
No. of schools	156	155	134	134	119	120	171	171	136	137	96	96	171	156	172	168
No. of students	3,150	3,026	4,360	4,251	3,488	3,465	3,848	3,848	3,843	3,879	1,985	1,985	2,469	2,250	10,562	10,234

Note: Robust standard errors (adjusted for school level clustering) in parentheses
*** p<0.01, ** p<0.05, * p<0.1.
Constant term (not reported) included

Table 5. Propensity score matching analysis. Treatment is pbuild.

Treatment: Pbuild	Australia (A) Adjusted math	Australia (B) Adjusted science	Belgium (C) Adjusted math	Belgium (D) Adjusted science	GB (E) Adjusted math	GB (F) Adjusted science	Italy (G) Adjusted math	Italy (H) Adjusted science	Japan (I) Adjusted math	Japan (J) Adjusted science	Netherlands (K) Adjusted math	Netherlands (L) Adjusted science	NZL (M) Adjusted math	NZL (N) Adjusted science	USA (O) Adjusted math	USA (P) Adjusted science
Nearest neighbor	-3.19	-1.69	-4.08	-0.42	4.47	7.75	-8.97	-0.52	-2.93	-5.42	-6.54	-7.55*	-11.06	-12.02	1.88	3.15
	(10.30)	(9.25)	(4.82)	(4.03)	(10.87)	(10.09)	(13.69)	(13.71)	(4.14)	(4.36)	(4.47)	(4.21)	(8.90)	(8.03)	(7.29)	(7.55)
Treated	88	88	67	67	81	81	127	127	66	66	58	58	68	68	79	79
Non-treated	38	38	39	39	26	26	31	31	28	28	22	22	44	44	43	43
Blocks	5	5	4	4	8	8	5	5	5	5	5	5	4	4	5	5
Bal. prop. (1%)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Common support	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Appendix: Appendix Tables

Table A1. Descriptive statistics. Control variables.

VARIABLES	Australia	Belgium	GB	Italy	Japan	Netherlands	NZL	USA
Books	3.46	3.00	3.21	2.62	2.77	3.10	3.22	2.95
	(1.16)	(1.07)	(1.24)	(1.20)	(1.07)	(1.14)	(1.18)	(1.22)
Observations	5090	9857	8537	4229	5290	2878	8289	18067
Native	0.84	0.93	0.84	0.96	0.98	0.93	0.84	0.79
	(0.36)	(0.25)	(0.36)	(0.20)	(0.12)	(0.25)	(0.37))	(0.41)
Observations	4925	9867	8492	3943	5240	2812	6909	18124
Foreign father	0.30	0.15	0.14	0.08	0.01	0.16	0.28	0.25
	(0.46)	(0.35)	(0.34)	(0.27)	(0.10)	(0.37)	(0.45)	(0.43)
Observations	4584	9817	8401	4110	5083	2715	6803	17872
Girl	0.50	0.50	0.50	0.48	0.49	0.49	0.49	0.50
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Observations	5219	10067	8757	4282	5322	2937	8502	18448
Number of students in math class	26.31	20.58	27.98	20.33	32.69	23.94	27.47	23.59
	(6.32)	(4.44)	(6.10)	(4.10)	(5.60)	(5.71)	(5.05)	(5.45)
Observations	4553	4938	6022	4278	4499	2728	3929	15097
Number of students in science class	26.67	20.60	27.99	20.33	32.72	23.94	27.73	23.98
	(6.32)	(4.29)	(5.87)	(4.10)	(5.60)	(5.71)	(4.91)	(6.02)
Observations	4440	4870	5960	4278	4535	2728	3596	14706
Size of cohort (divided by ten)	5.81	4.45	5.59	11.53	8.33	3.46	5.41	10.07
	(3.00)	(1.98)	(3.24)	(4.62)	(3.86)	(1.58)	(2.97)	(5.50)
Observations	5053	9829	6982	4282	5322	2472	8352	15989
Inhabitants in school district	3.24	4.01	3.44	3.76	2.70	3.68	3.14	3.51
	(1.74)	(1.14)	(1.42)	(1.34)	(1.37)	(1.17)	(1.72)	(1.51)
Observations	4958	9697	6684	4282	5210	2630	8062	15789
Teacher's tenure (years)	16.72	16.37	14.12	21.21	19.31	16.16	11.79	13.03
	(9.53)	(9.87)	(10.45)	(9.44)	(9.52)	(12.42)	(9.84)	(10.19)
Observations	4568	9777	6420	4282	5322	2698	7599	16430
Teacher's education	4.99	4	5.08	2.51	4.79	4.00	4.74	5.53
	(0.71)	(0)	(0.27)	(1.12)	(0.69)	(0.31)	(0.63)	(0.50)
Observations	4711	9849	6450	4282	5288	2705	7764	16318
License to teach (1: yes, 0: no)	0.82	0.98	0	0.97	0.98	0	0.86	(0.86)
	(0.39)	(0.15)	(0)	(0.18)	(0.15)	(0)	(0.34)	(0.34)
Observations	5219	10067	8757	4282	5322	2937	8502	18448
Teacher shortage	0.27	0.05	0.29	0.17	0.59	0.12	0.61	0.20
	(0.44)	(0.22)	(0.45)	(0.38)	(0.49)	(0.32)	(0.49)	(0.40)
Observations	5083	9671	7236	4282	5305	2553	8270	16255
Share of poor families in school	1.98	1.46	2.09	1.81	1.26	1.57	2.09	2.82
	(1.11)	(0.83)	(1.14)	(0.89)	(0.51)	(0.99)	(1.21)	(1.25)
Observations	5153	9857	6878	4282	5182	2673	7774	16557
Share of rich families in school	2.05	3.49	2.45	2.11	2.96	2.98	2.39	1.92
	(1.19)	(0.87)	(1.26)	(1.51)	(1.17)	(1.19)	(1.28)	(1.11)
Observations	4803	9633	6624	4282	5006	2542	7818	15072

Note: Standard deviation in parentheses. Books is a 1-6 index indicating the number of books in the student's home, where 1 is least and 6 is most books. The inhabitants in the school district index run from 1-6, where 1 indicates the largest and 6 the smallest districts. The teacher's education is a 1-6 index indicating the highest education the teacher has finished (1 is low, 6 is high). The teacher shortage dummy indicates whether the school to some extent suffers from teacher shortage, and equals one if the principal has reported value 2, 3 or 4 on question 23r. For more details, see the TIMSS background questionnaires: http://timss.bc.edu/timss2003i/context.html.

Table A2: Estimation of adjusted test scores

	Australia		Belgium		GB		Italy		Japan		Netherlands		NZL		USA	
Variables	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)	(M)	(N)	(O)	(P)
	Math	Science	Math	Science	Math	Science	Math	Science	Math	Science	Msth	Science	Math	Science	Math	Science
Books	10.38***	13.04***	8.739***	8.898***	12.41***	12.11***	5.760***	6.961***	15.70***	13.38***	9.654***	8.012***	13.85***	12.27***	9.372***	8.726***
	(1.042)	(1.158)	(0.822)	(0.822)	(0.795)	(0.786)	(0.917)	(0.959)	(1.177)	(1.140)	(0.864)	(0.881)	(1.137)	(1.190)	(0.627)	(0.609)
Native	5.357	3.392	21.38***	17.84***	44.85***	36.31***	24.75***	21.53***	48.07***	37.76***	5.421	8.546**	5.799	8.624*	39.90***	39.29***
	(3.660)	(3.738)	(3.660)	(3.653)	(2.747)	(2.471)	(6.292)	(6.919)	(10.08)	(12.05)	(4.212)	(3.597)	(3.886)	(4.395)	(1.959)	(1.925)
Foreign father	4.143*	0.116	-16.04***	-10.67***	-17.50***	-19.06***	-28.23***	-26.89***	3.912	-19.16	-14.54***	-17.10***	8.238***	7.969**	-3.302*	-8.982***
	(2.473)	(2.540)	(2.627)	(2.420)	(2.880)	(2.923)	(4.172)	(4.496)	(12.16)	(12.99)	(2.883)	(3.237)	(3.087)	(3.222)	(1.824)	(1.800)
Girl	-12.01***	-7.655***	-5.421***	-6.081***	-8.179***	-6.342***	-12.50***	-6.285***	-5.404**	-5.712**	-4.127**	-6.651***	-2.457	0.947	-9.329***	-6.786***
	(2.382)	(2.639)	(1.636)	(1.378)	(1.897)	(1.833)	(2.154)	(2.361)	(2.266)	(2.363)	(1.617)	(1.618)	(2.390)	(2.349)	(1.285)	(1.322)
Observations	4,480	4,480	9,592	9,592	8,285	8,285	3,852	3,852	5,014	5,014	2,675	2,675	6,473	6,473	17,510	17,510
No. of schools	203	0.051	0.067	0.067	0.126	0.110	0.047	0.040	0.068	0.053	0.069	0.070	0.056	0.046	0.109	0.101
R-squared	0.041	203	149	149	246	246	171	171	150	150	130	130	218	218	248	248

Note: Robust standard errors (adjusted for school level clustering) in parentheses
*** p<0.01, ** p<0.05, * p<0.1.
Constant term (not reported) and school fixed effects included

Table A3: Estimation of propensity score. Treatment is pbuild. Probit.

Treatment	Australia Pbuild	Belgium Pbuild	GB Pbuild	Italy Pbuild	Japan Pbuild	Netherlands Pbuild	NZL Pbuild	USA Pbuild
Size of cohort	0.449	-0.919	0.267	0.340	-0.00515	0.357	0.0594	-0.0568
	(0.299)	(0.850)	(0.475)	(0.270)	(0.338)	(0.760)	(0.263)	(0.114)
(Size of cohort)^2	-0.0620	0.215	-0.0487	-0.0227	-0.0180	-0.0744	-0.0150	-0.00112
	(0.0429)	(0.195)	(0.0778)	(0.0272)	(0.0394)	(0.162)	(0.0435)	(0.00811)
(Size of cohort)^3	0.00248	-0.0145	0.00255	0.000526	0.00106	0.00510	0.000698	7.75e-05
	(0.00181)	(0.0134)	(0.00372)	(0.000814)	(0.00143)	(0.0100)	(0.00209)	(0.000157)
Inhabitants in school	0.0915	-0.110	-0.0779	0.00983	-0.00356	-0.0667	-0.0392	0.0668
District	(0.0650)	(0.104)	(0.0917)	(0.0897)	(0.0954)	(0.139)	(0.0628)	(0.0704)
Teacher's tenure	0.00266	-0.00427	-0.00397	0.0126	0.0276**	-0.0121	0.00780	0.00429
	(0.0110)	(0.0111)	(0.0118)	(0.0117)	(0.0139)	(0.0113)	(0.0114)	(0.0103)
Teacher' education	-0.0157	()	-0.703	0.129	-0.273	()	0.0609	0.0260
	(0.154)	()	(0.552)	(0.121)	(0.185)	()	(0.174)	(0.215)
License to teach	-0.203	()	()	-0.0873	()	()	-0.454	-0.454
	(0.346)	()	()	(0.511)	()	()	(0.430)	(0.414)
Teacher shortage	0.687***	0.320	0.731**	0.326	0.645***	0.563	0.345*	0.803***
	(0.258)	(0.475)	(0.298)	(0.312)	(0.241)	(0.447)	(0.201)	(0.275)
Share of poor families	-0.0156	-0.0838	-0.0711	0.193	0.562**	-0.0741	-3.66e-05	0.138
in school	(0.115)	(0.150)	(0.137)	(0.134)	(0.254)	(0.173)	(0.100)	(0.106)
Share of rich families	-0.145	-0.194	0.0850	-0.0405	-0.00206	-0.0237	0.153	0.0562
in school	(0.101)	(0.138)	(0.126)	(0.0970)	(0.107)	(0.136)	(0.0945)	(0.113)
Students in class	0.0237	0.00709	-0.00443	-0.00750	0.00904	0.00149	0.0369	-0.00209
(average of science and math class)	(0.0185)	(0.0277)	(0.0223)	(0.0253)	(0.0271)	(0.0291)	(0.0233)	(0.0189)
No. Blocks	5	4	8	5	5	5	4	5
Bal. prop.y satisfied (1%)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Common support[1]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of treated	88	67	81	127	66	60	68	79
No. of untreated	71	68	42	44	71	36	113	95

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Constant term (not reported) included.

[1]This restriction implies that the test is performed only on the observations whose propensity score belongs to the intersection of the supports of the propensity score of schools with and without school with poor building conditions

supports of the propensity score of schools with and without school with poor building conditions