IC technology and learning. An impact evaluation of Cl@ssi2.0

Daniele Checchi[†*], Enrico Rettore[‡] and Silvia Girardi[♠]

(revised version: August 2018)

[†] FBK-Irvapp Trento, University of Milan and IZA – email: daniele.checchi@gmail.com – FBK-Irvapp, vicolo dalla Piccola 12 – 38122 Trento (Italy)
[‡] FBK-Irvapp Trento, University of Trento and IZA – email: enrettore@irvapp.it – FBK-Irvapp, vicolo dalla Piccola 12 – 38122 Trento (Italy)
[♠] LISER – email: silvia.girardi@liser.lu, Maison des Sciences Humaines 11, Porte des Sciences L-4366 Esch-sur-Alzette / Belval.

[*] Corresponding author

Abstract

In this paper we present a counterfactual evaluation of the effect on student achievements of a pilot study providing ICT resources at some Italian junior high schools. In 2009, 156 classes at 6th grade were endowed with additional resources earmarked for purchasing ICT equipment only. By selecting an equivalent number of classes in the same schools, we are able to conduct an evaluation of the causal effect of ICT on student achievements, controlling for their initial level. Despite a significant financial investment (in the order of 1500 euros per student over a three-year period), the estimated impact on literacy and numeracy achievement is negligible: if we take the most encouraging results, the average improvement associated with the programme would be 3 test points, corresponding to 17% of a standard deviation. Once we control for further outcomes (in terms of secondary school track attended the following year), we do not find any impact. Overall, we conclude that the intervention failed to identify innovative strategies for raising students' achievement.

Keywords: human capital; ICT educational support; quasi-experimental design
JEL code: I20; I28;

Such research on the impact of ICT is evidence-based and seeks to establish a causal relationship between input and impact. This is the 'holy grail' of government in many countries, of course, but it is not easy to isolate cause from effect, especially in education where there are so many variables in play. *(Balanskat et al 2006, p.23)*

1. Introduction

The introduction of IC technologies within schools may have significant repercussions on teaching. For example, Balanskat et al. (2006) conduct a review of 25 studies in the European context up to that date. However, the limited number of experimental studies makes it difficult to reach consensus on the causal impact of this technology, not to speak of the channels through which any impact is achieved. Several aspects contribute to the lack of a consensus view. First, the standard requirements for proper randomised trials are rarely satisfied, since teachers and students hardly attain to the non-contamination principle: schools are interactive communities, and any novelty tends to be shared with mates or colleagues. Second, teaching contents are differently adaptable to computerised technologies, mainly because they require diverse learning strategies. Consequently, in the same class some subjects may exhibit significant improvements in achievement while others remain unaffected. Third, access to new technologies is heterogeneous across pupils, being correlated to the availability of analogous resources at home. All taken into account, it is not surprising to find heterogeneous results in the (limited) literature on the topic.

This indeterminacy is exacerbated in any context where there is limited familiarity with IC technology. Despite its high level of per-capita income, Italy is characterised by limited use of ICT, both at home and in schools. According to the TIMSS 2007 database (reported in Eurydice 2011), the percentage of 4th graders using computers at school was 63% (60% for 8th graders), while corresponding figures for UK students were 86% and 79%. When corresponding teachers were asked whether they had ever used a computer in science classes, 40% of Italian teachers in 4th grade (and between 58% and 63% in 8th grade) provided a negative answer (corresponding figures for UK teachers are lower by at least 20 percentage points). In comparison to other European countries, in 2012 Italy ranked lowest in terms of computers in school per student, together with Greece (Table 2 in Bulman and Fairlie 2016). Thus by the time of the experiment discussed in this paper (2009-11), Italy was a country with great opportunities for expanding the use of ICT in teaching. This may also explain why the Ministry of Education decided to invest a significant amount of funds in order to foster the adoption of ICT in Italian schools.

This paper analyses the first (and to the best of our knowledge only) experiment conducted in Italy on the adoption of ICT in schools, using a counterfactual evaluation. The experiment was started in 2009 and involved 156 lower secondary schools (comprising grades 6 to 8), allocating a significant amount of money (30k euros) earmarked for ICT acquisition to the benefit of a single class (20-25 pupils) within a school. We matched the 156 treated classes to an equivalent number of control classes within the same school and analysed the potential effect of a large (as compared to the standard funds available to schools) ICT-earmarked fund on student achievement, as measured by literacy and numeracy tests, taken on entry to the 6th grade and three years later on exit from the 8th grade. The grant was restricted to the acquisition of hardware for students and/or teachers, thus purposely excluding teacher support and training, software and internet connection, hardware maintenance, which were considered as pre-conditions for applying for funds. The main difference with respect to other experiments studied in the literature is that teachers were left free to spend the funds on whatever hardware they deemed appropriate. Although most students and teachers received tablets, a significant proportion of classes purchased visual or sound equipment: consequently, our treatment is properly defined as "*availability of funds, in the order of 1500 euros per student, to increase the ICT endowment of the treated classes*", and it is by definition heterogeneous. Overall, our main finding is no effect of the treatment at the mean of the distribution of test scores, but some, albeit small, positive effects in the lower tail. We also find no impact on the secondary school track chosen by students in treated classes. Given the negligible impact on achievements (test scores) and attainment (track choice), this initial experiment proved a failure, and it was discontinued, while the original plan was to extend it to primary and upper secondary levels of schooling. Among the possible explanations for the absence of effect, in the robustness checks we considered possible contamination between treated and control, as well as heterogeneity of treatment associated with heterogeneous duration of treatment or varying degree of permanence of teachers in the same class/school, but in all cases we were unable to reject the lack of impact.

Our contribution to the literature is confirmatory: even when the treatment is financially significant and flexible enough to allow for different sorts of ICT equipment, student achievement in school does not improve, with some exception for low achievers in literacy. This is particularly discouraging when considering the Italian context, where ICT technology in schools is still limited; however, if there are complementarities among school inputs (for example between ICT equipment and quality of connectivity), this may not be surprising. At the same time, this experiment reveals that the absence of a well-specified protocol allows experimenting with alternative solutions but makes it rather difficult to identify through which channels the treatment could have been effective. The other contribution of the paper is that it represents the first application of a counterfactual evaluation in Italian schools, in which various governments have often invested significant resources (the new digital plan introduced by the law n.107/2015 aims to wire-connect all Italian schools in the next three years) without any evaluation of effectiveness and/or cost-benefit convenience.

The paper is organised as follows: the next section contains a literature review; section 3 introduces Cl@ssi2.0, presents our evaluation design and describes the data; section 4 sets out the main results while section 5 discusses possible threats to their validity and the robustness checks that we conducted to deal with them. Section 6 discusses the results, and a final section concludes.

2. Literature review

Bulman and Fairlie (2016) review recent evidence on the impact of ICT on educational outcomes. When considering the impact of classroom use of computers on student achievement, they stress that most experiments allow for the evaluation of ICT as an additional resource, without replacing more traditional ones, like books or teachers' aids. In practice there must be some substitutability in time use, as in the present case, where treated schools obtained additional resources thanks to the project analysed. While this design seems to favour finding a positive effect of technology, most studies analysed in their review find no impact on students' achievements: "*The results suggest that ICT does not generate gains in academic outcomes or that schools allow computer-based instruction to crowd out traditional instruction. Regardless, a null result in this context is a stronger result than if there was a binding constraint that required substitution away from investment and time allocated to other inputs.*" (Bulman and Fairlie 2016, p.250). The real problem is that new technologies do not only offer access to new information (mostly web resources); they also allow new teaching methods (for example learning objects, animation software for scientific experiments, etc.) which implicate all teaching subjects. Balanskat et al. (2006) distinguish between assessments and impact evaluations, in two main areas: learning outcomes and learners on the one hand, and teaching methodologies and teachers on the other. Their overall account suggests positive effects on learning in primary school (more in home language and science, less compelling in mathematics), though part of these qualitative effects are drawn from teachers' and/or parents' perceptions. In this perspective, the vast majority of teachers record an increase in attention and motivation: "*ICT has a strong motivational effect and positive effect on behaviour, communication and process skills*" (Balanskat et al. 2006, p.4). In addition, they suggest that ICT allows for greater responsibility of individual learning, as well as facilitating team work: "*ICT use at schools can help to minimise the social divide by reducing the digital divide*" (ibid.). On the teacher side, there are indications of increased enthusiasm and motivation; at a deeper level, there is also perception of a changing role in connection with the students: "*…with ICT, the teacher tends to become more of an advisor, critical dialogue partner and leader for specific subject domains*" (ibid.).

If we focus on the impact of computer provision on students' achievement, we have to distinguish between (randomised supply of) home computers and the introduction of new/additional computers in schools. While the former type of experiments tend to find no impact (Fairlie and London 2012, Beuermann et alt. 2013, Fairlie and Robinson 2013 – contrasting results can be found in Vigdor and Ladd 2010), possibly due the distracting effect of games: stronger positive effects are found in the latter approach where computers are distributed to schools, in some cases also allowing children to take computers home after class hours. While originally Angrist and Lavy (2002) did not find statistically significant effects of computer-aided instruction (which requires the expansion of computer availability in classes) on student test scores, Barrow et al (2009) produced evidence that the introduction of instructional computer software was effective in raising algebraic skills. In the same vein, Machin et al. (2007) found that random changes in ICT funding across schools produced an improvement in primary school achievements in literacy but not in numeracy. Moving from increased availability of computers to increased availability of internet access does not help to shed light on the issue of the potential causal impact of ICT: Goolsbee and Guryan (2006) study the impact of public subsidies on school demand for internet connection, finding a significant elasticity, without any detectable spill-over effect of student tests. By contrast, Hyland et al. (2013) find that a government programme providing broadband access to primary schools in Ireland is associated with higher average mathematics scores on standardised tests.

More ambiguous results are obtained when considering developing countries: Barrera-Osorio et al. (2009) did not find effects on student test scores of randomised distribution of free computers to schools in Colombia, even if accompanied by teacher training. Following teachers and student surveys, they argue that the lack of impact is attributable to the failure to incorporate these computers into the educational process. Similarly Cristia et al. (2012) do not find significant effects on literacy and numeracy achievements of the One Laptop per Child (OLPC) among randomly selected primary schools in (rural) Peru. Ferrando et al. (2012) considered the randomised distribution of computers to schools in Uruguay (Plan Ceibal), finding a positive impact on the math test scores of 6^{th} graders.

Summing up, the results of the literature on the impact of introducing/expanding computer use in schools seem rather inconclusive, mostly because it involves several changes that go beyond the mere presence of new technologies: teacher training; teaching practices; use of space; peer interaction; homework assignments; access to the Web. The experiment that we are studying here partly shares this indeterminacy, because the treatment consisted of a budgetary allocation, earmarked for *any* ICT that the teachers deemed useful for their teaching needs. Consequently, we take a conservative view, and we will interpret our results as the outcome of "increasing the ICT endowment in the treated class".

3. Method

3.1. The Cl@ssi 2.0 program

Cl@ssi2.0 is a programme launched in 2008 by the Italian Ministry of Education to foster ICT adoption in teaching. It is a large-scale experiment, meant to modify learning environments and to spread new teaching experiences. It provided a large grant to buy ICT for 156 Italian classes at the beginning of 6th grade, to be used over the next three years. The experiment was also repeated at a lower scale in primary and upper secondary school in subsequent years, but it was soon abandoned because of high costs and lack of evidence that it was effective.

In 2009, a call was launched to all Italian lower secondary schools, offering the opportunity to apply for extra funding for digital technologies. Conditions for eligibility of the school were the presence of a broadband connection, the internal approval of a project, and internal leadership by a certified ICT-trained teacher on a permanent contract. The projects were to be referred to one specific class within the school. 2361 schools responded to the call and applied, and 156 of them were selected, respecting a regional stratification. A graph illustrating the time sequence of the events is reported in Figure 1.

Schools within which treated classes were chosen were not randomly selected by the 20 regional committees. We could not gain access to the minutes of the selection process, but there is evidence that the selected schools were not representative of the pool of applicant schools (see below for some evidence supporting this view). At the beginning of the following school year 2009/10 the 156 classes identified in the projects of selected schools were entitled to receive an amount of 30,000 euros (approximately 40,000 US dollars) to be spent on IC technology only (training and maintenance expenses were not admissible). The size of the grant was far from negligible for most of the schools included in the programme (the median value of the grant as a proportion of total yearly resources available to the school is about 10%).

According to the original protocol, researchers in the field of education from local universities were in charge of coaching the selected classes to develop new teaching practices. In practice, some classes in a few regions received the support of university researchers, but field surveys that we conducted over the three-year period indicate that most of the sampled classes did not receive any external supervision.

3.2 The design of the impact evaluation

The Italian Ministry of Education did not plan a counterfactual evaluation of this study, but private sponsors made it possible to run an impact evaluation of the programme. The crucial aspect was identification of an appropriate control group. Two alternatives were discussed: an additional class within the same schools, exploiting *within-school* variation, or a similar class in schools not involved in the Cl@ssi 2.0 program, for example not applying to be included in the programme, or not selected despite having applied, thus exploiting *between-school* variation. Both strategies had pros and cons. The strategy of having treated and control classes in the same school was prone to the risk of contamination: control classes might gain some benefits from the ICT bought by the treatment classes, something that by design could not happen in the second alternative. In addition, there is the risk that control classes were strategically selected by school principals to magnify the difference in performance between treated and control classes, even if the amount of funding was independent of the achievement gains, and in fact school principals did not receive any official evaluation after the conclusion of the experiment. On the other hand, the second strategy was likely to incur a severe risk of selection bias driven by student sorting across schools by social origins and/or by perceived quality, a problem unlikely to be properly dealt with conditioning on observables (like parental education or occupation) to be collected in the benchmark survey.

Balancing pros and cons, we decided to go with the first alternative: the control group was defined as made up of *one additional class in each school selected for the project*. The school principals were required to identify as control classes those as similar as possible to the classes receiving the funding: "*choose the second class you would have chosen, had the first not been chosen*" was our request to the principals.

We included in the design of the study the collection of the information needed to detect the occurrence of the two threats to the validity of the design – non-random selection of the control group and contamination. For the former we ran a benchmark survey during the autumn of 2009, collecting information on the students' socio-economic background and conducting a competence-based test to measure literacy and numeracy of all students included in the study in order to check the degree of comparability of treatment and comparison groups. For the latter, we surveyed teachers and school principals, in order to monitor the correct implementation of the protocol. Our initial choice to select the control classes in the same schools as the treated ones was not without consequences because there is no way now to check the robustness of our results using as a control group classes from schools not involved in the experiment. Thus, our evaluation simply relies on within-school variability, and data limitations prevent us from considering alternative approaches.

Regarding the first source of concern - the comparability of treated and control groups at the start of the study - we report relevant statistics in Table 1. The average score is marginally higher in the treated classes for both literacy and numeracy. The unconditional magnitude of the difference is small (corresponding to one additional correct answer), but we can explicitly account for these residual differences between the two groups in the estimation of the causal effect by including the initial test score as additional control (more extensive discussion of the issue follows in Section 4.1). Note also that both treated and control classes achieve a score higher than the national average score, confirming that schools included in Cl@ssi 2.0 are by no means representative of the population of Italian schools. Apart from information on students' achievement and on their socio-economic background, we were not provided with any information on schools, in particular on pre-existing ICT endowment, and on teachers (teaching experience and qualifications). The same classes were re-tested three years later at the end of grade 8 to collect the information on students' achievement needed to measure the causal effect of the programme. The two tests are *not* anchored, so that we *cannot* measure the value-added at the individual level, but we can study the impact three years later conditional on the initial score. In addition, by normalising each survey with reference to its relevant population, we can also adopt a Diff-in-Diffs approach to analyse the impact of the new technologies.

The information that we recovered on teacher perceptions and practices is rich but irregular and dispersed, so that it does not lend itself to a formal impact evaluation. Teachers in the treated classes were required to compile three questionnaires, one per school year, on the teaching practices that they introduced in their class following the introduction of new technologies and on their *perceptions* of the impact that these new practices had on students. Their response rates were 80.1%, 64.1% and 69.9%, respectively. In addition, all teachers included in the study – either in a treated or in a control class – were required to compile a questionnaire by the end of the third year on their level of qualification and familiarity with ICT (475 respondents out of a potential population of 1500 teachers – school coverage reached 77.6%). School principals were surveyed twice, on technology adoption and on teacher turnover in the classes included in the study. Their response rate was 48.1% in the first survey and 99% in the second one. Eventually an external observer visited 40 schools in the period Dec. 2011-Jan. 2012, reporting on the context in which the programme had been implemented, on the targets that the schools identified at the outset, and on the practices they were implementing.

From the scattered information resulting from the implementation analysis we obtain two important findings:

i) The treatment had variable intensity due to delays in implementation. By March 2010 (first year), only 50% of the schools had already started using newly acquired technologies. The percentage rose to approximately 70% by the end of the first school year (June 2010) and to 90% by November 2010 (beginning of the second year). See Figure 2. Robustness of the main results to this heterogeneity of the treatment intensity is analysed in Section 5.2.

ii) The vast majority of schools bought notebooks/tablet and interactive whiteboards, but we do not know their exact amount because the schools were not requested to produce any cash statement at the end of the project. On the basis of interviews we obtain the picture shown in Table 2: a significant fraction of classes purchased equipment related to documenting (camcorders and cameras) and image manipulation (scanners and projectors); sounds and music were also considered, though at a lower rate. Some schools located in socially deprived areas also bought specific safe cases for portable computer protection (and recharge). Consequently, a lack of detailed data prevents us from studying the possibility of differential impacts of different equipment.

3.3 Sample description

Despite an initial sample population consisting of 6744 6th graders, distributed in 308 classes located in 155 schools, our working sample was reduced by one third.[1] The main reason for this reduction in sample size was the restriction we imposed to have two test records for each student, at the beginning and at the end of the programme. Closure of some schools (or merger with other schools) was responsible for the loss of track of approximately 10% of the classes included in the original sample (see Table 3). A change in the privacy protection policy of the Ministry[2] prevented a perfect linkage of all the student records at the benchmark survey to the corresponding records resulting from the achievement test by the end of the 8th grade. Migration of students and failures of some of them added some additional loss of track. Overall, we were left with 134 schools out of the original 156 in which both the treated and the control classes were available in the final sample, i.e. with information both at the start and the end of the programme. Table 4 provides some evidence on the underlying attrition process. Results from regressing the binary variable '*still in the sample by grade 8*' on students' characteristics collected at grade 6 show that controlling for schools fixed-effect, gender, age, students cohabiting with both parents, students with higher scores at the benchmark survey are less likely to attrition. This last effect is almost by construction, since good students are less likely to be failed. Note, however, that neither the treatment status nor its interaction with the literacy and numeracy score at the benchmark affect the probability of attrition,

[1] The interested reader can find descriptive statistics for the entire student population in Table A.1 in the Appendix.
[2] Until 2010 students were identified by their social security number. The new privacy protection policy replaced it by a new student identifier, which univocally corresponds to the original code while fully preserves anonymity. In the transition to the new system, some schools were unable to implement the new procedure, which explains the loss of one third of our students. We investigated whether this delay in implementation was systematically associated with any observable characteristics of the school without being able to find any consistent pattern. Thus we presume that the attrition is not systematically connected to our object of investigation (see Table 4).

meaning that attrition did *not* take place selectively on the treatment status at any level of literacy and numeracy skill.

Eventually, our working sample for the estimation of the impact on students' achievement of Cl@asse 2.0 was made up of 4487 students belonging to either the treatment or the control class in 134 schools. Table 5 presents summary statistics by treatment status. The test score is the number of correct answers to 100 questions. The two groups are statistically similar except for the initial test score in literacy, which is slightly higher in the treated classes (even if the difference is statistically significant only at level 0.10). As explained in the next section, we shall control for these differences when measuring the impact of the treatment. We now illustrate the results.

4. Results

4.1. On the comparability of the treatment and control groups

To begin with, we checked the degree of comparability of classes and students assigned to participate in the programme to classes and students in the same schools included in the control group. There are two separate threats to the comparability of the two groups: the possible selection bias induced by the way in which school principals chose the treatment and the control classes, and the bias induced by the attrition of some sample units out of the study, already discussed in the previous section.

We assessed the degree of comparability as of November 2009, i.e. at the beginning of the programme. Table 6 presents the results of regressing the binary treatment status on observable characteristics of the student and of his/her family. Column 1 contains a linear probability regression including only information on the students collected from the school register.[3] Columns 2 and 3 add sets of observable characteristics of the student's family reported by the student him/herself at the benchmark survey. Overall, there seems to be no systematic difference between the two groups of classes with respect to the observable characteristics of the students, except for a marginally significant difference with respect to age. Nevertheless, the students in the treated classes exhibited a better performance in tests at the start of the experiment. Table 7 estimates the mean difference in achievements, conditional on observable characteristics of the students, finding a

[3] Results were almost identical if we considered a probit model. Available from the authors. Also including school fixed effects does not change the picture, since by design there was one treated and one control class in (nearly) all schools; hence the OLS is equivalent to a fixed-effect estimator. We have anyhow kept the estimates with school fixed effect for completeness of information.

positive difference of 1.5 correct answers in literacy and 1.3 in numeracy: thus, the treatment group performed marginally better than the controls even at the benchmark.[4]

To account for these differences, the design that we envisaged was akin to the Difference-in-Differences model. We could not implement a pure Diff-in-Diffs estimation as such because the scale of the score was not invariant over time since the INVALSI test was not anchored. However, in the sequel we also consider the direct estimation of a Diff-in-Diff model after standardizing the scores with respect to their relevant distributions. We started by formalizing the problem with reference to the case of the raw literacy score independently measured in each survey. We now describe the solution that we adopted.

Let the literacy score observed on the i-th individual at time t be:

$$Y_{it} = a_t + b_t LIT_{it} + u_{it}, \quad t = 0,1$$

with $t = 0$ the pre-treatment period and $t = 1$ the post-treatment one; LIT_{it} is the true latent literacy ability of the i-th subject at time t; u_{it} is a measurement error. Here we allow the origin and the unit of measurement of the score to vary over time because of the lack of anchoring of the INVALSI test. If they were the same – i.e. if $a_0 = a_1$ and $b_0 = b_1$ – the standard Diff-in-Diff's on the observable score Y would be equivalent (up to sampling errors) to the unfeasible Diff-in-Diff's on the latent ability since the average value of the measurement error u_{it} is zero:

$$E[Y_{i1} - Y_{i0}|D = 1] - E[Y_{i1} - Y_{i0}|D = 0] = E[LIT_{i1} - LIT_{i0}|D = 1] - E[LIT_{i1} - LIT_{i0}|D = 0]$$

where D is the treatment status denoting exposure to the intervention. That is, the Diff-in-Diff's would identify the average causal effect on the treated under the usual common trend restriction. As applied to our case, the common trend restriction states that even if at $t = 0$ the average score for classes selected by school principals may be different from the corresponding average score for classes selected out, their variation over time in the absence of the intervention would be the same.

This is no longer the case if $b_0 \neq b_1$ because the scale of the score in $t = 1$ is not comparable to the scale in $t = 0$. We therefore used the pre-treatment score as a control. If the variance of the measurement error were negligible, controlling for Y_{i0} would be nearly the same as controlling for LIT_{0i}, thus getting rid of the selection bias at $t = 0$. If the measurement error is not negligible the sign of the bias depends on the correlation between the treatment status D and the true value of the control variable at $t = 0$, i.e. the sign of the selection bias at $t = 0$. In this specific case the evidence is that school principals induced a (small) positive selection bias, implying that our estimate might be slightly *upward* biased.

[4] The remaining coefficients are standard in the literature: the worst performers were foreign born students, students speaking dialect or a foreign language at home, students of less educated parents and/or with fewer educational resources at home, students with a single parent and/or many siblings, students with a working mother.

To anticipate the results in the next section, the residual selection bias (if any) due to the way in which school principals selected the Cl@ssi 2.0 *reinforces* our main conclusion that the programme had a minor impact on students' achievement (if any).

4.2. The impact of the treatment on student achievement

We now present the main results of the paper. Table 8 shows the estimates for the average causal effect of the programme from alternative specifications. The inclusion of control variables (as well as controlling or not for school fixed effects) does not affect the main conclusion that can be drawn from the exercise: *on average there is no impact of ITC technology on either literacy or numeracy*, but a modest positive effect on numeracy for children of college graduate parents. The coefficient on the test score at the benchmark is around 0.5. In the literature the coefficient for the test score lagged one year is typically between 0.7 and 0.8 (see for instance Vigdor and Nechyba, 2007). Since in our case the test score is lagged twice, the reference value falls in the range between 0.49 and 0.64, thus including our estimated coefficient. As a check on these results, we replicated the analysis on class averages. The basic results of these additional regressions are in line with those above.[5]

In Table 9 we present the Diff-in-Diff's estimate. After standardising the test scores in the two surveys with respect to the whole population - namely, without restricting the analysis to the working sample - we computed the first differences of the standardised test score and regressed it onto the treatment dummy along with the standard controls we used in Table 8. Not surprisingly also in this case we do not find any causal impact of the newly added IC technologies on students' achievements.

Allowing for potential heterogeneity of the effects, Table 10 (column 1 to 6) presents results for a regression in which the treatment status was interacted with the ranking of the students in the distribution of the score at the benchmark survey. The overall evidence is in line with the previous findings. The novelty here is that a small average positive impact of the programme sometimes emerges in the bottom tail of the distribution (1^{st} decile), which is partly offset by negative impact in the middle of the distribution (6^{th} decile for literacy, 7^{th} decile for numeracy) while the other parts of the distribution are nearly unaffected. These offsetting effects are difficult to interpret, since they could be the mere reflection of the uneven density distribution for the treated (see especially Figure 4). The results in Table 10 suggest that this causal effect could be heterogeneous among students: it is confined to poorly performing children, but it seems also to favour children from college educated parents, at least in the case of numeracy.

[5] See table A.2 in the Appendix.

Our overall conclusion is that *the average causal effect of newly introduced ICT on students' achievements is nil*. Some small positive effect can be detected in some specifications for pupils on the left tail of the distribution, which is partly offset by negative impact in the middle of the distribution.

4.3. The impact of the treatment on student career

Using students' identifiers, we were also able to track their subsequent follow-up one year later, after the conclusion of junior high school. The Italian secondary school (from grade 9 to grade 13) is divided into an academic oriented track (*licei*), a technical education track (*istituti tecnici*) and a vocational one (*istituti professionali*). During the 8th grade, students pre-enrol for the next year after the conclusion of the first term; then they receive advice from their teachers and possibly revise their choice before starting the 9th grade, conditional on passing the final exam (there is a small fraction that fail and have to repeat the year). We were able to reconstruct this information for almost all students of the initial sample (6303 out of 6744), thus overcoming the attrition that affects the achievement based on test scores. In Table 11 we estimate an ordered probit model, where the outcomes are ranked from the lowest (not found in upper secondary school national archives, possibly due to failure at the exit exam, migration abroad or simple drop-out) to the highest (being enrolled in the academic track, which is typically oriented to tertiary education). In column 1 we use the minimum set of information available for all students: girls, younger students and natives are more likely to be found in more academic oriented tracks. Having attended a treated class exhibits a positive contribution to this outcome, but it is simply due to the lack of adequate controls: as soon as we control for the test score in the benchmark survey in column 2, the treatment variable loses significance. The track attended by students is heavily affected by parental background, as testified by the strong impact of having a college-educated parent. More generally, best students reveal a lower inclination to disappear from the secondary school register or to attend a vocational track (consider the distribution of the dependent variable at the top of the table); this is confirmed by column 3 and 4, where we restrict the sample to the longitudinal component for which we have both test scores (at entry and exit). In conclusion, also on looking at the outcome after completion of junior high school, the availability of new technologies does not exhibit any causal impact on student careers in terms of better prospects.

We now discuss the threats to the validity of these results and how we conducted our robustness checks to deal with them.

5. Robustness checks

5.1. Contamination

Despite an explicit Ministry directive prohibiting the sharing of newly-bought technical equipment with other classes in the school, we collected qualitative evidence that, at least in some schools, the new technologies have been shared with other classes, possibly including control classes. It should also be noted that to some extent this is unavoidable since only the Italian language teacher is dedicated to a specific class, while all other teachers teach in more than one class. In personal interviews, some teachers defended their violation of the ministerial rules, on behalf of students: "*If I find a learning object that is effective in teaching, why should I refrain from using it with all other classes that I teach?*" commented a maths teacher in an interview. According to school principals, some contamination took place in no more than one out of ten cases. Inspection of Table 12 shows the extent of possible contamination, according to two different sources of information: it is larger according to teachers' opinions, but note that teachers were asked about sharing new ICT with *any other class* in the school, while school principals were explicitly asked about the use of the new ICT *by the control classes*.

To check whether contamination was biasing our results, we exploited the variability across schools in the location of the treated and control classes. All treated and control classes were identified by a school code and a building code, since an administrative unit (a school) may comprise more than one location (building), even in different municipalities. We created a dummy variable indicating 8 schools (out of 134) in which treatment and control classes were located in different buildings, making contamination much more difficult if not impossible. Nevertheless, either adopting sample split or interacting the treatment with this identifier, we did not find evidence of the significance of the treatment variable.[6]

An alternative source of contamination could be parental pressure to shift their children to the classes where new technologies were available. However, class changes are rare, since by law new classes are randomly formed at the start of junior high school, typically stratifying by gender, nationality and level of achievements. Once a class has been formed only exceptional events (like conflict of interests between a teacher and a child) are considered as possible reasons for class shift, and they are entirely under the discretionary power of the principal, who is also responsible for respecting existing rules for class size.

[6] If any, there is evidence of a negative impact of the treatment for literacy when treated and control classes are located in different buildings. See the estimates reported in Table A.4 in the appendix.

Parents could have asked for admission of their children in the future treated classes had they known the selection outcome in advance. However, new class formation typically takes place during the summer break (July and August), while the selection outcome was announced in September, making it impossible to manipulate the class formation.

A related confounding factor could be represented by parents in not-treated classes (including control classes) lobbying for their children to be "compensated" with equivalent IC technologies bought with other school funds. While we do not have direct information on the IC endowment of the classes in the schools that participated in this study, we can safely exclude that schools had equivalent amounts of monetary resources to match the Cl@assi2.0 allowance.

5.2. Heterogeneity of treatment intensity

According to the information that we collected in the implementation analysis, the treatment had variable intensity due to delays in the process of deciding, ordering and obtaining the intended technologies. By March 2010 (first year) only 50% of the classes had already started using the new technologies. The percentage rose to approximately 70% by the end of the first school year (June 2010) and to 90% by November 2010 (beginning of the second year – see Figure 1). To check the robustness of the main results to this heterogeneity of the treatment intensity we included in our main regression an interaction term between the treatment status and the months of effective adoption of newly purchased technologies, without finding any statistical effect.

A related factor that may have reduced the effectiveness of the treatment is the continuity of the teaching team. Although the programme required a commitment to permanence in the school of at least three years by the participating teams of teachers, at the end of the programme we cross-checked with the school administrative offices and found that on average one third (35.2%) of teachers had changed school in the sample period, thus potentially harming student achievement (see Barbieri et al. 2017). If the teachers' turnover could have been correlated to the availability of IC technologies, this could have biased our results. In order to check the potential contribution of this factor to the attenuation of the impact, we interacted this fraction with the treatment status, without finding any statistical impact.

5.3. Cheating

Since the school year 2011-2012, the test scores in the public use files provided by INVALSI have been rescaled by the factor [1–*probability of cheating*] where the probability of cheating is determined at the class - not the individual - level according to an INVALSI protocol.[7] We therefore distinguish between the *original* values (as we did in Table 8) and those *corrected* for cheating. To illustrate the difference between the *original* scores and the scores *corrected* by INVALSI, Figures 3 and 4 present the distribution of the literacy and numeracy score, respectively, for students in the control and treatment classes: it is evident that the algorithm used by INVALSI to predict potential cheating affects the distribution of test scores by compressing it towards the bottom (Angrist et al. 2014).

To check our results, in Table 13 we present estimates for the average causal effect using the *corrected* score instead of the *original* one, adopting the same specifications introduced in Table 8. The main difference between the two sets of results depends on the use that we make of the information on cheating: in Table 10 we use the *observed* individual test scores and control for the estimated probability of cheating at the class level, while in Table 13 we use of the individual test scores as *corrected* by INVALSI to account for cheating. Whilst using the *original* test score and controlling for cheating, the average impact is zero for both literacy and numeracy (except for a marginally significant positive effect on numeracy for children of college graduate parents), when using the test score *corrected* for cheating, a statistically significant average effect emerges on literacy in the range 2.0 to 3.5 points (i.e. number of correct answers) for children of uneducated parents, since the coefficient on the interaction between the treatment status and the dummy for college graduate parents nearly offsets the main coefficient. Conversely, no effect can be found for numeracy. These results are nearly unaffected by the inclusion in the regression of additional control variables and of school fixed effects. Working with class averages there is almost no effect when original test scores are considered, while we find some positive effect for treatment in literacy for corrected test scores, even if this effect is attenuated in classes where the share of children from college educated parents is higher (the effect is even reversed when the share of college educated parents exceeds 41%, which corresponds to the 7th decile in the class distribution, the median being 28%). This indicates potential heterogeneity of the effect of the treatment that we address in the next step.[8]

Table 11 presents results for a regression in which the treatment status was interacted with the ranking of the students in the distribution of the score at the benchmark survey (beginning of grade 6). As for the difference between using *original* (columns 1 to 6) vs. *corrected* (columns 7 to

---

[7] INVALSI estimates the probability of cheating using a statistical model that identifies surprisingly large average scores, low within-class variability, and suspicious patterns of missing values, all taken as potential indicators of teachers cheating to the test. See Quintano et al. (2009). There is evidence of reduction of cheating when external monitors are introduced (Bertoni et al. 2013).

[8] One can compare Table 13 (individual corrected scores) to Table A.3 in the Appendix (class averages of corrected scores).

12) scores the overall evidence is in line with the previous findings. The news here is that the average positive impact of the programme on literacy score detected by the regressions in Table 13 is driven by its causal effect in the bottom tail of the distribution (1^{st} decile) while the other parts of the distribution are nearly unaffected, although a negative impact is maintained in the middle of the distribution for numeracy as in the case of original scores. Thus the results in Table 13 confirm that a positive causal effect for literacy is confined to children of uneducated parents.

Summing up, when using the official test score measure provided by the Ministerial agency INVALSI instead of the original one, we find a small statistically significant positive effect of ICT on student achievements, especially for children of uneducated parents, who are also likely to be at the bottom of the initial distribution of ability. These effects are still visible – albeit to a lesser extent - in the original test score measure. A possible explanation for the larger effect we found using the official INVALSI score is that the probability of cheating turns out to be smaller for treated classes for literacy while there is no statistically significant difference between treated and control classes for numeracy (results available from the authors). Therefore, the INVALSI correction for cheating produces a larger estimated effect on the literacy score.

6. Discussion

Finding limited effect of experiments in teaching is not new in the Italian schools context. Imperfect compliance, weak enforcement of protocols, lack of appropriate outcome variables are endemic in the country, given the limited tradition of evidence-based policy-making.

As for the substantive results of the present experiment, two findings emerge forcefully: the impact of the programme on student achievement as measured by the test scores in literacy and numeracy – if any – is confined to the bottom tail of the distribution; and the effect is restricted to the literacy score only. When looking at subsequent stages of student careers, we do not find any impact on choices of tracks in upper secondary schools.

A possible qualitative interpretation of these findings is in the field analysis reported by the external observer who visited 40 Cl@ssi 2.0 schools during the last year of the programme (Dec. 2011 – Jan. 2012). According to his perception, the observed degree of cooperation between students in the Cl@ssi 2.0 was much stronger than in the standard classes: this was likely to result in a net gain for the weakest component of the class (disabled, immigrants and other weaker students). The external observer noted that the trigger activating this cooperation might have been a stronger commitment on the students' side to perform 'a good job' along with the awareness that when the task is rather complex doing a good job requires a great deal of cooperation among all those involved. According to the interviews with the teachers in the treated classes, the ICT available to the Cl@ssi 2.0 was an extremely powerful tool with which to enhance the effectiveness of the traditional way of teaching, for it allowed integrating together words, images, sounds in a stable reservoir easily and readily available to *all* students. Again, the weakest part of the class was likely to benefit the most from this since the other students would have been able to find their way to learning even in the standard scenario, possibly exploiting the resources they had available at home. All Cl@ssi 2.0 teachers stressed that what they did with disabled students would not have been possible in a standard class. There was also a widespread agreement on judging Cl@ssi 2.0 as a powerful tool for the integration of foreign students. Finally, teachers' perception of the benefit stemming from the programme was particularly strong in schools located in socially deprived areas.

We conclude with a suggested explanation for the evidence of a significant, albeit small, impact on literacy and not on numeracy. Exploiting the potential of ICT-enhanced teaching requires more intense interactions between teachers and students as compared to standard teaching practices. In this respect, teachers of language and literature have an advantage over teachers of other subjects since in the Italian junior high school they spend 12 hours per week in the same class (not to say that in most instances they are also in charge of teaching History and Geography) as compared to teachers of Maths and Science, who typically spend only 4 hours per week. Thus teachers of Italian language were in a favourable position to fully exploit this potential, as they apparently did with weaker students.

7. Concluding remarks

This paper has presented the results of a counterfactual evaluation of the effect of ICT additional resources invested at school on student achievements in Italy. Despite the huge investment of resources (in the order of 1500 euros per student over a three-year period), the results are almost negligible. Even if we take the most encouraging results (see columns 3 or 4 of Table 13, estimated on corrected scores), the average improvement associated with the programme would be 3 test points, corresponding to 17% of a standard deviation. Any cost-benefit evaluation would question the effectiveness of a programme costing 10,000 euros for each test point of improvement. When we look at further outcomes in terms of secondary school track attended, the treatment appears to have no impact on students' choices, which remain driven by their level of competences and their parental backgrounds.

The robustness test that we performed suggests that the potential threats to the internal validity of our results – contamination of the control classes, bias due to the way in which they were selected, heterogeneity of treatment – are negligible, making us unable to reject the no impact results.

Overall, this leaves open the issue of why the teachers interviewed were strongly convinced of the positive effect of ICT adoption, while recorded impact on achievement is at best weak. One possible interpretation is that teachers possess a wider and more complex view on student engagement in learning activity. Another possible interpretation is that new technologies help to develop non-cognitive skills (like cooperative attitudes among peers), which are not necessarily relevant to the outcomes that we have measured in the present paper.

Acknowledgements

Paper presented at SREE 2013 Spring conference (Washington). We thank two anonymous referees for helping us to make our contribution clearer and sharper. All remaining errors are ours. *Fondazione Agnelli* and *Fondazione per la Scuola* have generously funded the evaluation presented in this paper; Invalsi, the national agency for school assessment made the data on students' tests available to us. The interpretation of the results is entirely attributable to the authors and not to the sponsoring institutions nor to the Ministry of Education. Many people have made this evaluation possible, and we would like to thank them all, in particular Andrea Gavosto, Stefano Molina, Marco Giovannini and Gianfranco De Simone (*Fondazione Agnelli*); Annamaria Poggi and Claudia Mandrile (*Fondazione per la Scuola*); Piero Cipollone, Paolo Sestito, Roberto Ricci and Patrizia Falzetti (Invalsi); Gianna Barbieri and Antonio Martino (*Ufficio Statistico del MIUR*); Vittorio Campione and Filomena Morgia (*Ufficio del Consigliere per le Tecnologie*), Antonio Ronca, Giusy Cannella and Samuele Borri (Ansas). Eventually also Franco DeAnna, Valeria Pandolfini and Valentina Rastelli have provided valuable support at some stage of the research.

References

Angrist, J., and V. Lavy. (2002). New Evidence on Classroom Computers and Pupil Learning. *Economic Journal* 112: 735–765.

Angrist, J., E.Battistin and D.Vuri. (2014). In a Small Moment: Cheating and Class Size in Italian Primary Schools. NBER wp. 20173

Balanskat, Anja, Roger Blamire and Stella Kefala. (2006). The ICT Impact Report. A review of studies of ICT impact on schools in Europe. *European Schoolnet*

Barbieri, Gianna, Claudio Rossetti and Paolo Sestito, 2017. Teacher Motivation and Student Learning, *Politica economica*, issue 1, pages 59-72

Barrera-Osorio, Felipe, and Leigh L. Linden. (2009). The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia. Policy Research Working Paper 4836, Impact Evaluation Series No. 29, The World Bank.

Barrow, Lisa, Lisa Markman, and Cecilia E. Rouse. (2009). Technology's Edge: The Educational Benefits of Computer-Aided Instruction. *American Economic Journal: Economic Policy*, 1(1): 52-74

Bertoni, Marco, Brunello, Giorgio and Rocco, Lorenzo. (2013). When the cat is near, the mice won't play: the effect of external examiners in Italian schools. *Journal of Public Economics*, 104, pp. 65-77

Bulman, George and Robert W. Fairlie. (2016). Technology and Education: Computers, Software, and the Internet. in Eric A. Hanushek, Stephen Machin and Ludger Woessmann (eds) *Handbook of the Economics of Education* (vol.5). Elsevier North Holland: 239-280

Cristia, Julian P., Pablo Ibarraran, Santiago Cueto, Ana Santiago, and Eugenio Severin. (2012). Technology and Child Development: Evidence from the One Laptop per Child Program. Inter-American Development Bank Working Paper No. IDB-WP-304.

Eurydice (2011). *Key data on Learning and Innovation through ICT at school in Europe 2011*. Brussels

Fairlie, Robert and Jonathan Robinson. (2013). Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren. UC Santa Cruz working paper series

Fairlie, Robert and Rebecca London. (2012). The Effects of Home Computers on Educational Outcomes: Evidence from a Field Experiment with Community College Students. *Economic Journal* 122(June): 727-753

Ferrando, Mery, Alina Machado, Ivone Perazzo and Adriana Vernengo. (2012). Impact Assessment of the First OLPC National Model: The Uruguayan Plan Ceibal. mimeo

Goolsbee, Austan, and Jonathan Guryan. (2006). The Impact of Internet Subsidies in Public Schools. *The Review of Economics and Statistics* 88(2): 336-347.

Hyland, Marie; Layte, Richard; Lyons, Sean; McCoy, Selina; Silles, Mary. (2013). Are classroom internet use and academic performance higher after government broadband subsidies to primary schools? 2013 mimeo

Machin, Stephen, Sandra McNally, and Olmo Silva. (2007). New Technology in Schools: Is There a Payoff? *Economic Journal*, 117(522): 1145-1167, July.

Quintano, C., R. Castellano and S. Longobardi. (2009). A Fuzzy Clustering Approach to Improve the Accuracy of Italian Student Data. An Experimental Procedure to Correct the Impact of the Outliers on Assessment Test Scores. *Statistica & Applicazioni* ,Vol.VII(2),149–17

Vigdor, Jacob L., and Helen F. Ladd. (2010). Scaling the Digital Divide: Home Computer Technology and Student Achievement. NBER Working Paper No. 16078

Vigdor, Jacob L. and Thomas Nechyba. (2007), Peer effects in North Carolina public schools, in Ludger Woessman and Paul Petersen (eds), *Schools and the Equality of Opportunity Problem*, MIT Press.

White, Howard. (2010). A Contribution to Current Debates in Impact. *Evaluation* 16: 153-164

Table 1

Test score in grade 6 – beginning of school year 2009-10[1]

	schools	classes	Students	literacy (mean)	literacy (sd)	numeracy (mean)	numeracy (sd)
Controls[2]	152	152	3275	64.33	15.18	55.73	15.91
Treated	155	156	3469	65.63	15.10	56.79	16.40
Overall	155	308	6744	65.00	15.15	56.28	16.17
national INVALSI controlled sample (grade 5)	1296	2074	41539	60.80	11.00	50.90	19.00

[1] We tested the competences of 6th graders using an *ad hoc* test based on the test at exit from primary school. The score corresponds to the number of correct answers out of 100 questions.
[2] There are less control than treated classes because in three schools there was no other class than the treated one, while in one school there were two treated classes to which we chose to match just one control class.

Table 2

Distribution of treated schools by type of technology acquired

	according to school principals	*according to class coordinators*
tablet/pc/notebook	100.0	87.2
whiteboard interactive	89.7	73.8
camcorder	37.3	41.8
photographic equipment	28.0	35.5
multifunction printer/scanner	28.0	item not included
class furniture	20.0	item not included
recorder and/or player	17.3	item not included
Projector	10.7	39.0
broadband internet connection	10.7	34.0
number of respondent schools	75	141

Table 3

Attrition of classes and students between the entry test at grade 6 (school year 2009-10) and the final test at grade 8 (school year 2011-12)

	lack of match at class level	total classes	class attrition	lack of match at student level	total student	student attrition
Treated	17	156	10.90%	1141	3469	32.9%
Controls	16	152	10.46%	1116	3275	34.1%
Overall project cl@ssi2.0	33	308	10.68%	2257	6744	33.5%

Table 4

Probability of persisting in the sample after 3 years (from school year 2009-10 to school year 2011-12)

	1	2	3	4	5	6
Dependent variable: *still in the sample by grade 8th*	probit	linear probability	linear probability	linear probability	linear prob School FE	linear prob School FE
Female	0.211 [0.039]***	0.064 [0.012]***	0.057 [0.011]***	0.057 [0.011]***	0.045 [0.009]***	0.046 [0.009]***
birth year	0.023 [0.011]**	0.007 [0.002]***	0.006 [0.003]**	0.006 [0.003]**	0.004 [0.002]*	0.004 [0.002]*
foreign born	-0.282 [0.091]***	-0.094 [0.032]***	-0.054 [0.032]*	-0.055 [0.032]*	-0.05 [0.028]*	-0.052 [0.028]*
at least one college graduate parent	0.084 [0.062]	0.025 [0.019]	0.004 [0.017]	0.004 [0.017]	0.017 [0.011]	0.018 [0.011]*
speaking Italian at home	0.17 [0.062]***	0.056 [0.020]***	0.034 [0.021]*	0.034 [0.021]*	0.014 [0.014]	0.014 [0.014]
cohabiting with two parents	0.276 [0.063]***	0.088 [0.020]***	0.059 [0.019]***	0.059 [0.019]***	0.057 [0.018]***	0.056 [0.018]***
working mother	0.123 [0.045]***	0.038 [0.014]***	0.024 [0.014]*	0.024 [0.014]*	0.017 [0.011]	0.018 [0.010]*
working father	0.238 [0.099]**	0.08 [0.034]**	0.052 [0.033]	0.052 [0.033]	0.044 [0.030]	0.044 [0.030]
literacy test score at entrance (grade 6)			0.003 [0.001]***	0.003 [0.001]***	0.004 [0.000]***	0.004 [0.001]***
numeracy test score at entrance (grade 6)			0.003 [0.001]***	0.003 [0.001]***	0.003 [0.000]***	0.003 [0.001]***
treated class				0.007 [0.012]	0.013 [0.010]	0.083 [0.052]
treated class × literacy test score at entrance						0.00 [0.001]
treated class × numeracy test score at entrance						-0.001 [0.001]
Observations	5334	5334	5171	5171	5171	5171
Pseudo R^2 or R^2	0.06	0.07	0.11	0.11	0.50	0.50

Robust standard errors clustered at school level – intercept and regional controls included

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5

Descriptive statistics of the working sample

	overall sample	Students in treated classes	Students in control classes	pair wise comparison at the means
Nr. of observations	4487	2328	2159	
fraction of girls	50%	49%	50%	-0.001
birth year: 1994	0.04%	0.04%	0.00%	0.000
birth year: 1995	0.07%	0.04%	0.09%	0.000
birth year: 1996	0.60%	0.47%	0.74%	-0.002
birth year: 1997	3.68%	3.78%	3.57%	0.002
birth year: 1998 (regular students)	87.23%	87.06%	87.40%	-0.003
birth year: 1999	8.25%	8.47%	8.01%	0.004
foreign born	6%	6%	6%	-0.005
speaking Italian at home	79%	79%	79%	0.001
at least one parent with college education	31%	30%	32%	-0.017
cohabiting with two parents	87%	88%	87%	0.005
mean literacy score 6^{th} grade	66.98	67.38	66.55	0.83 *
(standard deviation in brackets)	(14.37)	(14.39)	(14.35)	(0.42)
mean numeracy score 6^{th} grade	58.26	58.54	57.95	0.58
(standard deviation in brackets)	(15.79)	(16.06)	(15.49)	(0.47)
mean literacy score 8^{th} grade, *original values*	75.43	75.65	75.18	0.46
(standard deviation in brackets)	(13.10)	(12.48)	(13.74)	(0.39)
mean numeracy score 8^{th} grade, *original values*	56.65	56.80	56.49	0.31
(standard deviation in brackets)	(17.41)	(17.39)	(17.44)	(0.52)

* significant at 10%

Table 6

Probability of being assigned to a treated class as a function of observable characteristics of students

	1	2	3
Dependent variable: *assigned to a treated class*	linear probability	linear probability	linear probability
female	0.006	0.003	0.003
	[0.010]	[0.013]	[0.013]
birth year	-0.001	-0.001	-0.002
	[0.002]	[0.002]	[0.002]
foreign born	-0.014	-0.016	-0.011
	[0.019]	[0.030]	[0.031]
disabled	0.01	0.052	0.032
	[0.027]	[0.036]	[0.042]
at least one college graduate parent		-0.02	-0.019
		[0.017]	[0.017]
speaking Italian at home		-0.007	-0.003
		[0.022]	[0.022]
book shelves at home		0.009	0.01
		[0.007]	[0.007]
number of siblings			0.003
			[0.008]
cohabiting with two parents			0.008
			[0.021]
working mother			-0.009
			[0.018]
working father			-0.026
			[0.035]
Observations	6731	5533	5263
R^2	0.00	0.00	0.00

Robust standard errors clustered at school level – intercept and regional controls included
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7

Regression of the literacy and numeracy score at the benchmark survey (6th grade – beginning of school year 2009-10) on the treatment status - OLS

	1	2	3	4	5	6
	Literacy score	Literacy score	Literacy score	Numeracy score	Numeracy score	Numeracy score
treated class	**1.328**	**1.459**	**1.47**	**1.115**	**1.31**	**1.344**
	[0.484]* **	**[0.478]*** **	**[0.489]*** **	**[0.619]* **	**[0.639]** **	**[0.635]** **
Female	1.363	0.869	0.795	-1.454	-2.027	-1.995
	[0.371]***	[0.370]**	[0.383]**	[0.445]***	[0.437]***	[0.439]***
birth year	0.151	0.055	0.049	0.202	0.1	0.105
	[0.098]	[0.060]	[0.060]	[0.095]**	[0.059]*	[0.058]*
foreign born	-14.884	-8.831	-9.089	-9.535	-4.978	-5.166
	[0.968]***	[0.974]***	[0.943]***	[0.901]***	[0.958]***	[0.973]***
at least one college graduate parent		1.787	1.074		1.322	0.596
		[0.474]***	[0.470]**		[0.556]**	[0.556]
speaking Italian at home		2.251	1.917		2.018	1.841
		[0.624]***	[0.591]***		[0.532]***	[0.512]***
book shelves at home		3.204	2.942		3.077	2.834
		[0.181]***	[0.185]***		[0.193]***	[0.201]***
number of siblings		-1.417	-1.235		-1.079	-0.974
		[0.225]***	[0.216]***		[0.284]***	[0.272]***
cohabiting with two parents		2.896	2.672		3.301	2.986
		[0.589]***	[0.594]***		[0.594]***	[0.590]***
working mother		0.98	0.796		1.161	0.843
		[0.447]**	[0.434]*		[0.478]**	[0.472]*
working father		1.616	1.497		0.568	0.359
		[1.069]	[1.092]		[0.992]	[0.981]
School fixed-effect	no	no	yes	no	no	yes
Observations	5970	5106	5106	5999	5134	5134
R^2	0.10	0.19	0.25	0.10	0.17	0.24

Robust standard errors clustered at school level – intercept and regional controls included

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8

The effect of Cl@sse 2.0 on test scores by the end of 8th grade (original test scores) – individual data - OLS

	Literacy				Numeracy			
	1	2	3	4	5	6	7	8
	Basic, full sample	Basic, excluding outliers, school FE	Basic, excluding outliers, school FE	Extended excluding outliers, school FE	Basic, full sample	Basic, excluding outliers, school FE	Basic, excluding outliers, school FE	Extended excluding outliers, school FE
test score at entrance (6th grade)	0.515 [0.017]***	0.526 [0.015]***	0.526 [0.015]***	0.500 [0.015]***	0.526 [0.021]***	0.532 [0.018]***	0.532 [0.018]***	0.516 [0.018]***
treated class	**-0.306 [0.738]**	**0.369 [0.528]**	**0.525 [0.578]**	**0.582 [0.560]**	**-0.587 [1.336]**	**0.389 [0.893]**	**-0.129 [0.972]**	**-0.175 [0.994]**
probability of cheating (estimated at class level)	16.581 [1.953]***	11.452 [1.155]***	11.524 [1.165]***	11.745 [1.163]***	38.52 [3.812]***	34.244 [5.177]***	34.193 [5.137]***	33.852 [5.374]***
Female	0.776 [0.348]**	0.639 [0.332]*	0.638 [0.332]*	0.506 [0.345]	-0.566 [0.453]	-0.488 [0.416]	-0.486 [0.415]	-0.736 [0.432]*
birth year	0.144 [0.095]	0.128 [0.100]	0.127 [0.100]	0.116 [0.100]	-0.036 [0.119]	-0.076 [0.096]	-0.075 [0.095]	-0.091 [0.097]
foreign born	-1.415 [0.790]*	-1.5 [0.783]*	-1.507 [0.784]*	-0.784 [0.877]	-4.318 [1.035]***	-3.599 [0.961]***	-3.568 [0.958]***	-1.931 [1.069]*
at least one college graduate parent	1.539 [0.343]***	1.277 [0.366]***	1.522 [0.496]***	0.77 [0.531]	2.305 [0.528]***	2.224 [0.495]***	1.396 [0.670]**	0.746 [0.696]
at least one college graduate parent × treatment			**-0.484 [0.659]**	**-0.524 [0.677]**			**1.629 [0.915]***	**1.624 [0.952]***
fraction of permanent teachers (still in the school after 3 years)	1.854 [1.782]	2.312 [2.118]	2.265 [2.120]	1.976 [2.051]	4.208 [3.184]	1.502 [3.766]	1.63 [3.778]	1.458 [3.727]
teacher in charge teaches language	1.674 [0.987]*	-0.091 [0.856]	-0.094 [0.854]	-0.062 [0.806]	-0.079 [1.615]	-2.812 [1.734]	-2.804 [1.734]	-2.525 [1.747]
teacher in charge teaches mathematics	-0.067 [0.859]	-1.403 [0.854]	-1.408 [0.853]*	-1.318 [0.812]	0.911 [1.834]	0.859 [1.412]	0.874 [1.414]	1.106 [1.432]
speaking Italian at home				0.677 [0.422]				1.607 [0.559]***
book shelves at home				0.827 [0.151]***				0.715 [0.206]***
number of siblings				-0.549 [0.198]***				-0.366 [0.239]
cohabiting with two parents				0.969 [0.502]*				1.655 [0.580]***
working mother				0.755 [0.385]*				1.164 [0.512]**
working father				-0.107 [0.806]				-0.753 [1.091]
Observations	4171	4171	4171	3917	4174	4174	4174	3920
R^2	0.42	0.50	0.50	0.50	0.39	0.51	0.51	0.51

robust standard errors clustered at school level – constant and regional controls included

Outlier students are defined as those with a score smaller than 10

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9

Diff-in-Diff's estimate of the effect of Cl@ssi 2.0; standardised test scores (original test scores) – individual data.

	Literacy				Numeracy			
	1	2	1	2	1	2	1	2
	Basic, full sample	Basic, excluding outliers, school FE	Basic, full sample	Basic, excluding outliers, school FE	Basic, full sample	Basic, excluding outliers, school FE	Basic, full sample	Basic, excluding outliers, school FE
treated class	-0.043	0.034	0.054	0.055	-0.04	0.036	0.018	0.002
	[0.059]	[0.041]	[0.046]	[0.045]	[0.087]	[0.060]	[0.064]	[0.066]
probability of cheating (estimated at	1.077	0.844	0.853	0.874	2.316	1.968	1.967	1.985
class level)	[0.155]***	[0.088]***	[0.089]***	[0.084]***	[0.251]***	[0.288]***	[0.286]***	[0.294]***
Female	0.032	0.03	0.03	0.029	0.027	0.039	0.039	0.037
	[0.029]	[0.028]	[0.028]	[0.029]	[0.030]	[0.028]	[0.028]	[0.029]
birth year	0.007	0.006	0.006	0.007	-0.009	-0.011	-0.01	-0.009
	[0.004]*	[0.004]	[0.004]	[0.004]	[0.008]	[0.006]*	[0.006]*	[0.006]
foreign born	0.219	0.2	0.199	0.177	-0.004	0.033	0.034	0.045
	[0.066]***	[0.063]***	[0.063]***	[0.071]**	[0.069]	[0.065]	[0.065]	[0.071]
at least one college graduate parent	-0.012	0.007	0.038	0.036	-0.013	0.03	0.001	0.024
	[0.029]	[0.029]	[0.040]	[0.043]	[0.032]	[0.030]	[0.044]	[0.045]
at least one college graduate			-0.06	-0.069			0.057	0.06
parent × treatment			[0.053]	[0.055]			[0.057]	[0.059]
fraction of permanent teachers (still in	0.045	0.042	0.036	0.024	0.059	-0.21	-0.205	-0.218
the school after 3 years)	[0.134]	[0.163]	[0.163]	[0.160]	[0.187]	[0.235]	[0.236]	[0.237]
teacher in charge teaches language	0.134	-0.045	-0.045	-0.032	-0.011	-0.19	-0.19	-0.148
	[0.082]	[0.064]	[0.064]	[0.062]	[0.099]	[0.099]*	[0.098]*	[0.100]
teacher in charge teaches	-0.041	-0.172	-0.172	-0.16	-0.002	-0.032	-0.032	-0.004
mathematics	[0.073]	[0.069]**	[0.069]**	[0.067]**	[0.123]	[0.098]	[0.098]	[0.100]
speaking Italian at home				0.012				0.057
				[0.037]				[0.039]
book shelves at home				-0.013				-0.044
				[0.013]				[0.013]***
number of siblings				-0.013				0.009
				[0.016]				[0.017]
cohabiting with two parents				0.018				0.018
				[0.042]				[0.039]
working mother				0.038				0.051
				[0.032]				[0.035]
working father				0.016				-0.017
				[0.068]				[0.072]
Observations	4171	4171	4171	3917	4174	4174	4174	3920
R^2	0.08	0.20	0.20	0.20	0.25	0.38	0.38	0.38

robust standard errors clustered at school level – constant and regional controls included
Outlier students are defined as those with an original score smaller than 10
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 10

The effect of Cl@ssi2.0 on test scores at 8th grade (original and corrected test scores) conditional on initial ranking of the student – individual data - OLS

	1	2	3	4	5	6	7	8	9	10	11	12
	original test scores						corrected test scores					
	Literacy	Literacy, school FE	Literacy, school FE	Numeracy	Numeracy, school FE	Numeracy, school FE	Literacy	Literacy, school FE	Literacy, school FE	Numeracy	Numeracy, school FE	Numeracy, school FE
test score at entrance (6th grade)	0.528	0.543	0.516	0.547	0.552	0.542	0.461	0.529	0.5	0.531	0.533	0.524
	[0.022]***	[0.018]***	[0.017]***	[0.028]***	[0.024]***	[0.023]***	[0.035]***	[0.022]***	[0.022]***	[0.031]***	[0.026]***	[0.025]***
treatment×1st decile 6th grade	**2.231**	**2.791**	**2.708**	**1.894**	**1.826**	**2.885**	**5.48**	**5.194**	**4.833**	**1.373**	**1.642**	**2.681**
	[1.629]	**[1.397]****	**[1.362]****	**[1.905]**	**[1.430]**	**[1.459]****	**[2.074]*****	**[1.806]*****	**[1.774]*****	**[2.242]**	**[1.577]**	**[1.588]***
treatment×2nd decile 6th grade	-0.002	0.329	0.183	2.888	2.031	2.738	2.342	1.586	0.503	2.304	1.712	2.406
	[1.422]	[1.147]	[1.078]	[1.855]	[1.445]	[1.380]**	[1.790]	[1.400]	[1.213]	[2.145]	[1.547]	[1.468]
treatment×3rd decile 6th grade	0.032	-0.047	0.073	-0.159	-0.454	-0.07	0.821	1.498	1.323	-0.862	-0.767	-0.326
	[1.024]	[0.842]	[0.811]	[1.490]	[1.325]	[1.296]	[1.608]	[1.208]	[1.211]	[1.739]	[1.355]	[1.328]
treatment×4th decile 6th grade	0.244	0.238	0.103	-1.692	-1.983	-1.337	1.829	1.46	0.777	-3.744	-2.863	-2.302
	[0.972]	[0.763]	[0.729]	[1.273]	[1.061]*	[1.061]	[1.441]	[1.152]	[1.163]	[1.704]**	[1.175]**	[1.160]**
treatment×5th decile 6th grade	0.18	0.099	-0.011	-1.398	-1.212	-0.651	1.909	1.651	1.473	-1.873	-1.296	-0.731
	[0.936]	[0.788]	[0.797]	[1.244]	[0.943]	[0.928]	[1.474]	[1.273]	[1.345]	[1.475]	[1.109]	[1.087]
treatment×6th decile 6th grade	-1.075	-1.604	-1.666	-1.222	-1.256	-0.449	-0.128	-0.42	-0.892	-0.771	-1.285	-0.472
	[0.838]	[0.730]**	[0.735]**	[1.211]	[0.962]	[0.905]	[1.571]	[1.178]	[1.158]	[1.319]	[1.013]	[0.945]
treatment×7th decile 6th grade	0.574	0.519	0.498	-1.847	-2.603	-2.094	1.493	2.176	1.969	-1.645	-2.731	-2.155
	[0.643]	[0.566]	[0.555]	[1.219]	[0.934]***	[0.903]**	[1.507]	[1.185]*	[1.249]	[1.442]	[1.003]***	[0.989]**
treatment×8th decile 6th grade	0.697	0.597	0.632	-0.969	-1.875	-1.37	1.754	1.017	0.768	-0.6	-1.789	-1.221
	[0.704]	[0.691]	[0.682]	[1.269]	[1.042]*	[1.040]	[1.531]	[0.989]	[1.043]	[1.363]	[1.068]*	[1.064]
treatment×9th decile 6th grade	-1.054	-1.242	-1.079	-0.26	-1.28	-0.629	0.118	0.219	-0.098	0.015	-1.151	-0.601
	[0.593]*	[0.561]**	[0.564]*	[1.239]	[1.094]	[1.064]	[1.614]	[1.032]	[1.078]	[1.349]	[1.121]	[1.097]
probability of cheating (estim. at class level)	16.734	11.831	11.992	38.747	33.414	33.209						
	[1.986]***	[1.149]***	[1.139]***	[3.697]***	[5.181]***	[5.482]***						
Female	0.766	0.641	0.504	-0.58	-0.476	-0.727	0.744	0.943	0.853	-0.449	-0.431	-0.665
	[0.350]**	[0.331]*	[0.343]	[0.454]	[0.413]	[0.432]*	[0.509]	[0.407]**	[0.434]*	[0.482]	[0.413]	[0.433]
birth year	0.148	0.139	0.129	-0.049	-0.076	-0.092	0.131	0.148	0.147	-0.026	-0.069	-0.085
	[0.099]	[0.103]	[0.105]	[0.117]	[0.090]	[0.093]	[0.143]	[0.108]	[0.119]	[0.133]	[0.091]	[0.092]
foreign born	-1.552	-1.622	-0.893	-4.492	-3.694	-2.095	-1.631	-1.564	-0.669	-3.977	-3.523	-1.977
	[0.794]*	[0.782]**	[0.869]	[1.029]***	[0.953]***	[1.069]*	[0.894]*	[0.781]**	[0.868]	[1.013]***	[0.972]***	[1.118]*
at least one parent with college	1.498	1.41	0.484	2.296	1.244	1.55	1.206	1.838	0.385	2.341	1.27	1.553
	[0.340]***	[0.483]***	[0.392]	[0.533]***	[0.651]*	[0.531]***	[0.683]*	[0.676]***	[0.415]	[0.639]***	[0.652]*	[0.550]***
at least one parent with college×treatm		-0.313			1.879			-1.507			1.715	
		[0.602]			[0.831]**			[1.017]			[0.850]**	
Italian spoken at home			0.675			1.584			1.022			1.377
			[0.420]			[0.565]***			[0.497]**			[0.584]**
number of books at home			0.8			0.7			0.761			0.66
			[0.150]***			[0.208]***			[0.177]***			[0.206]***
number of siblings			-0.529			-0.403			-0.508			-0.392
			[0.197]***			[0.240]*			[0.223]**			[0.236]*
living with two parents			0.959			1.62			0.639			1.627
			[0.505]*			[0.580]***			[0.557]			[0.585]***
mother working			0.804			1.176			0.597			1.336
			[0.380]**			[0.510]**			[0.427]			[0.514]***
father working			-0.161			-0.951			-1.586			-1.009
			[0.802]			[1.101]			[1.000]			[1.091]
Observations	4171	4171	3917	4174	4174	3920	4171	4171	3917	4174	4174	3920
R^2	0.42	0.5	0.5	0.39	0.51	0.51	0.26	0.51	0.52	0.28	0.47	0.48

Robust standard errors clustered at school level – intercept and regional controls included
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 11

The effect of Cl@ssi2.0 on student status one year after completion of the 8th grade – individual data - ordered probit

	1	2	3	4
	original sample without initial test	original sample with initial test	reduced sample final test original score	reduced sample final test corrected score
Dependent variable:				
- not in secondary school (migrated, retained in grade 8, drop-out)	10.95%	9.17%	0.20%	
- enrolled in vocational track	14.50%	13.08%	13.35%	
- enrolled in technical track	39.43%	40.26%	44.36%	
- enrolled in academic track	35.13%	37.49%	42.08%	
Literacy test score at entrance (6th grade)		0.028 [0.002]***	0.015 [0.002]***	0.019 [0.002]***
Numeracy test score at entrance (6th grade)		0.019 [0.002]***	0.011 [0.002]***	0.011 [0.002]***
Literacy test score at exit (8th grade) – original			0.019 [0.003]***	
Numeracy test score at exit (8th grade) – original			0.007 [0.002]***	
Literacy test score at exit (8th grade) - corrected for cheating				0.01 [0.002]***
Numeracy test score at exit (8th grade) - corrected for cheating				0.011 [0.002]***
attended a treated class	**0.089 [0.046]***	**0.054 [0.045]**	**0.059 [0.055]**	**0.035 [0.054]**
Female	0.293 [0.031]***	0.347 [0.036]***	0.254 [0.043]***	0.268 [0.043]***
birth year	0.022 [0.009]**	0.01 [0.005]**	0.005 [0.005]	0.006 [0.005]
foreign born	-0.621 [0.059]***	-0.175 [0.072]**	-0.105 [0.087]	-0.105 [0.088]
at least one parent with college		0.47 [0.039]***	0.467 [0.048]***	0.475 [0.048]***
Observations	6303	5199	3992	3992
Pseudo R^2	0.02	0.15	0.15	0.14
Loglikelihood	-7754.34	-5414.89	-3416.95	-3439.02

Robust standard errors in brackets clustered at class level - region controls included
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 12

Percentage of schools in which the Cl@ssi 2.0 ICT have been shared with other classes in the school

*The newly acquired ICT technologies under the project Cl@ssi2.0 have been shared by **other** classes in the school*	percentage of positive answers
respondent: school principals	33.3
respondent: teachers (1st questionnaire)	20.8
respondent: teachers (2nd questionnaire)	63.0
Combined information	58.9
*The newly acquired ICT technologies under the project Cl@ssi2.0 has been shared by the **control** class in the school*	
school principals	11.3

Table 13

The effect of Cl@ssi2.0 on test scores by the end of 8th grade (test scores corrected for cheating) – individual data - OLS

	Literacy				Numeracy			
	1	2	3	4	5	6	7	8
	Basic, full sample	Basic, excluding outliers, school FE	Basic, excluding outliers, school FE	Extended, excluding outliers, school FE	Basic, full sample	Basic, excluding outliers, school FE	Basic, excluding outliers, school FE	Extended, excluding outliers, school FE
test score at entrance (6th grade)	0.429	0.496	0.497	0.475	0.524	0.522	0.521	0.504
	[0.029]***	[0.019]***	[0.019]***	[0.019]***	[0.024]***	[0.018]***	[0.018]***	[0.019]***
treated class	**2.01**	**2.53**	**3.276**	**3.475**	**-0.498**	**0.039**	**-0.427**	**-0.189**
	[1.746]	**[1.414]***	**[1.510]****	**[1.541]****	**[1.774]**	**[1.006]**	**[1.096]**	**[1.227]**
Female	0.706	0.968	0.962	0.835	-0.454	-0.489	-0.486	-0.677
	[0.508]	[0.407]**	[0.407]**	[0.434]*	[0.478]	[0.423]	[0.423]	[0.436]
birth year	0.122	0.132	0.13	0.123	-0.018	-0.078	-0.077	-0.087
	[0.132]	[0.108]	[0.105]	[0.111]	[0.134]	[0.094]	[0.093]	[0.096]
foreign born	-1.426	-1.508	-1.546	-0.665	-3.882	-3.435	-3.408	-1.848
	[0.901]	[0.783]*	[0.786]*	[0.888]	[1.017]***	[0.983]***	[0.980]***	[1.117]*
at least one college graduate parent	1.31	1.063	2.257	1.898	2.366	2.165	1.406	0.9
	[0.696]*	[0.380]***	[0.676]***	[0.740]**	[0.627]***	[0.492]***	[0.679]**	[0.717]
at least one college graduate parent			**-2.344**	**-2.841**			**1.504**	**1.364**
× treatment			**[1.077]****	**[1.114]****			**[0.941]**	**[0.973]**
fraction of permanent teachers (still in	-3.313	-7.903	-8.089	-9.997	1.645	3.222	3.304	-1.981
the school after 3 years)	[4.234]	[5.757]	[5.697]	[5.727]*	[4.046]	[3.804]	[3.828]	[4.675]
teacher in charge teaches language	1.459	-1.396	-1.406	-1.183	-0.813	-3.273	-3.276	-2.858
	[1.808]	[2.296]	[2.280]	[2.268]	[2.114]	[1.787]*	[1.787]*	[1.875]
teacher in charge teaches	-1.377	-0.818	-0.851	-1.748	1.433	1.44	1.446	1.677
mathematics	[2.316]	[2.391]	[2.394]	[2.567]	[2.026]	[1.539]	[1.545]	[1.610]
speaking Italian at home				1.044				1.418
				[0.494]**				[0.572]**
book shelves at home				0.773				0.671
				[0.180]***				[0.205]***
number of siblings				-0.519				-0.352
				[0.221]**				[0.235]
cohabiting with two parents				0.626				1.65
				[0.558]				[0.585]***
working mother				0.576				1.338
				[0.421]				[0.513]***
working father				-1.502				-0.812
				[0.980]				[1.080]
Observations	4171	4146	4146	3917	4174	4124	4124	3920
R^2	0.26	0.50	0.50	0.52	0.28	0.45	0.45	0.48

Robust standard errors clustered at school level – constant and regional controls included
Outlier students are defined as those with a score smaller than 10
* significant at 10%; ** significant at 5%; *** significant at 1%

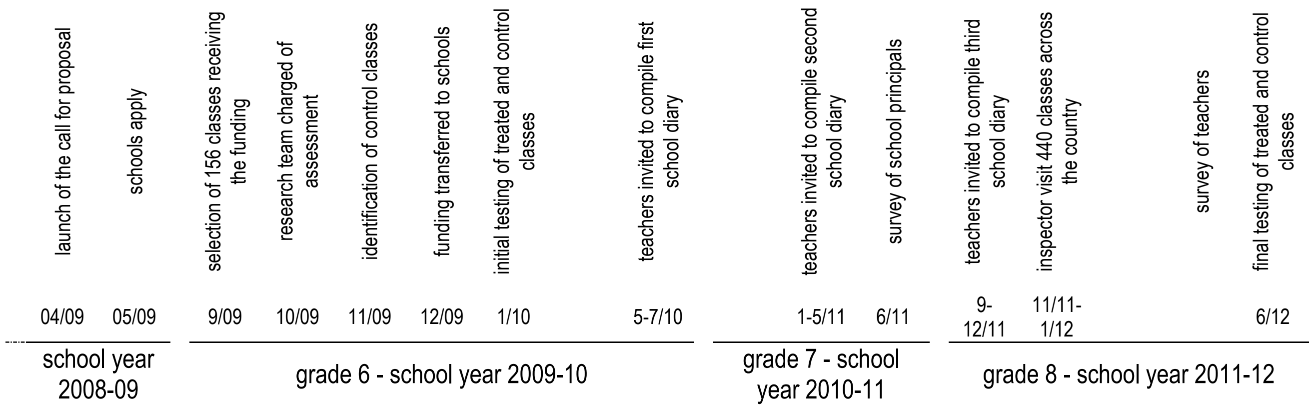

Figure 1: Time sequence of the experiment

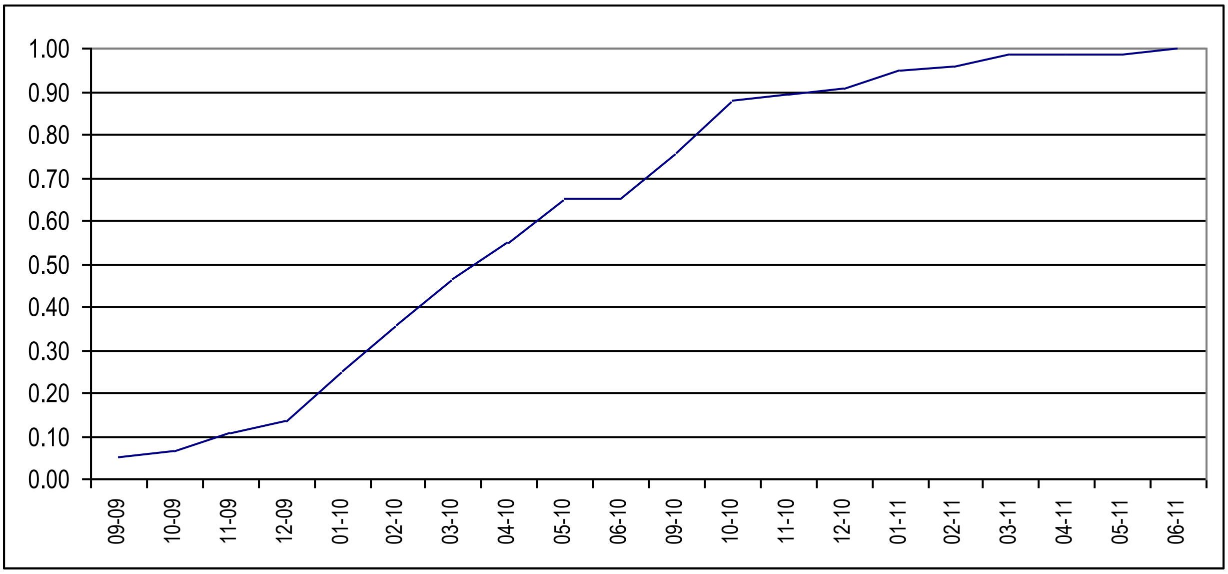

Figure 2: Cumulative fraction of schools introducing newly acquired ICT by month (75 respondent schools)

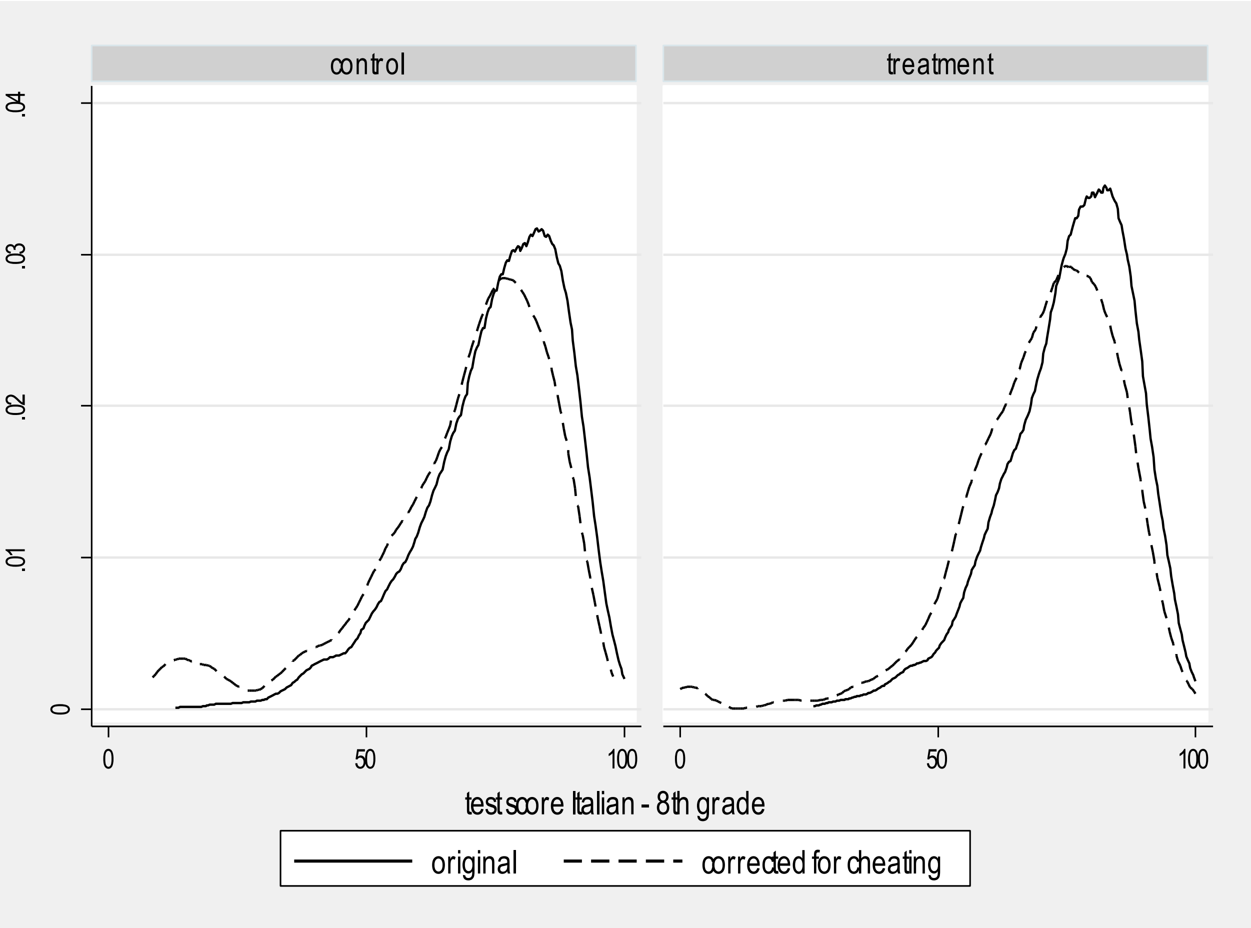

Figure 3: Distribution of the literacy score by treatment status - end of grade 8 (sch.year 2011-12)

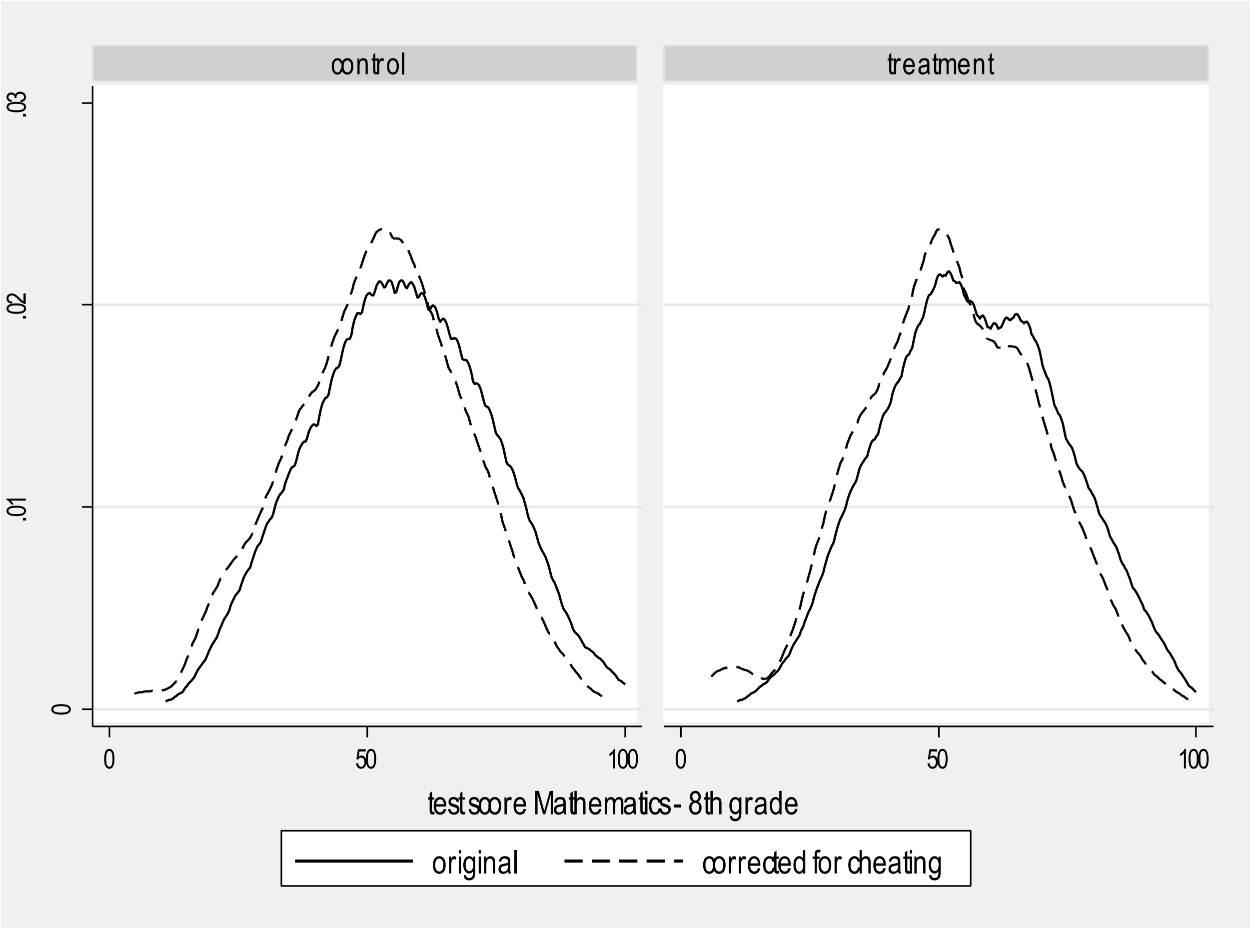

Figure 4: Distribution of the numeracy score by treatment status - end of grade 8 (sch.year 2011-12)

Appendix

Table A.1

Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Treatment	6744	0.51	0.50	0	1
Female	6744	0.47	0.50	0	1
birth year	6731	97.90	2.35	9	99
foreign born	6744	0.07	0.26	0	1
Disable	6744	0.03	0.18	0	1
at least one parent with college education	5680	0.31	0.46	0	1
Italian spoken at home	6105	0.77	0.42	0	1
number of books at home	6091	2.89	1.23	1	5
number of siblings	6189	1.26	0.91	0	4
two parent	6160	0.86	0.35	0	1
mother working	6025	0.70	0.46	0	1
father working	6002	0.96	0.20	0	1
test score Mathematics - Nov 2009 - 6th grade	6007	56.28	16.17	0.00	100
test score Mathematics - June 2012 - 8th grade – original	4880	56.73	17.41	10.87	100
test score Mathematics - June 2012 - 8th grade - corrected for cheating	4880	52.24	17.13	4.84	100
test score Italian - Nov 2009 - 6th grade	5978	65.00	15.15	0.00	100
test score Italian - June 2012 - 8th grade – original	4862	75.31	13.09	12.94	100
test score Italian - June 2012 - 8th grade - corrected for cheating	4880	69.39	17.15	0.00	100

Table A.2

The effect of the treatment Cl@ssi2.0 on test scores at 8th grade (original test scores) – class averages - OLS

original test scores	Literacy				Numeracy			
	1	2	3	4	5	6	7	8
	basic	extended	extended school FE	extended school FE	basic	extended	extended school FE	extended school FE
test score at entrance (6th grade – class average)	0.415	0.299	0.408	0.387	0.419	0.397	0.355	0.335
	[0.095]***	[0.099]***	[0.083]***	[0.080]***	[0.088]***	[0.125]***	[0.152]**	[0.154]**
treated class	**0.333**	**0.455**	**0.262**	**1.858**	**-0.13**	**-0.248**	**-0.053**	**1.618**
	[0.643]	**[0.628]**	**[0.503]**	**[1.199]**	**[1.058]**	**[1.067]**	**[1.003]**	**[2.005]**
probability of cheating (estimated)	18.009	17.749	11.754	12.468	40.719	36.75	29.166	29.359
	[2.311]***	[2.273]***	[1.622]***	[1.658]***	[3.749]***	[3.872]***	[7.008]***	[7.103]***
girls (share)		3.687	0.852	0.87		0.146	-1.452	-1.641
		[2.560]	[2.976]	[2.873]		[4.170]	[5.706]	[5.700]
birth year (class average)		-2.822	-2.336	-2.578		-0.182	-2.169	-2.329
		[3.146]	[4.116]	[4.048]		[5.564]	[7.955]	[8.005]
foreign born (share)		-5.447	-4.787	-4.397		-12.421	-3.019	-2.445
		[3.797]	[6.546]	[6.503]		[7.143]*	[18.231]	[18.466]
at least one college graduate parent (share)		8.334	8.017	10.938		5.994	6.631	9.76
		[2.211]***	[2.692]***	[3.187]***		[3.491]*	[5.691]	[6.490]
at least one college graduate parent (share) × treated class				**-5.081**				**-5.38**
				[3.130]				**[4.959]**
fraction of permanent teachers (still in the school after 3 years)		1.464	2.003	1.573		4.789	2.339	2.013
		[2.106]	[3.120]	[3.131]		[3.495]	[5.595]	[5.594]
Observations	268	268	268	268	268	268	268	268
R^2	0.48	0.52	0.83	0.83	0.42	0.48	0.78	0.79

robust standard errors clustered at school level – intercept and regional controls included

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A.3

The effect of the treatment Cl@ssi2.0 on test scores at 8th grade (test scores corrected for cheating by Invalsi) – class averages - OLS

corrected test scores	Literacy				numeracy			
	1	2	3	4	5	6	7	8
	basic	extended	extended school FE	extended school FE	basic	extended	extended school FE	extended school FE
test score at entrance (6th grade – class average)	0.043 [0.123]	-0.034 [0.131]	0.281 [0.195]	0.202 [0.194]	0.536 [0.101]***	0.397 [0.155]**	0.324 [0.189]*	0.297 [0.186]
treated class	**1.622** **[1.437]**	**1.935** **[1.461]**	**1.948** **[1.371]**	**8.084** **[3.077]***	**-0.38** **[1.302]**	**-0.318** **[1.359]**	**-0.064** **[1.335]**	**2.083** **[2.480]**
girls (share)		0.038 [5.126]	9.39 [8.521]	8.919 [8.018]		2.625 [4.902]	4.849 [6.355]	4.513 [6.421]
birth year (class average)		-0.087 [1.451]	1.3 [1.214]	1.852 [1.153]		0.095 [1.773]	-2.926 [1.478]**	-2.741 [1.486]*
foreign born (share)		2.425 [6.784]	-0.182 [11.643]	1.958 [11.636]		-8.748 [6.931]	1.311 [16.916]	2.237 [17.082]
at least one college graduate parent (share)		8.032 [4.658]*	7.857 [5.422]	19.064 [8.130]**		6.584 [4.937]	7.817 [8.577]	11.796 [8.952]
at least one college graduate parent (share) × treated class				**-19.654** **[8.543]****				**-6.894** **[5.621]**
fraction of permanent teachers (still in the school after 3 years)		-4.164 [4.542]	-10.029 [7.937]	-11.297 [7.738]		2.048 [4.449]	-3.479 [7.672]	-3.87 [7.751]
Observations	268	268	268	268	268	268	268	268
R^2	0.22	0.23	0.64	0.66	0.15	0.19	0.67	0.68

robust standard errors clustered at school level – intercept and regional controls included
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A.4

The effect of Cl@sse 2.0 on test scores by the end of 8th grade (original test scores) – individual data – decomposition by class location - OLS

	Literacy				Numeracy			
	1	2	3	4	5	6	7	8
	Classes in the same building	Classes in different buildings	Full sample	Full sample, school FE	Classes in the same building	Classes in different buildings	Full sample	Full sample, school FE
test score at entrance (6th grade)	0.512 [0.017]***	0.601 [0.088]***	0.514 [0.017]***	0.526 [0.015]***	0.521 [0.022]***	0.636 [0.115]***	0.525 [0.022]***	0.532 [0.018]***
treated class	-0.316 [0.761]	-2.338 [2.655]	-0.211 [0.747]	0.398 [0.534]	-0.722 [1.372]	2.971 [3.713]	-0.509 [1.343]	0.426 [0.903]
treated and control classes in different buildings			-3.2 [1.615]**	-0.727 [1.726]			-2.529 [3.139]	-0.96 [3.419]
probability of cheating (estimated at class level)	16.411 [1.960]***	42.744 [57.901]	16.445 [1.946]***	11.461 [1.154]***	37.827 [3.710]***	346.507 [22.94]***	38.41 [3.805]***	34.236 [5.174]***
Female	0.726 [0.359]**	1.524 [1.598]	0.766 [0.349]**	0.639 [0.332]*	-0.56 [0.458]	-0.926 [1.768]	-0.576 [0.454]	-0.489 [0.416]
birth year	0.145 [0.095]	-0.078 [1.601]	0.146 [0.096]	0.128 [0.101]	-0.039 [0.117]	2.816 [2.370]	-0.034 [0.119]	-0.076 [0.096]
foreign born	-1.401 [0.805]*	-1.153 [3.882]	-1.435 [0.789]*	-1.497 [0.783]*	-4.077 [1.047]***	-1.758 [5.409]	-4.329 [1.032]***	-3.595 [0.960]***
at least one college graduate parent	1.435 [0.345]***	2.337 [1.753]	1.488 [0.341]***	1.273 [0.366]***	2.357 [0.542]***	1.569 [3.319]	2.263 [0.532]***	2.219 [0.496]***
fraction of permanent teachers (still in the school after 3 years)	1.731 [1.838]	27.466 [8.750]***	1.791 [1.754]	2.327 [2.122]	2.892 [3.302]	-4.901 [13.871]	4.16 [3.168]	1.524 [3.765]
teacher in charge teaches language	1.754 [0.994]*	-0.176 [2.966]	1.661 [0.982]*	-0.112 [0.858]	0.142 [1.632]	0.668 [3.776]	-0.087 [1.618]	-2.84 [1.736]
teacher in charge teaches mathematics	0.093 [0.876]	1.169 [2.910]	0.127 [0.839]	-1.371 [0.852]	1.382 [1.965]	1.62 [3.695]	1.059 [1.848]	0.9 [1.408]
Observations	3980	191	4171	4171	3983	191	4174	4174
Number of schools	126	8	134	134	126	8	134	134
R^2	0.42	0.48	0.43	0.5	0.39	0.49	0.39	0.51

robust standard errors clustered at school level – intercept and regional controls included

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A.5

The effect of Cl@sse 2.0 on test scores by the end of 8th grade (original test scores) – individual data – heterogeneity of treatment by duration and by teachers' turnover - OLS

	1	2	3	4
	Literacy	Numeracy	Literacy	Numeracy
test score at entrance (6th grade)	0.527	0.505	0.516	0.526
	[0.024]***	[0.033]***	[0.017]***	[0.021]***
treated class	**0.095**	**-1.311**	**-1.658**	**3.072**
	[1.449]	**[2.903]**	**[2.702]**	**[4.312]**
treated class × short treatment	**-0.89**	**1.163**		
(less than 16 months)	**[1.650]**	**[2.552]**		
treated class × long treatment	**0.516**	**0.604**		
(more than 22 months)	**[1.297]**	**[2.500]**		
treated class × fraction of			**2.064**	**-5.585**
permanent teachers			**[3.782]**	**[6.423]**
probability of cheating (estimated at	18.076	39.206	16.575	38.523
class level)	[2.486]***	[5.758]***	[1.957]***	[3.854]***
Female	0.509	-0.77	0.765	-0.54
	[0.492]	[0.609]	[0.349]**	[0.457]
birth year	0.112	-0.033	0.144	-0.036
	[0.085]	[0.046]	[0.095]	[0.120]
foreign born	-0.09	-4.054	-1.388	-4.377
	[1.223]	[1.544]***	[0.785]*	[1.029]***
at least one college graduate	1.134	1.491	1.53	2.326
parent	[0.491]**	[0.759]*	[0.340]***	[0.533]***
fraction of permanent teachers (still	2.243	3.032	0.846	6.937
in the school after 3 years)	[3.067]	[4.727]	[2.872]	[4.368]
teacher in charge teaches language	2.527	-0.147	1.673	-0.076
	[1.541]	[2.250]	[0.992]*	[1.590]
teacher in charge teaches	-2.177	-0.999	-0.036	0.829
mathematics	[1.481]	[2.737]	[0.856]	[1.827]
Observations	2025	2026	4171	4174
Number of schools	64	64	134	134
R^2	0.43	0.42	0.42	0.39

robust standard errors clustered at school level – constant and regional controls included
* significant at 10%; ** significant at 5%; *** significant at 1%.
The duration of treatment is considered in columns 1 and 2.
Teachers' continuity is considered in columns 3 and 4.