social policy institute

www.employment.gov.sk/isp

Ministry of Labour, Social Affairs and Family of the Slovak Republic

Computer Availability in Slovak Primary Schools

A Potential Boost for Disadvantaged Children

Marcela Veselkova

October/2022

Abstract

This paper examines the impact of computer availability on educational achievement of Slovak fourth grade students in mathematics, science and reading, as measured by IEA's Trends in International Mathematics and Science Study 2019 and Progress in International Reading Literacy Study 2016. The conditional average treatment effect was estimated using a Bayesian additive regression trees model with a propensity score. All individual treatment effects of computer availability in mathematics and science lessons are statistically uncertain, i.e. it is impossible to rule out that computer availability has no impact on student performance. Computer availability in reading lessons has on average a medium positive impact on student performance: it improves reading scores by 0.10 to 0.13 standard deviations. The effect is higher for disadvantaged children, suggesting that computer availability has a potential to narrow educational inequalities.

Acknowledgments

I am grateful to Mária Balberčáková (Education Policy Institute), Matěj Bělín (Social Policy Institute), Imrich Berta (VšZP - General Health Insurance Company), Vincent Dorie (Code for America), Ján Komadel (Social Policy Institute), Marián Šaling (Institute for Financial Policy) and Katarína Vančíková (Matej Bel University) for helpful comments.

isp@employment.gov.sk

1 INTRODUCTION

Slovak primary and secondary schools were significantly computerized in the past two decades: the student-computer ratio fell from 12.2 students per computer in 2004 to 2.3 in 2021 (CVTI 2022). However, the existing research suggests that computer availability per se does not help students learn (Bulman and Fairlie 2016). What matters is how computers are used in the classroom (Comi et al. 2017; Falck et al. 2018) or whether they are used at all (Barrera-Osorio and Linden 2009).

This paper examines the impact of computerization by examining computer availability during mathematics, science and reading lessons in Slovakia, using the data from the 2019 round of eTIMSS and the 2016 round of PIRLS testing. Assignment to treatment was based on the teacher's answer to the question: Do the students in this class have computers available to use during their lessons?

The conditional average treatment effect was estimated using a Bayesian modeling procedure, known as Bayesian Additive Regression Trees (BART). This approach was originally developed by Chipman et al. (2010) and first used for estimation of a causal effect by Hill (2011). This method was previously used to estimate the impact of computer use in Spanish and Italian schools using OECD's PISA data (Cabras and Horrillo 2015, 2016, Ferraro 2018). However, regularized models originally designed for prediction can bias causal estimates in the presence of confounding, especially when this confounding is driven by selection into treatment based on potential outcomes (Hahn et al. 2018, Hahn et al. 2020). Targeted selection is of concern because computer availability in schools is likely related to the socioeconomic status of the neighborhood: the more affluent the neighborhood, the better the information technology equipment. Vice versa, teachers may assign the computer use to lower achieving students more frequently in the hope that these students obtain more feedback through educational software (Papanastasiou et al. 2003). To control for the possible bias in the presence of confounding, I use Bayesian additive regression trees model with a propensity score (Dorie and Hill 2020).

To anticipate the results, individual treatment effect estimates of computer availability in mathematics and science lessons are statistically uncertain: their 90% credible intervals contain zero, i.e. it is impossible to rule out the possibility of no effect. However, computer availability for reading lessons has a positive impact on student performance. It improves the reading scores by 0.10 to 0.13 standard deviations. The effect is the highest for children attending schools, in which all children were eligible for free school meal.

The rest of the paper is organized as follows. Section 2 reviews the existing theoretical and empirical literature on the impact of computer availability and computer use in schools. Section 3 discusses other factors affecting student outcomes. Section 4 describes the Bayesian additive regression trees model and its application in the causal inference setting. Sections 5 to 7 discuss the estimated treatment effects of computer availability in mathematics, science and reading lessons. Finally, section 8 concludes.

2 IMPACT OF COMPUTER USE IN THE CLASSROOM

A standard model of education production views educational outcomes, such as test scores or earnings, as a function of family and school-related inputs (Bulman and Fairlie 2016; Hanushek 2020). Holding all other inputs constant, ICT increases the overall study time (Barrow et al. 2009). In the traditional classroom, the teacher has to divide his time between group and individual instruction. Although the teacher's time allocated to individual instruction of one student cannot be allocated to individual instruction of another student, educational software can supplement the teacher during periods when he instructs others. Furthermore, educational software enables students to work at their own pace, which is beneficial particularly for low-achieving students who might struggle during group instruction (Koedinger et al. 1997; Jacob et al. 2016).

These positive effects may be offset by unproductive computer use (Belo et al. 2014). For example, children may distract themselves with videos, social networking and other activities, which do not improve their performance. Furthermore, if the existing mix of inputs is optimal, investment in computer hardware, software and connectivity will lead to reallocation of funds from more effective inputs, such as teachers or traditional instructional materials (Bulman and Fairlie 2016).

Empirical evidence mirrors the theoretical indeterminacy. Early studies suggest a positive impact of ICT on student scores (for a review, see Kirkpatrick and Cuban 1998). However, majority of these studies fail to address the endogeneity issue: students who use ICT and those who do not may have different observed and unobserved characteristics that may also influence their outcomes. For example, family background may influence both the computer use and student performance. Likewise, students themselves decide how frequently they will use the computer at home and this decision is most likely correlated with their individual characteristics.

Recent studies therefore exploit exogenous variation in ICT investment (Angrist and Lavy 2002; Goolsbee and Guryan 2006; Leuven 2007; Malamud and Pop-Eeches 2010) or computer use (Cabras and Horrillo 2016; Crook et al. 2014; Grimes and Warschauer 2008; Silvernail and Gritter 2007; Shapley et al. 2011; Spiezia 2010; Suhr et al. 2010), analyze randomized experiments (Banerjee et al. 2007; Barrera-Osorio and Linden 2009; Barrow et al. 2009; Lai et al. 2012; 2013; 2015; Rouse et al. 2004) or control for the endogeneity of ICT investment and use through instrumental variable techniques (Machin et al. 2007; Belo et al. 2014).

This body of literature shows mixed results. Studies, which examine investment in ICT tend to document no or even negative impact (Angrist and Lavy 2002; Leuven et al. 2007; Goolsbee and Guryan 2006; Belo et al. 2014). This supports the view that it is not computer availability per se that matters but rather how computers are used in the classroom (Comi et al. 2017; Falck et al. 2018), or whether they are used at all (Barrera-Osorio and Linden 2009). When looking specifically at computer use by students, effects tend to be mostly positive (Cabras and Horrillo 2016; Crook et al. 2014; Grimes and Warschauer 2008; Silvernail and Gritter 2007; Spiezia 2010; Suhr et al. 2010; but see Shapley et al. 2011). However, randomized controlled trials of educational software demonstrate that both the type of the technology and what is being substituted influence effect size and direction (Linden 2008; Mo et al. 2014).

Research has also focused on the impact of computer use on children's social development. On the one hand, computer use substitutes personal interactions and may reallocate time from sleep and recreational activities, such as sports or social activities (OECD 2015: 43). On the other hand, computer use may facilitate subsequent interpersonal interactions, especially for individuals with social anxiety or people with autism spectrum disorder (Pierce 2009; Ziv and Kiasi 2016; Valencia et al. 2019). (Quasi-)experimental studies tend to support the view that computer use has little impact on children's participation in extra-curricular activities such as sports, music or arts (Bauernschuster et al. 2014; Fairlie and Kalil 2017; Malamud and Pop-Eleches 2011). However, there is some evidence that technology use has a negative impact on social skills. In one study, elementary school children spent five days in an educational camp without any electronic devices (Uhls et al. 2014). Children in the experimental group had more in-person interactions and scored significantly higher at reading emptions compared to the control group. Fostering traditional, face-to-face interactions among students may, therefore, help develop their social skills (Savina et al. 2017).

2.1 Impact of Computer Availability

Angrist and Lavy (2002) examine the impact of a large-scale computerization policy in elementary and middle schools in Israel and find that an increase in computer availability did not translate into higher test scores. In fact, the effect on the math scores of fourth graders was negative. Authors hypothesize that computer use may have displaced more effective educational activities or consumed school resources that might have prevented a decline in achievement.

Leuven et al. (2007) come to a similar conclusion. They exploit a policy that provided subsidy for the purchase of computers and software to schools with more than seventy percent disadvantaged students in the

Netherlands. Extra funds for computers and software did not have a positive impact on student scores and even had a negative effect on language and arithmetic scores.

Mixed effects were documented also in developing countries. For example, Cristia et al. (2014) found no effect of hardware and software availability in Peruvian primary schools on grade repetition, dropout rates or enrollment into secondary schools. Similarly, Barrera-Osorio and Linden (2009) conducted a randomized evaluation of a project, which aimed to integrate computers into the teaching of language in Colombian public schools. Computer availability failed to improve scores in Spanish language, most likely because teachers did not incorporate the computers into their curriculum. Nevertheless, computer availability improved student performance in computer science classes.

Investment in better connectivity is not associated with improved student performance either. Goolsbee and Guryan (2006) explore a US program, which subsidized schools' investment in internet and communication. They find that an increase in internet connections did not have any significant impact on students' performance. Likewise, Belo et al. (2014) show that students in Portuguese schools with higher levels of broadband score significantly lower in standardized tests. The negative effect is stronger in schools, which do not block distracting content, such as YouTube.

Positive impact of ICT investment was documented by Machin et al. (2006) who exploit a change in how ICT funds are allocated to primary schools in England. They observe a positive impact on student performance in English and science but not in mathematics. However, the authors note that the results may reflect the fact that resources were directed to schools with better educational standards and therefore more likely to use them efficiently.

Evaluations of computer availability have a shortcoming in that they do not take into account how computers are used – or whether they are used at all. These evaluations may, therefore, underestimate the true impact of integrating technology into curriculum. A more promising approach to evaluating the impact of ICT in schools is to look explicitly at computer use by students.

2.2 Impact of Computer Use

Large one-to-one laptop programs were implemented in several U.S. states with mixed outcomes. Silvernail and Gritter (2007) evaluate a Maine's Middle School Laptop Program, which provided students and their teachers with laptops and assisted teachers with integrating laptops into the educational process. Laptop use improved writing performance by one third of a standard deviation and the effect was stronger for students who used laptops more frequently. In Texas, laptop use failed to improve middle school students' reading or mathematics achievement (Shapley et al. 2011). In Farrington School District in California, test scores declined in the first year of computer use but increased in the second year, offsetting the initial decline (Grimes and Warschauer 2008; Suhr et al. 2010).

In Australia, being schooled with 1:1 laptops had a positive impact on student attainment, with a medium effect size in physics and small effect sizes in biology and chemistry (Crook et al. 2014). A follow-up survey among students and teachers revealed that computers were used more often for simulations and spreadsheets in physics classes compared to biology or chemistry classes.

Another set of computer use evaluations draws on OECD's Programme for International Student Assessment (PISA), which includes questions about the computer use and provides rich amount of information on student, school and parent background. Spiezia (2010) finds that greater frequency of computer use reported in the 2006 round of PISA is associated with higher science scores in OECD countries. However, this effect is stronger when computer is used at home rather than at school. Cabras and Horrillo (2016) compare Spanish students who use a computer at school and those who do not. They report a moderate positive effect on math scores as measured in the 2012 round of PISA, although it is impossible to rule out the possibility of no effect. The effect is stronger in case of students from a low socioeconomic background, suggesting that computers help to reduce the performance gap stemming from differences in the socio-economic background. Replication of this study using Italian data from the 2012 round of PISA testing supports the view that the general computer use in school improves mathematics scores (Ferraro 2018). The novelty of these studies is in their use of a non-parametric modeling approach Bayesian additive regression trees to estimate the causal effect (Cabras and Horrillo 2015, 2016, Ferraro 2018). A possible shortcoming is related to the fact that both studies fail to control for possible confounding.

2.3 Impact of Computer-Assisted Learning

Computer-assisted learning (CAL) aims to enhance learning with educational software, such as tutorials, drills, exercises or immediate feedback. CAL is designed to improve skills in a specific domain and can be easily randomized at a classroom or school level. However, one should note that many experiments evaluate supplemental use of educational software, often after the school (Banerjee et al. 2007; Lai et al. 2012; 2013; 2015; Rouse and Krueger 2004). In other words, any effect has to be interpreted as some combination of substituting CAL for traditional instruction and increasing instructional time (Bulman and Fairlie 2016: 248-250; 253-6).

The evidence from developed countries suggests that CAL may be substituting resources from more productive traditional inputs. For example, evaluations of several reading and math software products in the United States find no or very limited effects (Campuzano et al. 2009; Dynarski et al. 2007; Rouse and Krueger 2004). However, Barrow et al. (2009) document positive effects of prealgebra and algebra software in the United States. In line with expectations, the effect is larger in bigger classes, in which computers increase individual instruction time.

In developing countries, where CAL substitutes teachers with less training or motivation and where computers are often unavailable at home, impact tends to be positive (Banerjee et al. 2007; Lai et al. 2015). Banerjee et al. (2007) tested an intervention in India, in which children used computer-based math program two hours per week, one hour during the class time and the other outside of school hours. The impact on math scores was positive, although it faded over time. To separate the effect of CAL as a substitute for or supplement to the existing curriculum, Linden (2008) tested two interventions in India, one delivered in-school and the other after-school. He found that the impact of the supplemental program was heterogeneous: it improved math scores performance of the poorest performing students but was ineffective for the rest of the children. These results could reflect the design of the learning program which emphasized reinforcement of material presented in class but did not offer much space for self-paced discovery of uncovered material. A pull-out in-school program had a negative impact on math scores of all types of participating students. This suggests that CAL was a poor substitute for the existing instruction, probably because the program was tested at high-quality NGO-run schools.

In contrast to Linden (2008), Mo et al. (2014) tested the impact of CAL in China's rural schools with shortages of qualified teachers or learning materials. There, the impact was positive, suggesting that CAL did not replace productive classroom activities run by teachers. Out-of-school CAL programs implemented in China had a positive impact on student learning (Lai et al. 2012; 2013; 2015). The effect was stronger in case of children with lower socio-economic background (Lai et al. 2013; 2015) and low-performing students (Lai et al. 2012), suggesting that well-designed and implemented CAL has a potential to reduce socio-economic inequalities.

2.4 Impact of Computer-Based Teaching Practices

Why does computer use have positive impact in some cases and null or even negative impact in other cases? This question motivated some researchers to examine the impact of specific pedagogical uses of digital technology.

Falck et al. (2018) argue that the little to no effect of classroom computers on student achievement reflects the heterogeneous effect of various computer-based teaching practices. Using data from the 2011 Trends in International Math and Science Study (TIMSS), they find positive effects when computers are used for activities that do not have an alternative in the traditional classroom, such as looking up information on the Internet. In contrast, when computers substitute activities with a potentially more effective conventional teaching alternative, such as practicing skills, the impact is negative.

This view is supported by Comi et al. (2017) who examined various teaching practices in the Italian setting. They find that computer-based teaching practices improve student score when they are aimed at increasing students' awareness of ICT use and at improving their critical skills. In contrast, practices requiring an active role of students in the use of ICT have a negative impact, particularly in the case of Italian language.

3 OTHER FACTORS AFFECTING STUDENT OUTCOMES

In observational studies, such as IEA's TIMSS and PIRLS, potential outcomes are typically not independent of the treatment assignment. If students from better-off families are allocated to schools with computers, then it is the socio-economic status rather than treatment that improves their performance in the test. If schools with disadvantaged students are more likely to receive subsidy for the purchase of computers, the effect will be driven by the socio-economic status again. In order to ensure the strong ignorability of treatment assignment, it is necessary to control for all the possible confounding variables, such as school factors or factors beyond the school system.

3.1 Factors beyond the School System

OECD's PISA reveals that gender matters in explaining the student performance in a wide range of countries (OECD 2016). On average, boys perform better than girls in mathematics but worse in reading, although the gender gap in reading narrowed somewhat between 2009 and 2015. The variation in the gender-gap across countries implies that there are no differences in innate abilities between girls and boys. Rather, the gender-gap reflects gender stereotypes and is eliminated in more gender-equal cultures (Guiso et al. 2008; Fryer and Levitt 2010; González de San Román and de la Rica 2012). However, most of the existing studies fail to distinguish between the impact of institutional constraints and the culture. Therefore, Nollenberger et al. (2016) compare the test scores of second-generation immigrants who were exposed to the same institutions in the host country. They find that immigrant girls whose parents come from more gender-equal countries perform better than immigrant girls whose parents come from less gender-equal countries. In other words, the gender-gap reflects the cultural beliefs of their parents on the role of women in society.

Student performance is influenced also by the family's socio-economic status (Willms & Somers, 2001; Woessmann, 2004; OECD, 2016, Jæger and Breen 2016). On average, children of better educated parents in better-paid jobs are more likely to succeed at school because they benefit from various financial, cultural and social resources, such as tutoring, active parenting or networks (OECD 2016: 206). However, the effect varies between countries and tends to be stronger in educational systems, which sort students into different schools based on their abilities at an early age (Hanushek and Woessmann, 2006; Pfeffer, 2008; Schuetz et al., 2008; Woessmann, 2009). In countries with an early age of tracking, such as Slovakia, the future of the child may be decided as early as the age of 11.

The skills today are built upon skills acquired in the past (Cunha and Heckman 2007, 2009). Parental preschool literacy and numeracy practices (such as reading, counting, playing words games or playing with number toys), are associated with increased literacy and numeracy skills (Gustafsson et al. 2013; Kleemans et al. 2012; LeFevre et al. 2009; LeFevre et al. 2010; Skwarchuk et al. 2014), as well as later performance in school (Ersan & Rodriguez 2020; Dunst et al. 2017; Mullis 2020).

Inadequate home learning environment can be compensated for by pre-primary education, which has a considerable positive short-term impact on cognitive skills although this effect fades out as children grow older (Burger 2010; Elango et al. 2015a). In the long-run, pre-primary education boosts graduation rates of females and employment of males, as well as lowers criminal activity (Heckman et al. 2010a, 2013; Campbell et al. 2014, and Elango et al. 2015b). Investment in early childhood education is particularly effective for disadvantaged children as it compensates for social inequalities (Black et al. 2014; Cornelissen et al. 2018; Drange a Havnes 2019; Felfe et al. 2015; Felfe a Lalive 2018; Fitzpatrick 2008; Gormley and Gayer 2005; Gormley et al. 2005; Havnes & Mogstad 2011, 2015; Hustedt et al. 2008; Wong et al. 2008). However, these children are the least likely to participate in early education and childcare in Slovakia (Buchel et al. 2022) and this is particularly true for Romani children from marginalized communities (Grauzelová & Markovič, 2020; Hellebrandt et al. 2019).

3.2 School Factors

The relationship between educational resources and student performance is not linear. Among moderately wealthy economies with per capita GDP up to around 50 000 USD, the greater country's wealth is associated with a higher mean score on the PISA reading test (OECD 2016c: 184-186). Once this minimum level has been reached, the evidence that additional resources have a strong impact on student outcomes is mixed (OECD 2016c: 186-193). At the school level, the direct effect of additional financial resources is debated as well. Whereas some document positive relationship between school funding and student performance (Battistin et al. 2013; Falzetti et al. 2012; Gibbons et al. 2012; Machin et al. 2010; Papke 2005), others find no or negative effect of additional school funding on student performance (Bénabou et al. 2009, Leuven et al. 2007; Van der Klaauw 2008). The latter is true particularly for smaller campaigns (Cellini et al. 2010; Martorell et al. 2015). This implies that how money is spent is more important than how much is spent. In the absence of detailed information about school resources, a simple student-teacher ratio is a useful proxy for resources devoted to education (OECD 2019: 377).

Teacher quality is an important contributor to student academic performance (Aaronson et al. 2007; Chetty et al. 2014a, Chetty et al. 2014b; Gerritsen and Plug 2017; Hanushek, 2010; Nye et al. 2004; Rivkin et al. 2005; Rockoff, 2004). Furthermore, teachers who improve test scores do not simply teach to the test but improve their students' outcomes also in adulthood: students of more effective teachers are more likely to attend college, earn higher salaries and less likely to have children as teenagers (Chetty et al. 2014b).

The most important observed teacher characteristic that matters for student performance is teacher experience, especially in the initial years of the teaching career (Gerritsen et al. 2016; Staiger and Rockoff, 2010; Hanushek, 2011; Chetty et al., 2011). However, it must be noted that the negative relationship between the teacher support and mathematics performance in some countries suggests that teachers might give higher levels of support to the lowest-achieving students, a fact that needs to be taken into account by interpretation of results (OECD 2010: 112).

Class size reduction could boost learning through more individualized instruction. However, empirical evidence about class-size effects from experimental (Krueger 1999; Nye et al. 2000; Nye and Konstantopoulos 1999; Chetty et al. 2011) or quasi-experimental studies (Angrist and Lavy 1999; Hoxby, 2000; Leuven et al., 2008; Leuven and Løkken 2017) is mixed. Positive effects of smaller class size on student performance was documented mostly by studies of the Tennessee STAR experiment (Nye et al. 2000; Nye and Konstantopoulos 1999; Chetty et al. 2011), which randomly assigned one cohort of more than 11 500 children and 1300 teachers to small (13 to 17 students) or regular (22 to 25 students) classes with or without a paid aide. Smaller classes improved test scores (Krueger 1999) and college attendance but not earnings (Chetty et al. 2011). Although the design and implementation of the STAR experiment was questioned (Hanushek 1999), the results of STAR experiment inspired efforts to reduce class sizes in the United States at both the federal and the state level.

4 METHODOLOGY

4.1 Problem Setup

Within Neyman-Rubin potential outcomes framework, each student is assumed to have two potential outcomes (Neyman 1923, Rubin 1974). Let Y_i denote a test score of *i*th student, for a population of *i* = 1, … *n*. Suppose that the *i*th student is characterized by a length *p* vector of observed control variables x_i, such as gender, socio-economic status or language spoken at home. A binary treatment variable Z takes the value 1 if the student used the computer at school (treatment) and 0 otherwise (control). The causal effect of computer availability for student *i* is then defined as a difference between $Y_i(1)$ and $Y_i(0)$, which would be observed under Z = 1 and Z = 0, respectively.

Note that it is impossible to observe both outcomes for a single student simultaneously because no student can both use and not use a computer in school. It is only possible to observe outcomes, which correspond to the observed treatment, i.e. $Y_i = Y_i(1)Z_i + (1 - Z_i)Y_i(0)$. As there is only one observed outcome per student, it is necessary to consider multiple students. However, students who use and who do not use computers may differ with regard to pretreatment characteristics X_i that also affect their test scores. Therefore, identification of treatment effects from observed data requires researchers to assume strong ignorability (Rosenbaum and Rubin 1983), which stipulates that $Y_i(0), Y_i(1) \perp Z_i \mid X_i$, and that $0 < \Pr(Z_i = 1 \mid x_i) < 1$. The first condition (ignorability) assumes that treated and control students with identical values on all observed covariates *x* have the same chance of receiving the treatment, and the second condition (common support or overlap) assumes that empirical counterfactuals exist for all observations. Under these assumptions, $E(Y_i(z) \mid x_i) = E(Y_i \mid x_i, Z_i = z)$ and the conditional treatment effect can be expressed in terms of observed (rather than potential) outcomes, i.e. $\tau(X_i) \coloneqq E(Y_i \mid x_i, Z_i = 1) - E(Y_i \mid x_i, Z_i = 0)$. These conditional expectations can be estimated using various regression models, which model $E(Y_i \mid x_i, Z_i = z_i) = f(x_i, z_i)$ directly.

4.2 Bayesian Additive Regression Trees

The response surfaces $E[Y_i(1)|x_i]$ and $E[Y_i(0)|x_i]$ will be estimated using Bayesian additive regression trees model (Chipman et al. 2010) for causal inference (Hill 2011). In contrast to the popular linear regression, which makes restrictive parametric assumptions regarding linearity and additivity, BART can flexibly fit even highly nonlinear response surfaces without making unwarranted parametric assumptions. In simulation studies, BART was documented to outperform several machine learning competitors such as random forests, boosting or neural nets (Chipman et al. 2010).

BART is a sum-of-trees model with a regularization prior on the model parameters and posterior computed with Bayesian backfitting Markov chain Monte Carlo algorithm (Chipman et al. 2010; for a recent review of development of the original algorithm, see Hill et al. 2019). A sum-of-trees model can be formally written as follows:

$$Y = \sum_{j=1}^{m} g\left(x; T_j, M_j\right) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

where Y is the vector of responses, x is a p-dimensional vector of predictors, and ε is a vector of noise. T denotes a tree structure consisting of interior nodes with decision rules and terminal nodes which do not split further. An example is shown in Figure 1, with interior nodes depicted as rectangles and terminal nodes depicted as ovals. M = (μ_1, μ_2, … μ_b) denotes a set of parameter values associated with each of the *b* terminal nodes of T. For continuous variables x_i, decision rules take the form $x_i < c$, with c acting as a splitting value, which sends observations either left or right. This process continues until the terminal node (otherwise known as a leaf) is reached and x value is assigned the μ_i value associated with that terminal node. Thus, function g(x; T_j, M_j)

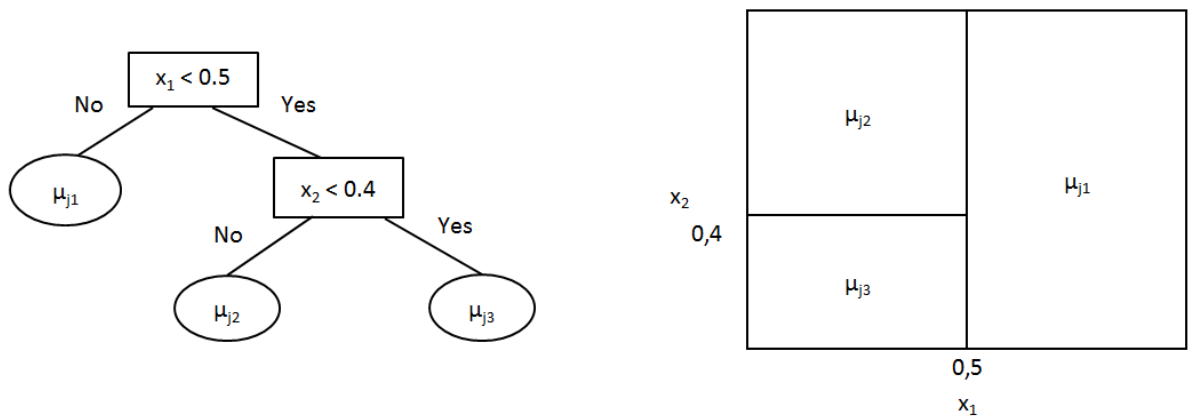
assigns a $\mu_{ij} \in M_j$ to x for each binary regression tree T_j and its associated terminal node parameters M_j. The sum of all the terminal node μ_{ij}'s assigned to x by the g(x; T_j, M_j)'s represents predicted mean of outcome Y given x , i.e. E(Y|x).

To avoid overfitting, a regularization prior constrains each tree to contribute only a small part to the overall fit. The T_j prior limits complexity of any single tree by enforcing shallow tree structures. The probability that a node at depth *d* is further split into two children nodes is given by

$$\alpha(1+d)^{-\beta}, \quad \alpha \in (0,1), \beta \in (0,\infty)$$

Default values α = 0.95 and β = 2 proposed by Chipman et al. (2010) set the probability of a tree with 1, 2, 3, 4 and over 5 terminal nodes at 0.05, 0.55, 0.28, 0.09 and 0.03, respectively, favoring small tree sizes of 2 or 3 over deep, complex trees. Decision rules occur in two steps. First, the predictor is randomly selected to serve as the splitting variable. Second, the splitting value is selected from the range of the selected predictor via the discrete uniform distribution.

Figure 1 An example binary tree, which uses two predictors to split the data into subgroups in the *j*th tree of an ensemble.



Note: Left: An example binary tree, with internal nodes labelled by their splitting rules and terminal (leaf) nodes labelled with the corresponding parameters μ_{j1}. Right: The corresponding partition of the sample space and the step function $g(x;\ T_j;\ M_j)$.

The $\mu_{ij}|T_j$ prior shrinks the leaf parameters towards the center of the distribution of the outcome. Chipman et al. (2010) first rescale the response variable Y to lie between -0.5 and 0.5 and set the mean of μ_{ij} to 0. The prior on each terminal node parameter is given by

$$\mu_{ij} \sim N\left(0, \sigma_\mu^2\right) \text{ where } \sigma_\mu = 0.5/(k\sqrt{m})$$

Thus, increasing *k* shrinks the value of σ_μ^2 and leads to a stronger regularization of μ_{ij}. The default value of k = 2 proposed by Chipman et al. (2010) assigns a 95% prior probability that the expected value of the response variable lies between the specified range of -0.5 to 0.5.

The σ prior, which controls the error variance, is given by

$$\sigma^2 \sim InvGamma(v/2, v\lambda/2)$$

where λ is determined from the data so that there is a q = 90% chance that the BART model will improve upon the RMSE from a linear regression fit to the data. Chipman et al. (2010) recommend the default setting $(v,q) = (3, 0.90)$, which avoids extremes.

The default hyperparameters α, β, v and q provide good performance compared to other popular machine learning algorithms (Chipman et al. 2010; Hill 2011). However, when hyperparameters are chosen via cross-validation from plausible values, BART performs better than its competitors (Chipman et al. 2010).

4.3 BART for Causal Inference

Hill (2011) proposed to use BART for causal inference by treating the treatment variable *z* as a covariate and fitting the model $y_i = f(x_i, z_i) + \epsilon_i$. Estimation of the causal effect requires fitting BART to the observed data and making predictions $E[Y_i(1)|x_i] = f(x_i, 1)$ for treated and $E[Y_i(0)|x_i] = f(x_i, 0)$ for control students. In the next step, the treatment status is switched so that treated students are now assumed to be in the control group and vice versa, and another set of predictions is made. The difference between the predicted outcomes under the treatment and under the control for each student represents an individual treatment effect. A simple mean of these individual treatment effects represents an average treatment effect (Hill 2011).[1]

This approach was documented to correctly infer causal effects in simulation studies (Dorie et al. 2019; Hill 2011, Hill and Su 2013; Wendling et al. 2018). However, Hahn et al. (2020) observe that BART, which was originally designed for prediction, may produce biased estimates of causal effects in the presence of moderate to strong confounding. A natural solution to this problem is to include the propensity score as a covariate passed to the BART model. There are two such modifications of BART, PS-BART (Dorie et al. 2019) and BCF, short for Bayesian Causal Forest (Hahn et al. 2020). Both methods outperformed their competitors in data analysis challenges (Dorie et al. 2019, Hahn et al. 2020). PS-BART and BCF had the lowest estimation error for CATE and PS-BART edged BCF in terms of coverage, although BCF had shorter intervals (Hahn et al. 2020).

4.4 Modeling Strategy

The treatment effect was estimated using R package "bartCause", version 1.0-5 (Dorie and Hill 2020) in R version 4.2.0 (R Core Team 2022). The response surface was fit using BART; a vector of propensity scores was supplied.

The propensity score model should include all variables related to the outcome to decrease the variance of an estimated exposure effect without increasing bias (Brookhart et al. 2006). In this paper, variable selection for propensity score estimation was automated via a variable selection model (see Bon 2022; Hahn et al. 2020). I regressed outcome against covariates (excluding treatment variable) using R package "BART", version 2.9 (McCulloch et al. 2021) to select a subset of covariates most associated with the outcome. Then I used these covariates to estimate propensity scores $\Pr(Z = 1|X)$ using covariate balancing propensity score (CBPS) methodology, which models treatment assignment while optimizing the covariate balance (Imai & Ratkovic 2014). The propensity scores were estimated using R package "WeightIt", 0.13.1 (Greifer 2022). Balance diagnostics is available in Appendix 1. Estimated propensity scores were then included among covariates passed to the BART model used to estimate the conditional average treatment effect.

Cross-validation was used to choose the optimal end-node prior parameter *k* (see Dorie and Hill 2020). The following values of *k* were selected: k = 0.8534836 for mathematics, k = 1.665375 for science and k = 2.398489 for reading. Other than that I used default prior specification with 75 trees, which has been found to work well across a wide variety of settings (Chipman 2010; Hill 2011). I ran 10 chains with 10 000 iterations each (in addition to 2 000 burn-in iterations). The chain convergence was assessed using the package "stableGR", version 1.1, which calculates stable Gelman-Rubin convergence diagnostic for Markov chain Monte Carlo (Knudson and Vats 2021; see also Vats and Knudson 2018). In all cases, the potential scale reduction factor was close to 1, indicating that the sample collected by the Markov chain has converged to the target distribution. Convergence was achieved also according to the function *n.eff*, which calculates effective samples size for a

[1] This approach was used by Cabras and Horrillo (2015, 2016) and Ferraro (2018) in their evaluation of computer use by 15-year-olds in Spain and Italy.

set of Markov chains using lugsail variance estimators. Further visual convergence diagnostics is available in Appendix 2. Common support was enforced by excluding observations with high posterior uncertainty, using 1 standard deviation rule (Hill and Su 2013).

The causal estimand is the conditional average treatment effect for individual i (henceforth “ICATE”), which is defined as a difference between the (samples from the posterior of) the expected value under the treated and the control condition. Results of estimation were summarized with posterior means and 90% credible intervals, which were computed as highest density interval (HDI) from Markov chain Monte Carlo (MCMC) samples. Individual HDIs were computed with R package “bayestestR”, version 0.12.1 (Makowski et al. 2020).

5 DATA

The impact of computer availability on student achievement in mathematics, science and reading is assessed using international assessments undertaken by the International Association for the Evaluation of Educational Achievement. Trends in International Mathematics and Science Study (TIMSS) is a quadrennial assessment of mathematics and science achievement of the 4th and 8th grade students. Progress in International Reading Literacy Study (PIRLS) is a quinquennial assessment of reading achievement of the 4th grade students. Both databases contain background information about students' home environment for learning, school climate and resources, as well as teaching practices.

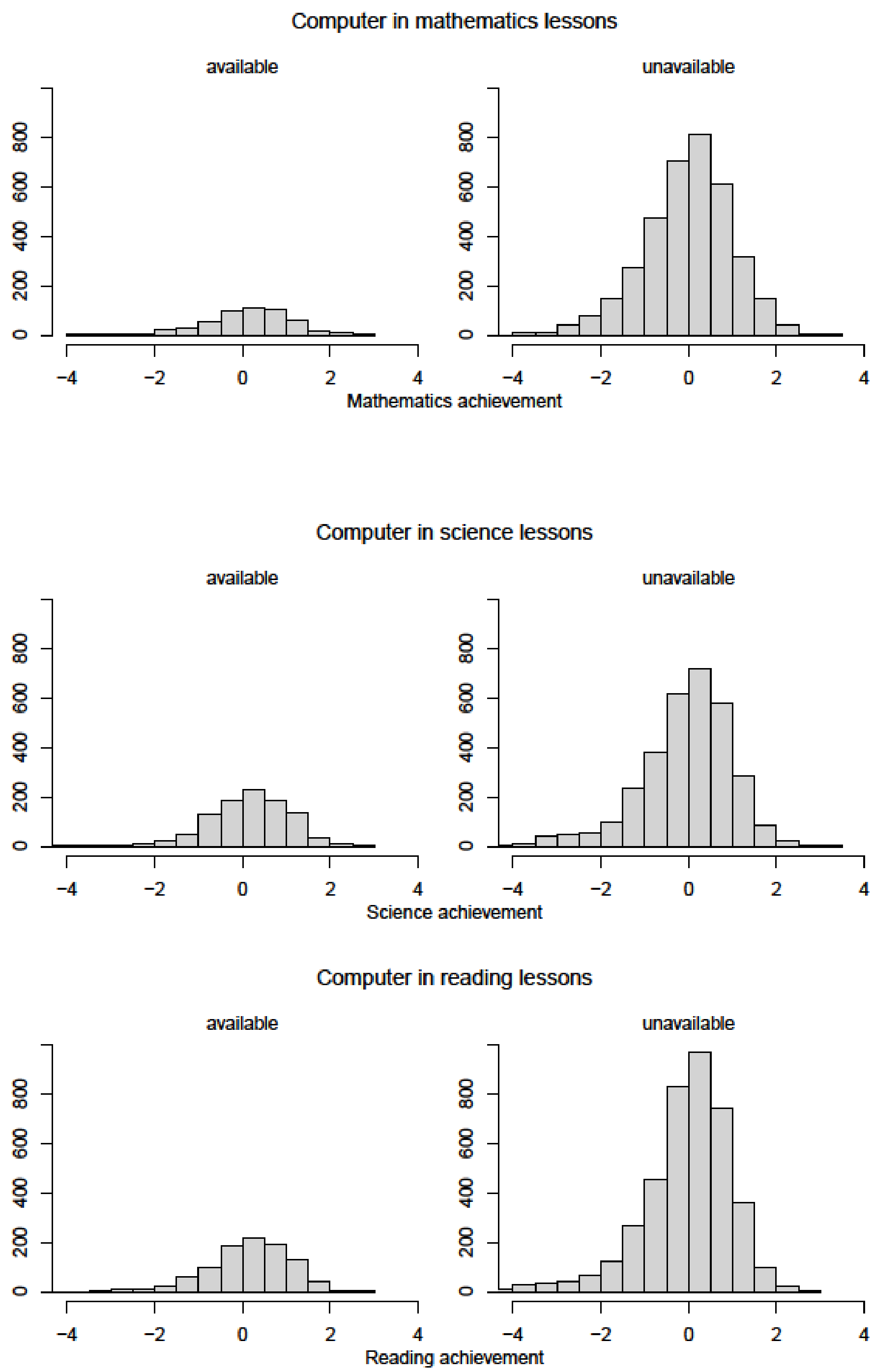
This paper uses the latest 2019 round of eTIMSS and the 2016 round of PIRLS. Assignment to treatment was based on questions ATBM04A and ATBS03A in the eTIMSS Teacher Questionnaire and question ATBR14A in the PIRLS Teacher Questionnaire: “Do the students in this class have computers (including tablets) available to use during their mathematics/science/reading lessons?” Pupils, whose teacher indicated that the computer or tablet was available during their lessons, were considered treated. Pupils of teachers who answered “No” were assigned to the control group. In the subsequent question (ATBM04CA/ATBS03CA), teachers were asked how often they used activities on computers during mathematics/science lessons to support learning for the whole class. Those who answered “never or almost never” were excluded from the sample because the frequency of “never or almost never” for the treatment variable for this group does not imply no computer availability as it does for those not in the intervention group.

All students with missing information about computer availability and other characteristics were excluded from the sample. Following the matching, two teachers per mathematics class were left in case of one school (ID = 5031). This school was excluded from the sample as well. In the final sample, computers were available to 12.4% of 4 168 students in mathematics lessons, 23.9% of 4 158 students in science lessons and 19.4% of 5 008 students in reading lessons. Figure 2 shows the sample distribution of the students’ test score conditional on treatment status.

The dependent variable was defined as the plausible value in mathematics, science or reading achievement. These plausible values represent the posterior distribution of student proficiency given achievement and contextual data. Using only one of plausible values or an average of plausible values as outcome may lead to some estimation errors. Therefore, average treatment effects were estimated for all five plausible values in every subject. Plausible values were standardized, such that the mean equals to zero and standard deviation to one. This allows for comparison across plausible values and subjects, as well as with results from other studies.

Selection of potential confounders was based on the overview of literature in the second section of this paper. Several proxies of socio-economic background were included. First, eTIMSS dataset includes variable “Home Resources for Learning”, which is based on students' and parents' responses concerning the availability of books in the home, the highest level of education of parents and the highest level of occupation of parents. Other proxies of home environment support include early literacy and numeracy activities before beginning primary school, a dummy variable of preschool attendance, and a variable, which indicates whether students speak language of test at home. Apart from these variables, I controlled also for the gender of the child. Finally, I controlled for the computer use for schoolwork (including classroom tasks, homework, studying outside of class), either specifically for mathematics or science (eTIMSS 2019) or in general at home, in school or elsewhere (PIRLS 2016). It must be noted that there were children who reported computer use for schoolwork (presumably outside the classroom) although their teacher did not indicate computer availability during lessons and vice versa, there were children who indicated that they did not use computer for schoolwork although their teacher reported computer availability during lessons. Categorical variables (e.g., language spoken at home) were turned into binary variables. Outliers for continuous variables were winsorized at 0.05 and 0.95 percentile.

Figure 2 Histogram of the test score for those students who use computer in mathematics, science and reading lessons (left panel) and those who do not (right panel).



Note: Plausible values are standardized for each subject, such that the mean equals to zero and standard deviation to one.

Students and teachers were matched using variable "Teacher ID and Link" (IDTEALIN), which links each teacher to a class and to certain students within this class. This enables analysis of student outcomes in relation to teacher-level variables, such as teacher's experience measured in years he or she has been teaching. Information about the class size was taken from the teacher's questionnaire as well. However, this variable was not used in the main analysis due to high share of missing data. For example, in case of mathematics, inclusion of class size would reduce the sample size by 16%.

Finally, to capture the socio-economic background of the school, I controlled for the school composition by socioeconomic background reported by the school principal in the eTIMSS 2019 School Questionnaire (affluent, average or disadvantaged) or the share of pupils eligible for free school lunch (none, some, all) in the PIRLS 2016 School Questionnaire. In 2015, all Slovak children were eligible for free school lunch in schools with at least 50 percent of children from families receiving social assistance benefit in material and social deprivation or whose household income did not exceed subsistence minimum. Finally, I controlled for whether instruction was affected by math, science or reading resource shortages.

Descriptive statistics by treatment status is listed in Appendix 3. On average, children whose teachers use computers in mathematics lessons score slightly higher in mathematics achievement. They are more likely to attend a school with neither disadvantaged nor affluent student body and are less likely to speak the language of instruction at home only some of the time. Children whose teachers use computers in science lessons also score higher in science achievement. Furthermore, they are more likely to attend schools with affluent student body. Finally, children whose teachers use computers in reading lessons score slightly higher in reading achievement. They are also more likely to attend schools in which no children are eligible for free school lunch. To control for possible selection into treatment, I include propensity score as a covariate. Additional information about data sources and variable construction is available in Appendix 4.

6 RESULTS

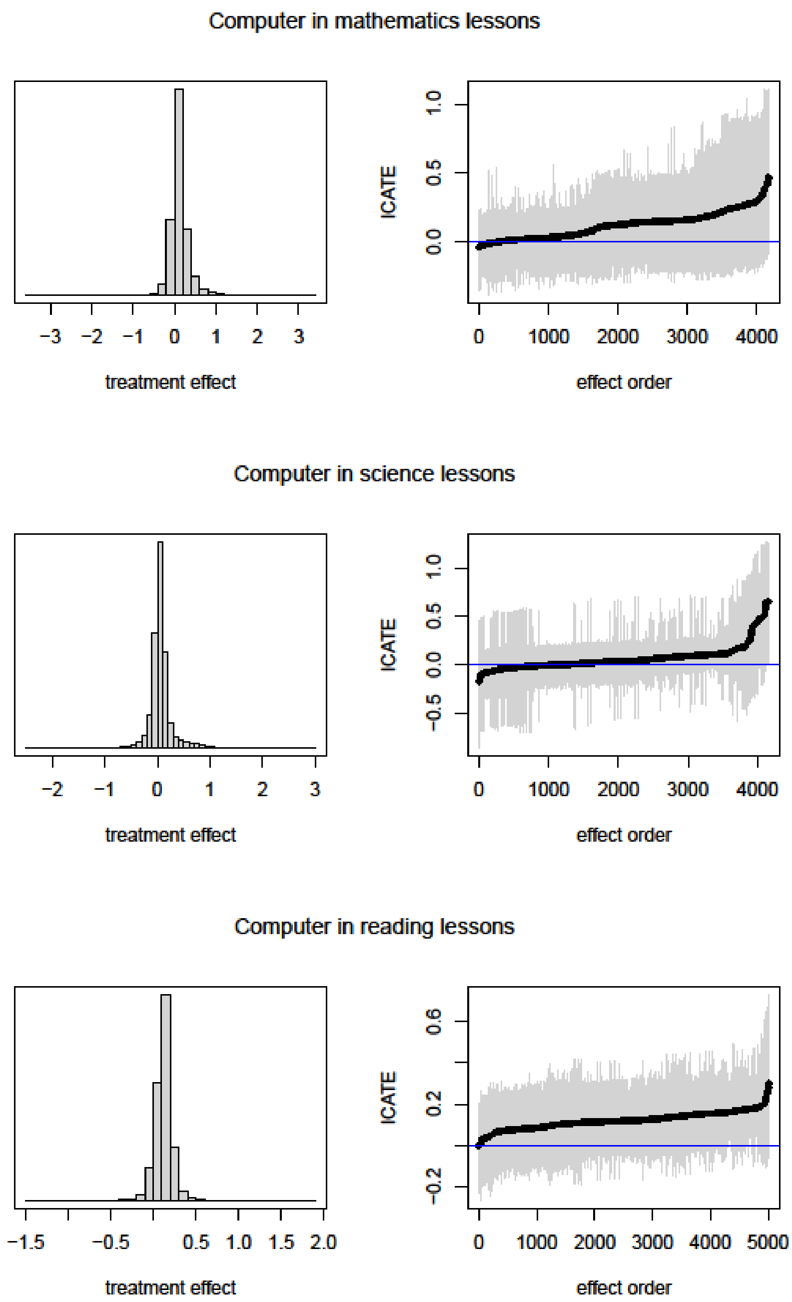
Computer availability during mathematics and science lessons has no impact on student performance (Table 1). Positive impact was estimated for mathematics only using plausible value 5 and for science only using plausible value 1. However, even in these two cases, almost all individual treatment effects are statistically uncertain: their 90% credible intervals contain zero, i.e. it is impossible to rule out the possibility of no effect. Computer availability for reading lessons has on average a positive impact on student performance: it improves the reading scores by 0.10 standard deviations with a 90% credible interval of 0.05 to 0.16 (plausible value 3) to 0.13 standard deviations with a 90% credible interval of 0.07 to 0.19 (plausible value 2); 24% (plausible value 1) to 63% (plausible value 5) of individual treatment effects are statistically certain, i.e. their credible intervals do not contain zero.

Table 1 Estimates of Treatment Effect of Computer Availability Using Various Plausible Values

Plausible Value	Mathematics	Science	Reading
PV1	0.04 (-0.04, 0.13)	0.07 (0.01, 0.13)	0.12 (0.06, 0.18)
PV2	0.07 (-0.01, 0.16)	0.05 (-0.02, 0.11)	0.13 (0.07, 0.19)
PV3	0.01 (-0.08, 0.09)	0.03 (-0.03, 0.09)	0.10 (0.05, 0.16)
PV4	0.05 (-0.03, 0.14)	0.04 (-0.02, 0.10)	0.11 (0.05, 0.17)
PV5	0.13 (0.04, 0.21)	0.05 (-0.01, 0.11)	0.12 (0.06, 0.17)

Since estimated treatment effects of computer availability in reading lessons are almost identical regardless of the plausible variable considered, all results shown in the remainder of the paper are based only on the first plausible value. In case of mathematics and science, results are shown for the fifth plausible value in mathematics and the first plausible value in science. This decision was motivated by the subsequent subgroup analysis: when the average treatment effect is zero, one should not read too much into the evidence of subgroup differences. Nevertheless, the findings of the subgroup analysis of computer availability in mathematics and science lessons should be still regarded as exploratory or hypothesis generating.

Figure 3 Estimated treatment effect

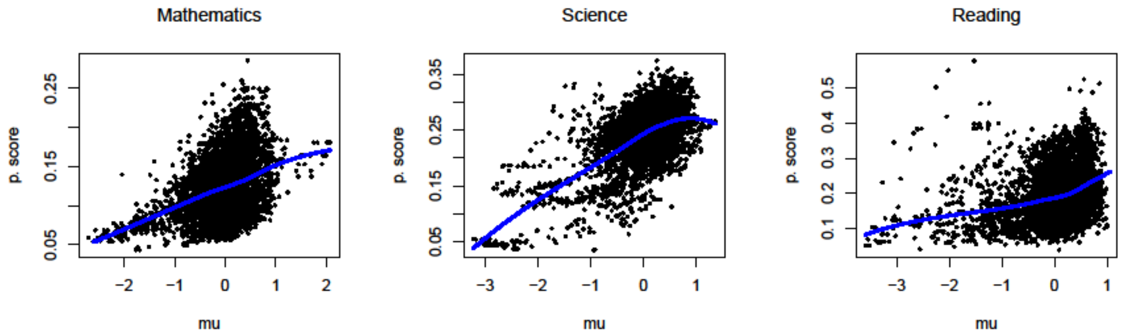


Note: Left panel: Posterior distribution of treatment effect of computer availability. Right panel: Point estimates of individual treatment effects (taken as the average of their posterior samples) ordered from smallest to largest. The gray lines correspond to the posterior 90% credible intervals.

The estimated BART model explains around 40% of the variability of mathematics achievement, 48% of the variability of science achievement and 47% of the variability of reading achievement. For comparison, linear regression model explains around 31% of the variability of mathematics achievement, 40% of the variability of science achievement and 38% of the variability of reading achievement. Comparison of BART fit to linear model is available in Appendix 5.

I also investigate the presence of confounding due to targeted selection (Hahn et al. 2020). Figure 4 shows a LOWESS trend between the estimated propensity score and a prediction of the outcome in the absence of treatment $\mu(x) = E(Y|Z = 0, x)$. The trend indicates that a higher expected test score is predictive of computer availability.

Figure 4 Targeted selection



Note: Each black dot depicts the estimated propensity score and a prediction of the outcome in the absence of treatment. The blue line depicts a locally weighted scatterplot smoothing (LOWESS) trend fit to these points.

To analyze the effect heterogeneity, I fit a regression tree for the individual effects given all covariates (using a recursive partitioning algorithm in the R package “rpart”, version 4.1.16 by Therneau et al. (2019a)).[2] Subgroup differences were quantified by plotting the posterior histogram of the difference between the rightmost and leftmost nodes of the *rpart* tree. The posterior mass above zero indicates that there are significant differences between the two subgroups.

The subgroup analysis suggests that the effect is higher for children attending schools, in which instruction is very little or very much affected by mathematics resource shortages, schools with a more disadvantaged student body, and schools in which free lunch is provided either for all or for no students (Figure 5). Teacher experience and actual use of computer for schoolwork are also important moderating variables.

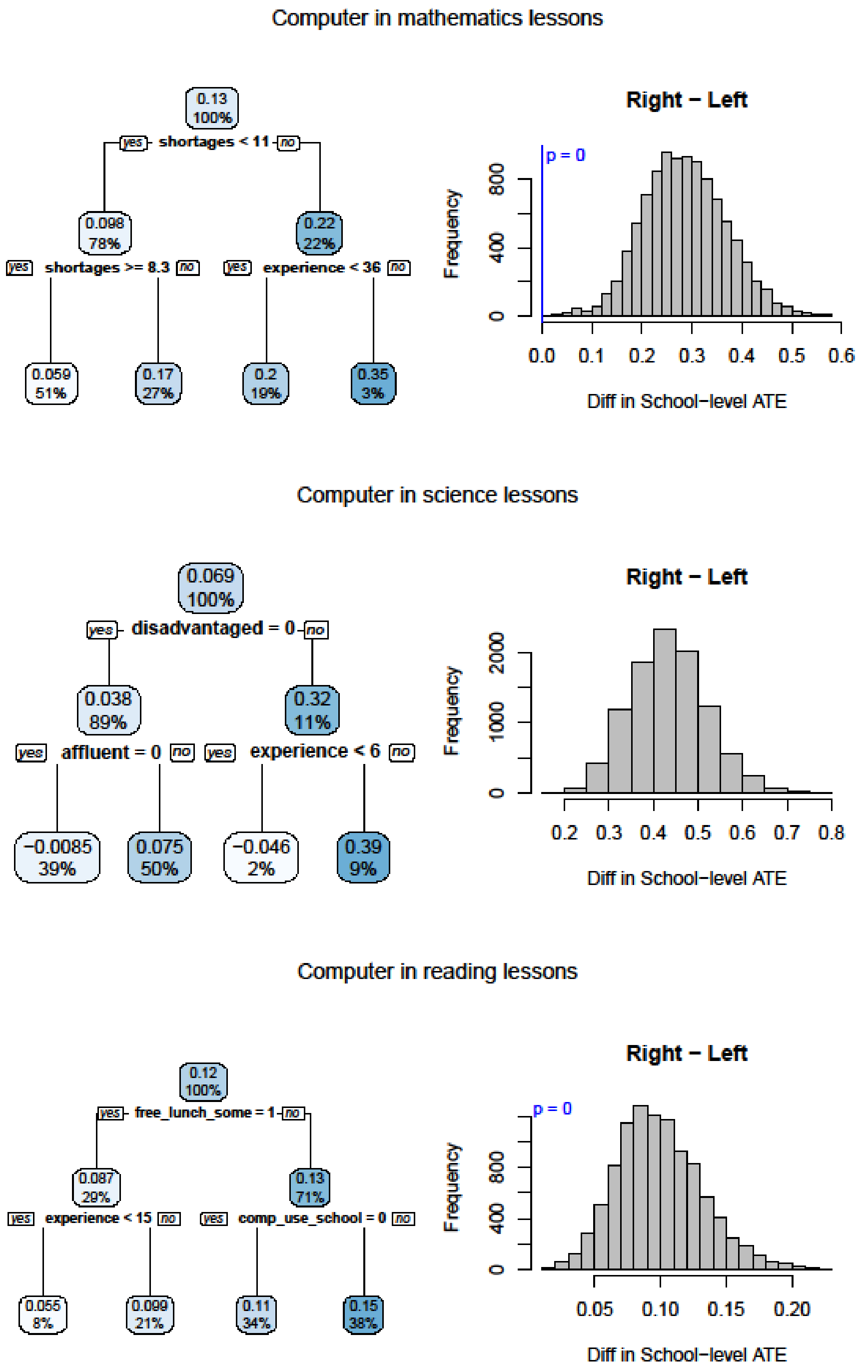
Mathematics resource shortages scale was flagged as the most important moderating variable of the treatment effect of computer availability in mathematics lessons. This scale is based on principals’ reports about shortages of teachers with a specialization in mathematics, shortages of software, applications and calculators for mathematics instruction or library resources relevant to mathematics instruction. A higher score indicates that mathematics instruction is less affected by resources shortages. Schools “affected a lot” by resource shortages score below 6.7, schools “not affected” by resource shortages score above 11.3.

It should be noted that the treatment effect conditional on shortages is not linear and is higher for children attending schools both affected and unaffected by shortages and lower for children attending schools only somewhat affected by shortages (see also Appendix 7.1). Teacher experience is another important moderating variable. The impact of computer availability in mathematics lessons equaled to 0.35 standard deviations for children attending schools, which scored above 11 on the mathematics resource shortages scale and were taught by teachers with more than or equal to 36 years of teaching experience. In contrast, the treatment effect equaled to 0.06 standard deviations for children attending schools only somewhat affected by resource shortages. There was also a significant difference between children with or without a preschool experience (see

[2] For more information regarding Bayesian “fit-the-fit” posterior summarization strategies, see Carnegie et al. (2019), Dorie (2020: 13), Dorie et al. Hahn and Carvalho (2015), Sivaganesan et al. (2017) and Starling et al. (2020).

Appendix 7.1). The treatment effect equaled to 0.28 s.d. for children who attended the preschool and 0.12 s.d. for those who did not.

Figure 5 Subgroup analysis



Note: Left panel: A summarizing regression tree fit to posterior point estimates of individual treatment effects. Each node contains the average subgroup treatment effect in that partition of the population and the percentage of observations contained in that terminal node. Right panel: Distribution of differences in subgroup average treatment effect for the leftmost and rightmost rpart nodes of the tree in the left panel.

In case of science lessons, the *rpart* model identified disadvantaged student body as the main cut off point. The treatment effect equaled to 0.32 standard deviations for those children who attended school with a more disadvantaged student body and to 0.04 standard deviations for those who attended schools with less disadvantaged student body. Again, teacher experience was flagged as an important moderator. The treatment effect equaled to 0.39 standard deviations for children who attended schools with a disadvantaged student body and were taught by a teacher with more than or equal to 6 years of teaching experience. In contrast, the effect was negative (-0.01 standard deviations) for children, who attended a school with an affluent student body.

Finally, school provision of free meals for some students was flagged as the most important moderating variable of the treatment effect of computer availability for reading lessons. The impact of computer availability for reading lessons equaled to 0.13 standard deviations for children who did not attend these schools and 0.09 standard deviations for those who did. Furthermore, the impact was slightly higher for children who not only did not attend these schools but also used computer for schoolwork (including classroom tasks, homework, or studying outside of class) at school (0.15 s.d.). In contrast, treatment effect equaled to 0.06 standard deviations for those children who attended these schools and were taught by a teacher with less than 15 years of teaching experience. Further heterogeneity analysis suggests that it was children attending schools, in which all children were eligible for free school lunch, who benefited the most from computer availability (see Appendix 7.3). The treatment effect equaled to 0.09 s.d. for children attending schools, which provided free meal to some students, 0.13 s.d. points for children attending schools, which did not provide free meal to any students and 0.25 points for children attending schools, which provided free meal to all students. The treatment effect for children attending schools, which provided free meal to all students, was significantly higher using all five plausible values.

When using plausible value 5 as the dependent variable, language spoken at home was identified as the most important moderator of the treatment effect of computer availability in reading lessons (see Appendix 6). The treatment effect was the highest for children who never spoke the language of test at home (0.17 s.d.) and the lowest for children who always spoke the language of test at home (0.11 s.d.). The effect was further moderated by teacher experience. Children who always spoke the language of test at home and were taught by less experienced teachers benefited the least from computer availability (0.09 s.d.). In contrast, the treatment effect equaled to 0.16 s.d. for children who did not always speak the language of test at home and were taught by teachers with more than (or equal to) 15 years of experience.

Teacher experience was identified as the most important moderator when using plausible value 2. The treatment effect equaled to 0.17 s.d. for children taught by teachers with more than (or equal to) 17 years of teaching experience. In contrast, it equaled only to 0.05 s.d. for children taught by teachers with less than 17 years of experience. Teaching experience was flagged as an important moderator when using all plausible values.

It is important to note that one should be cautious to read too much into the evidence of subgroup differences when the average effect is zero, as is the case of treatment effect of computer availability in mathematics and science lessons. However, differences in treatment effects of computer availability for reading lessons across subgroups suggest that it is the most disadvantaged children who benefit the most from computer availability during lessons.

7 CONCLUSIONS

This paper examined the impact of computer availability in mathematics, science and reading lessons on performance of Slovak fourth graders, using the data from the 2019 round of eTIMSS and the 2016 round of PIRLS testing. The impact of computer availability in mathematics and science lessons is statistically uncertain. However, computer availability in reading lessons had a positive impact: it improved the overall reading score by 0.10 to 0.13 standard deviations. Based on the fingings of recent meta-analyses of education interventions, this can be considered a medium effect size (Kraft 2020).

These findings are in line with the existing literature, which typically reports stronger effect of receiving an intervention than of offering an intervention (Kraft 2020). Studies examining the impact of computer availability (i.e. offering an intervention) tend to report no effect (e.g. Angrist and Lavy 2002; Leuven et al. 2007; Cristia et al. 2014; but Machin et al. 2006 documents a positive impact). In contrast, computer use (i.e. receiving an intervention) was documented to have a positive effect on student performance in some instances (e.g. Silvernail and Gritter 2007 or Crook et al. 2014). To get an idea about the effect size, Silvernail and Gritter (2007) document that in the United States, laptop use improved writing performance by one third of a standard deviation; Crook et al. (2015) document that computer use in Australia improved physics scores by 0.38 standard deviations, biology scores by 0.26 standard deviations and chemistry scores by 0.23 standard deviations.

The subgroup analysis suggests that the effect of computer availability in reading lessons was moderated by the socioeconomic composition of the school. Test scores improved by 0.09 s.d. for children attending schools, which provided free meal to some students, 0.13 s.d. points for children attending schools, which did not provide free meal to any students and 0.25 points for children attending schools, which provided free meal to all students (based on the plausible value 1; similar results were obtained for the remaining 4 plausible values). The effect size greater than 0.20 can be considered large (Kraft 2020). The treatment effect was higher also for children who never spoke the language of test at home (based on plausible value 5). This implies that computer availability has a potential to reduce educational inequalities (Lai et al. 2012; 2013; 2015), which affect student performance (Willms & Somers, 2001; Woessmann, 2004; OECD, 2016, Jæger and Breen 2016).

Treatment effect was higher also for children taught by teachers with more teaching experience. However, it is not clear how to interpret this effect. On the one hand, more experienced teachers could be better at integrating the technology in classroom. On the other hand, the impact of computer use on student performance is positive only if it substitutes less effective resources or teaching practices (Bulman and Fairlie 2016; Comi et al. 2017; Falck et al. 2018). Thus, it is also possible that computers substitute ineffective teaching practice. Further research is needed to uncover the link between teacher experience and computer use in classroom.

This study has several limitations stemming from the nature of the data. First, TIMSS and PIRLS datasets provide information on computer availability in classroom but do not provide information on how exactly the computers are used. Furthermore, the dataset does not include information on whether children had access to distracting content on the internet during lessons. Finally, the data is cross-sectional, i.e. it is impossible to track children over time. Further research is needed to identify types of educational software or computer-based teaching practices, which would deliver on the promise of digital transformation of Slovak schools.

References

Aaronson, D., Barrow, L. and Sander, W. (2007). 'Teachers and Student Achievement in the Chicago Public High Schools', *Journal of Labor Economics*, 25(1), pp.95–135. doi: 10.1086/508733.

Abadie, A. and Imbens, G.W. (2006). 'Large Sample Properties of Matching Estimators for Average Treatment Effects', *Econometrica,* 74(1), pp.235-292. doi: 10.1111/j.1468-0262.2006.00655.x.

Angrist, J.D., and Lavy, V. (1999). 'Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement', *Quarterly Journal of Economics,* 114(2), pp.533–75. doi: 10.1162/003355399556061.

Angrist, J.D. and Lavy, V. (2002). 'New Evidence on Classroom Computers and Pupil Learning', *Economic Journal* 112(482), pp.735–765. doi: 10.1111/1468-0297.00068.

Banerjee, A.V., Cole, S., Duflo, E. and Linden, L. (2007). 'Remedying Education: Evidence from Two Randomized Experiments in India', *Quarterly Journal of Economics,* 122(3), pp.1235-1264. doi: 10.1162/qjec.122.3.1235.

Barrera-Osorio, F., and Linden, L.L. (2009). *The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia*, World Bank Policy Research Working Paper 4836, Impact Evaluation Series No. 29. Available at: http://nrs.harvard.edu/urn-3:HUL.InstRepos:8140109 (Accessed: 26 September 2022).

Barrow, L., Markman, L. and Rouse, C.E. (2009). 'Technology's Edge: The Educational Benefits of Computer-Aided Instruction', *American Economic Journal: Economic Policy,* 1(1), pp.52-74. doi: 10.2139/ssrn.1083781.

Battistin, E. and Meroni, E.C. (2013). 'The Effects of Instruction Time on Low Achieving Students: Evidence From Southern Italy', *Economics of Education Review*, 55, pp.39-56. doi: 10.1016/j.econedurev.2016.08.003.

Bauernschuster, S., Falck, O., and Woessmann, L. (2014). 'Surfingalone? The internet and social capital: evidence from an unforeseeable technological mistake', *Journal of Public Economics*, 117, pp.73–89. doi: 10.1016/j.jpubeco.2014.05.007.

Belo, R., Ferreira, P. and Telang, R. (2014). 'Broadband in School: Impact on Student Performance', *Management Science,* 60(2), pp.265-282. doi: 10.1287/mnsc.2013.1770.

Bénabou, R., Kramarz, F. and Prost, C. (2009). 'The French zones d'éducation prioritaire: Much ado about nothing?', *Economics of Education Review,* 28(3), pp.345–356. doi: 10.1016/j.econedurev.2008.04.005.

Black, S.E., Devereux, P.J., Vellesen Løken, K. and Salvanes, K.G. (2014). 'Care or Cash? The Effect of Child Care Subsidies on Student Performance', *Review of Economics and Statistics,* 96(5), pp.824–37. doi: 10.2139/ssrn.2056731.

Bon, J.J. (2019). 'Using the tidytreatment package with BART.' https://cran.r-project.org/web/packages/tidytreatment/vignettes/use-tidytreatment-BART.html

Breiman, L. (2001). 'Random forests', *Machine learning,* 45(1), pp.5–32. doi: 10.1023/A:1010933404324.

Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Glynn, R.J., Avorn, J. and Stürmer, T. (2006). 'Variable selection for propensity score models', *American Journal of Epidemiology*, 163(12), pp.1149-1156. doi: 10.1093/aje/kwj149.

Buchel, O., Gajdošová, K. and Veselková, M. (2022). *Who takes care of the children? Under-reporting participation in formal childcare in EU-SILC*. Ministry of Labour, Social Affairs and Family of the Slovak Republic. Unpublished.

Bulman, G. and Fairlie, R. (2016). 'Chapter 5 - Technology and Education: Computers, Software, and the Internet', *Handbook of the Economics of Education. Volume 5*. Edited by Eric A. Hanushek, Stephen Machin and Ludger Woessmann, pp.239-280. doi: 10.1016/B978-0-444-63459-7.00005-1.

Burger, K. (2010). 'How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds', *Early Childhood Research Quarterly,* 25(2), pp.140–165. doi: 10.1016/j.ecresq.2009.11.001.

Cabras, S. and Tena Horrillo, J.D. (2015). *A Bayesian model to estimate causality in PISA scores: a tutorial with application to ICT*, UC3M Working Paper, Statistics and Econometrics, Universidad Carlos III de Madrid, Departamento de Estadística. Available at: http://hdl.handle.net/10016/21456 (Accessed: 26 September 2022).

Cabras, S. and Tena Horrillo, J.D. (2016). 'A Bayesian non-parametric modeling to estimate student response to ICT investment', *Journal of Applied Statistics*, 43(14), p2627-2642. doi: 10.1080/02664763.2016.1142946.

Cabras, S., Fidrmuc, J. and Tena Horrillo, J.D. (2017). 'Minimum Wage and Employment: Escaping the Parametric Straitjacket', *Economics: The Open-Access, Open-Assessment E-Journal,* 11, pp.1–20 [online]. Available at: http://dx.doi.org/10.5018/economics-ejournal.ja.2017-15 (Accessed: 26 September 2022).

Campbell, F., Conti, G., Heckman. J.J., Moon, S.H., Pinto, R., Pungello, E. and Pan, Y. (2014). 'Early childhood investments substantially boost adult health', *Science,* 343(6178), pp.1478-1485. doi: 10.1126/science.1248429.

Campuzano, L., Dynarski, M., Agodini, R. and Rall, K. (2009). *Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts – Executive Summary*. NCEE 2009-4041. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Available at: https://ies.ed.gov/ncee/pubs/20094041/pdf/20094042.pdf (Accessed: 26 September 2022).

Carnegie, N.B., Harada, M. and Hill, J. (2017). *treatSens: A Package to Assess Sensitivity of Causal Analyses to Unmeasured Confounding*. R package version 2.1.2. Available at: https://CRAN.R-project.org/package=treatSens (Accessed: 26 September 2022).

Carnegie, N.B., Dorie, V. and Hill, J. (2019). 'Examining treatment effect heterogeneity using BART', *Observational Studies,* 5(2), pp.52-70. doi: 10.1353/obs.2019.0002.

Cellini, S., Ferreira, F. and Rothstein, J. (2010). 'The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design', *Quarterly Journal of Economics,* 125(1), pp.215-261. doi: 10.1162/qjec.2010.125.1.215.

Chetty, R., Friedman J.N., Hilger, N., Saez, E., Schanzenbach, D.W. and Yagan, D. (2011). 'How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star', *The Quarterly Journal of Economics,* 126(4), pp.1593–1660. doi: 10.1093/qje/qjr041.

Chetty, R., Friedman, J.N. and Rockoff, J.E. (2014a). 'Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates', *American Economic Review,* 104(9), pp.2593–2632. doi: 10.1257/aer.104.9.2593.

Chetty, R., Friedman, J.N. and Rockoff, J.E. (2014b). 'Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood', *American Economic Review,* 104(9), pp.2633–2679. doi: 10.1257/aer.104.9.2633.

Chipman, H.A., George, E.I. and Mculloch, R.E. (2010). 'BART: Bayesian Additive Regression Trees', *Annals of Applied Statistics,* 4(1), pp.266–298. doi: 10.1214/09-AOAS285.
Chipman, H.A. and McCulloch, R.E. (2016). *_BayesTree: Bayesian Additive Regression Trees_*. R package version 0.3-1.4. Available at: https://CRAN.R-project.org/package=BayesTree.

Comi, S. L., Argentin, G., Gui, M., Origo, F. and Pagani, L. (2017). 'Is it the way they use it? Teachers, ICT and student achievement', *Economics of Education Review*, 56, pp.24-39. doi: 10.1016/j.econedurev.2016.11.007.

Cornelissen, T., Dustmann, C., Raute, A. and Schönberg, U. (2018). 'Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance', *Journal of Political Economy* 126(6), pp.2356–2409. doi: 10.1086/699979.

Crook, S. J., Sharma, M. D., and Wilson, R. (2015). 'An evaluation of the impact of 1: 1 laptops on student attainment in senior high school sciences', *International Journal of Science Education*, 37(2), pp.272-293. doi: 10.1080/09500693.2014.982229.

Cunha, F. and Heckman, J. (2007). 'The technology of skill formation', *American Economic Review,* 97(2), pp.31-47. doi: https://doi.org/10.1257/aer.97.2.31.

Cunha, F. and Heckman, J. (2009). 'The economics and psychology of inequality and human development', *Journal of the European Economic Association* 7(2–3), pp.320–364. doi: 10.1162/JEEA.2009.7.2-3.320.

Diamond, A. and Sekhon, J.S. (2013). 'Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies', *Review of Economics and Statistics,* 95(3), pp.932-946. doi: 10.1162/REST_a_00318.

Dorie, V. and Hill, J. (2020). *Package 'bartCause'.* R package version 1.0-4 Available at: https://cran.r-project.org/web/packages/bartCause/bartCause.pdf (Accessed: 26 September 2022).

Dorie, V., Harada, M., Carnegie, N.B. and Hill, J. (2016). 'A flexible, interpretable framework for assessing sensitivity to unmeasured confounding', *Statistics in Medicine,* 35(20), pp.3453-3470. doi: 10.1002/sim.6973.

Dorie, V., Hill, J., Shalit, U., Scott, M. and Cervone D. (2019). 'Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition', *Statistical Science*, 34(1), pp.43–68. doi: 10.1214/18-STS667.

Drange, N. and Havnes, T. (2019). 'Early childcare and cognitive development: Evidence from an assignment lottery', *Journal of Labor Economics*, 37(2), pp.581-620. doi: 10.1086/700193.

Dunst, C.J., Hamby, D.W., Wilkie, H. and Scott Dunst, K. (2017). 'Meta-analysis of the relationship between home and family experiences and young children's early numeracy learning', in Phillipson, S., Gervasoni, A., Sullivan, P. (ed.) *Engaging Families as Children's First Mathematics Educators. Early*

Mathematics Learning and Development, Singapore: Springer, pp.105-125. doi: 10.1007/978-981-10-2553-2_7.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R. et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort.* Washington, D.C.: U.S. Department of Education, Institute of Education Sciences. Available at: https://files.eric.ed.gov/fulltext/ED496015.pdf (Accessed: 26 September 2022).

Elango, S., García, J.L., Heckman, J.J. and Hojman, A. (2015a). *Early Childhood Education*. Cambridge: NBER Working Paper No. 21766. Available at: https://www.nber.org/papers/w21766 (Accessed: 26 September 2022).

Elango, S., García, J.L., Heckman, J.J., Hojman, A., Ermini, D., Rados, M.J., Shea, J. and Torcasso, J.C. (2015b). *The internal rate of return and the benefit-cost ratio of the Carolina Abecedarian Project.* University of Chicago, Department of Economics. Unpublished.

Ersan, O. and Rodriguez, M.C. (2020). 'Socioeconomic status and beyond: A multilevel analysis of TIMSS mathematics achievement given student and school context in Turkey', *Large-scale Assessments in Education*, *8*(1), pp.1-32. doi: 10.1186/s40536-020-00093-y.

Fairlie, R. W. and Kalil, A. (2017). 'The effects of computers on children's social development and school participation: Evidence from a randomized control experiment', *Economics of Education Review*, 57, pp.10-19. doi: 10.1016/j.econedurev.2017.01.001.

Falck, O., Mang, C. and Woessmann, L. (2018). 'Virtually no effect? Different uses of classroom computers and their effect on student achievement', *Oxford Bulletin of Economics and Statistics*, 80(1), pp.1-38. doi: 10.1111/obes.12192.

Falzetti, P., Palmerio, L. and Vidno, D. (2012). 'Reducing the gap? The effects of EU funds for improving education in Southern Italy', *Improving skills: Evidence from secondary analysis of international surveys conference,* 15-16 November 2012. Limassol, Cyprus.

Felfe, C., Nollenberger, N. and Rodríguez-Planas, N. (2015). 'Can't Buy Mommy's Love? Universal Childcare and Children's Long-Term Cognitive Development', *Journal of Population Economics*, 28(2), pp.393-422. doi: 10.1007/s00148-014-0532-x.

Felfe, C. and Lalive, R. (2018). 'Does early child care affect children's development?' *Journal of Public Economics*, 159, pp.33-53. doi: 10.1016/j.jpubeco.2018.01.014.

Ferraro, S. (2018). *Is information and communication technology satisfying educational needs at school?* MPRA Paper No. 86175. Available at: https://mpra.ub.uni-muenchen.de/86175/ (Accessed: 26 September 2022).

Fitzpatrick, M.D. (2008). 'Starting School at Four: The Effect of Universal Pre-kindergarten on Children's Academic Achievement', B.E. Journal of Economic Analysis and Policy, 8(1), pp.1-40. doi: 10.2202/1935-1682.1897.

Fryer, R.G., and Levitt, S.D. (2010). 'An empirical analysis of the gender gap in mathematics', *American Economic Journal: Applied Economics*, 2(2), pp.210-240. doi: 10.1257/app.2.2.210.

Ganzeboom, H.B.G. (2010). *Do's and Don't's of Occupation Coding with an Extension on ISCO-08*. Working Paper of the Department of Social Research Methodology, VU Amsterdam. Available at: http://www.harryganzeboom.nl/Pdf/2010%20-%20DO-and-DONTS-Occupation-coding-(paper-version4).pdf (Accessed: 26 September 2022).

Gelman, A. (2006). 'Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)', *Bayesian Analysis*, 1(3), pp.515–534. doi: 10.1214/06-BA117A.

Gerritsen, S. and Plug, E. (2017). 'Teacher Quality and Student Achievement: Evidence from a Sample of Dutch Twins', *Journal of Applied Econometrics,* 32(3), pp.643-660. doi: 10.1002/jae.2539.

Gibbons, S., McNally, S. and Viarengo, M. (2012). *Does additional spending help urban schools? An evaluation using boundary discontinuities*. IZA discussion paper No. 6281. Bonn: Institute for the Study of Labor. Available at: https://docs.iza.org/dp6281.pdf (Accessed: 26 September 2022).

Gormley, W.T., Jr., and Gayer, T. (2005). 'Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program', *Journal of Human Resources,* 40 (3), pp.533–58. doi: 10.3368/jhr.XL.3.533.

Gormley, W. T., Jr., Gayer, T., Phillips, D. and Dawson, B. (2005). 'The Effects of Universal Pre-K on Cognitive Development', *Developmental Psychology,* 41(6), pp.872– 84. doi: 10.1037/0012-1649.41.6.872.

Grauzelová, T. and Markovič, F. (2020). *Príjmy a životné podmienky v marginalizovaných rómskych komunitách. Vybrané ukazovatele zo zisťovania EU SILC MRK 2018*. Bratislava: Úrad splnomocnenca vlády SR pre rómske komunity. Available at: https://www.minv.sk/swift_data/source/romovia/EU%20SILC_MRK_2018_FINAL.pdf (Accessed: 26 September 2022).

Green, D. P. and Kern, H. L. (2012). 'Modeling heterogeneous treatment effects in survey experiments with Bayesian additive regression trees', *Public opinion quarterly*, 76(3), pp.491-511. doi: 10.2307/41684581.

Grønmo, L.S., Lindquist, M., Arora, A. and Mullis, I.V.S. (2013). 'Chapter 1. TIMSS 2015 Mathematics Framework', in Mullis, I.V.S. and Martin, M.O. (ed.) *TIMSS 2015 Assessment Frameworks.* Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, pp.11-27. Available at: https://timssandpirls.bc.edu/timss2015/downloads/T15_FW_Chap1.pdf (Accessed: 26 September 2022).

González de San Román, A. and de La Rica, S. (2016). 'Gender Gaps in PISA Test Scores: The Impact of Social Norms and the Mother's Transmission of Role Attitudes /Brechas de género en los resultados de PISA: El impacto de las normas sociales y la transmisión de roles de género de madres a hijas', *Estudios de Economía Aplicada*, 34(1), pp.79-108.

Goolsbee, A. and Guryan, J. (2006). 'The Impact of Internet Subsidies in Public Schools', *The Review of Economics and Statistics,* 88(2), pp.336-347. doi: 10.1162/rest.88.2.336.

Grimes, D. and Warschauer, M. (2008). 'Learning with laptops: a multi-method case study', *Journal of Educational Computing Research,* 38(3), p305-332. doi: 10.2190/EC.38.3.d.

Guiso, L., Monte, F., Sapienza, P. and Zingales, L. (2008). 'Culture, Gender, and Math', *Science*, 320(5880), pp.1164–1165. doi: 10.1126/science.1154094.

Gustafsson, J.E., Hansen, K.Y. and Rosén, M. (2013). 'Chapter 4. Effects of home background on student achievement in reading, mathematics, and science at the fourth grade', in Mullis, I.V.S. and Martin, M.O. (ed.) *Timss and pirls 2011: Relationships Among Reading, Mathematics, and Science Achievement at the Fourth Grade—Implications for Early Learning*, Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, pp.181-287. Available at: https://timssandpirls.bc.edu/timsspirls2011/downloads/TP11_Chapter_4.pdf (Accessed: 28 September 2022).

Hahn, P.R. and Carvalho, C.M. (2015). 'Decoupling shrinkage and selection in Bayesian linear models: a posterior summary perspective', *Journal of the American Statistical Association*, 110(509), pp.435–448. doi: https://doi.org/10.1080/01621459.2014.993077.

Hahn, P.R., Carvalho, C.M., Puelz, D. and He, J. (2018). 'Regularization and confounding in linear regression for treatment effect estimation', *Bayesian Analysis*, 13(1), pp.163-182. doi: 10.1214/16-BA1044.

Hahn, P.R., Murray, J.S. and Carvalho, C.M. (2020) 'Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects (with Discussion)', *Bayesian Analysis*, 15(3), pp.965-1056. doi: 10.1214/19-BA1195.

Hahn, P.R., Dorie, V. and Murray, J.S. (2019). 'Data Analysis Challenge 2017', *Atlantic Causal Inference Conference (ACIC)*. McGill University, 24 May. Montreal. Available from https://arxiv.org/pdf/1905.09515.pdf (Accessed: 26 September 2022).

Hanushek, E.A. (1999). 'Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects', *Educational Evaluation and Policy Analysis,* 21(2), pp.143-163. doi: 10.3102/01623737021002143.

Hanushek, E.A. and Woessmann, L. (2006). 'Does Early Tracking Affect Educational Inequality and Performance? Differences-in-Differences Evidence across Countries', *Economic Journal,* 116(510), pp.C63–C76. doi: 10.1111/j.1468-0297.2006.01076.x.

Hanushek, E.A. (2010). 'The difference is teacher quality', in Weber, K. (ed.) *Waiting for "Superman": How we can save America's failing public schools.* New York, NY: Public Affairs, pp.81–100.

Hanushek E.A. (2011). 'The economic value of higher teacher quality', *Economics of Education Review,* 30(2), pp.466–479. doi: 10.1016/j.econedurev.2010.12.006.

Hanushek, E.A. (2020). 'Chapter 13. Education production functions', in Bradley, S. and Green, C. (ed.) *The Economics of Education*. A Comprehensive Overview. Second Edition. Cambridge: Academic Press, pp.161-170.

Havnes, T. and Mogstad, M. (2011). 'No child left behind: Subsidized child care and children's long-run outcomes', *American Economic Journal: Economic Policy*, 3(2), pp.97-129. doi: 10.1257/pol.3.2.97.

Havnes, T. and Mogstad, M. (2015). 'Is universal child care leveling the playing field?', *Journal of public economics*, 127, pp.100-114. doi: 10.1016/j.jpubeco.2014.04.007.

Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P. and Yavitz, A. (2010). 'Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program', *Quantitative Economics,* 1 (1), pp.1–46. doi: 10.3982/QE8.

Heckman, J.J., Pinto, R. and Savelyev, P.A. (2013). 'Understanding the mechanisms through which an influential early childhood program boosted adult outcomes', *American Economic Review,* 103 (6), pp.2052–2086. doi: 10.1257/aer.103.6.2052.

Hellebrandt, T., Bednárik, M., Salomonsová, L. and Markovič, F. (2019). *Groups at Risk of Poverty or Social Exclusion Spending Review. Final Report*. Bratislava: Ministry of Finance of the Slovak Republic. Available at: https://www.mfsr.sk/files/archiv/56/ReviziavydavkovnaohrozeneskupinyZSverziaFINALENG.pdf (Accessed: 26 September 2022).

Higgins, S., Xiao, Z. and Katsipataki, M. (2012). *The impact of digital technology on learning: A summary for the education endowment foundation*. Durham: Education Endowment Foundation and Durham University. Available at: https://larrycuban.files.wordpress.com/2013/12/the_impact_of_digital_technologies_on_learning_full_report_2012.pdf (Accessed: 26 September 2022).

Hill, J. (2011). 'Bayesian Nonparametric Modeling for Causal Inference', *Journal of Computational and Graphical Statistics,* 20(1), pp.217–240. doi: 10.1198/jcgs.2010.08162.

Hill, J., Linero, A. and Murray, J. (2020). 'Bayesian Additive Regression Trees: A Review and Look Forward', *Annual Review of Statistics and Its Application*, 7(1), pp.251-278. doi: 10.1146/annurev-statistics-031219-041110.

Hill, J. and Yu-Sung Su (2013). 'Assessing lack of common support in causal inference using Bayesian nonparametrics: Implications for evaluating the effect of breastfeeding on children's cognitive outcomes,' *The Annals of Applied Statistics*, 7(3), pp.1386-1420. doi: 10.1214/13-AOAS630.

Hill, J. L., Weiss, C., and Zhai, F. (2011). 'Challenges with propensity score strategies in a high-dimensional setting and a potential alternative', *Multivariate Behavioral Research*, 46(3), pp.477-513. doi: 10.1080/00273171.2011.570161.

Hustedt, J.T., Barnett, W.S., Jung, K., and Figueras, A. (2008). *Impacts of New Mexico Pre-K on Children's School Readiness at Kindergarten Entry: Results from the Second Year of a Growing Initiative*. New Jersey: National Institute for Early Education Research, Rutgers University. Available at: https://nieer.org/wp-content/uploads/2008/06/NewMexicoRDD0608.pdf (Accessed: 26 September 2022).

Hoxby, C.M. (2000). 'The effects of class size on student achievement: New evidence from population variation', *The Quarterly Journal of Economics,* 115(4), pp.1239–1285. doi: 10.1162/003355300555060.

International Association for the Evaluation of Educational Achievement (IEA) (2015). *TIMSS 2015 International Database*. [online]. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and IEA. Available from: https://timssandpirls.bc.edu/timss2015/international-database/ (Accessed: 26 September 2022).

Jacob, B., Berger, D., Hart C. and Loeb, S. (2016). 'Can Technology Help Promote Equality of Educational Opportunities?' *RSF: The Russell Sage Foundation Journal of the Social Sciences,* 2(5), pp.242–271. doi: 10.7758/rsf.2016.2.5.12.

Jæger, M.M. and Breen, R. (2016). 'A dynamic model of cultural reproduction', *American Journal of Sociology,* 121(4), pp.1079-1115. doi: 10.1086/684012.

Kirkpatrick, H. and Cuban, L. (1998). 'Computers Make Kids Smarter — Right?', *TECHNOS,* 7(2), pp.26-31.

Kleemans, T., Peeters, M., Segers, E. and Verhoeven, L. (2012). 'Child and home predictors of early numeracy skills in kindergarten', *Early Childhood Research Quarterly*, 27(3), pp.471-477. doi: 10.1016/j.ecresq.2011.12.004.

Knudson, C.P. and Vats, D. (2021). *stableGR: A Stable Gelman-Rubin Diagnostic for Markov Chain Monte Carl.* R package version 1.0. Available at: https://cran.r-project.org/web/packages/stableGR/index.html (Accessed: 26 September 2022).

Koedinger, K.R., Anderson, J.R., Hadley, W.H., and Mark, M.A. (1997). 'Intelligent Tutoring Goes To School in the Big City', *International Journal of Artificial Intelligence in Education*, 8, pp.30-43.

Kraft, M.A., (2020). Interpreting effect sizes of education interventions, *Educational Researcher*, 49(4), pp.241-253. Available at: https://doi.org/10.3102/0013189X20912798

Krueger, A.B. (1999). 'Experimental Estimates of Education Production Functions', *Quarterly Journal of Economics,* 114, pp.497–532.

Kruschke, J.K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan (2.ed.).* Amsterdam: Elsevier, Academic Press.

Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, Y., Boswell, M. and Rozelle, S. (2012). *Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China*. Stanford, CA: Rural Education Action Program (REAP) Working Paper No. 237. Available at: https://sccei.fsi.stanford.edu/reap/publications/does_computerassisted_learning_improve_learning_outcomes_evidence_from_a_randomized_experiment_in_public_schools_in_rural_minority_areas_in_qinghai_china (Accessed: 26 September 2022).

Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Boswell, M. and Rozelle, S. (2013). 'Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in Shaanxi', *Journal of Development Effectiveness*, 5(2), pp.208-231. doi: 10.1080/19439342.2013.780089.

Lai, F., Luo, R., Zhang, L., Huang, X. and Rozelle, S. (2015). 'Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing', *Economics of Education Review*, 47, pp.34-48. doi: 10.1016/j.econedurev.2015.03.005.

LeFevre, J.A., Skwarchuk, S.L., Smith-Chant, B.L., Fast, L., Kamawar, D. and Bisanz, J. (2009). 'Home numeracy experiences and children's math performance in the early school years', *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 41(2), pp.55-66. doi: 10.1037/a0014532.

LeFevre, J.A., Polyzoi, E., Skwarchuk, S.L., Fast, L. and Sowinski, C. (2010). 'Do home numeracy and literacy practices of Greek and Canadian parents predict the numeracy skills of kindergarten children?', *International Journal of Early Years Education*, 18(1), pp.55-70. doi: 10.1080/09669761003693926.

Leonti, M., Cabras, S., Weckerle, C.S., Solinas, M.N. and Laura Casu (2010). 'The Causal Dependence of Present Plant Knowledge on Herbals – Contemporary Medicinal Plant Use in Campania (Italy) Compared to matthioli (1568)', *Journal of Ethnopharmacology,* 130 (2), pp.379–391. doi: 10.1016/j.jep.2010.05.021.

Leuven, E., Lindahl, M., Oosterbeek, H. and Webbink, D. (2007). 'The Effect of Extra Funding for Disadvantaged Pupils on Achievement', *Review of Economics and Statistics,* 89(4), pp.721-736. doi: 10.1162/rest.89.4.721.

Leuven, E., Oosterbeek, H. and Rønning, M. (2008). 'Quasi-experimental estimates of the effect of class size on achievement in Norway', *Scandinavian Journal of Economics,* 110(4), pp.663–693. doi: 10.1111/j.1467-9442.2008.00556.x.

Leuven, E. and Løkken, S.A. (2017). *Long Term Impacts of Class Size in Compulsory School*. IZA Discussion Papers no. 10594. Bonn: Institute for the Study of Labor. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931331 (Accessed: 26 September 2022).

Machin, S., McNally, S. and Silva, O. (2007). 'New Technology in Schools: Is There a Payoff?', *Economic Journal,* 117(522), pp.1145-1167. doi: 10.1111/j.1468-0297.2007.02070.x.

Machin, S., McNally, S. and Costas Meghir (2010). 'Resources and standards in urban schools', *Journal of Human Capital,* 4(4), pp.365–393. doi: 10.1086/658634.

Makowski, D., Lüdecke, D., Ben-Shachar, M.S., Wilson, M.D., Bürkner, P., Mahr, T., Singmann, H. and Gronau, Q.F. (2020). *Package 'bayestestR'*. R package version 0.13.0. Available at: https://cran.r-project.org/web/packages/bayestestR/bayestestR.pdf (Accessed: 26 September 2022).

Malamud, O. and Pop-Eleches, C. (2011). 'Home Computer Use and the Development of Human Capital', *Quarterly Journal of Economics,* 126(2), pp.987-1027. doi: 10.1093/qje/qjr008.

Martorell, P., Stange, K.M. and McFarlin, I. (2015). *Investing in Schools: Capital Spending, Facility Conditions, and Student Achievement.* Cambridge: NBER Working Paper No. 21515. Available at: https://www.nber.org/system/files/working_papers/w21515/revisions/w21515.rev0.pdf (Accessed: 26 September 2022).

McConnell, K. J. and Lindner, S. (2019). 'Estimating treatment effects with machine learning', *Health services research*, 54(6), pp.1273-1282. doi: 10.1111/1475-6773.13212.

Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., Qiao, Y., Boswell, M. et al. (2014). 'Integrating computer-assisted learning into a regular curriculum: Evidence from a randomised experiment in rural schools in Shaanxi', *Journal of development effectiveness*, 6(3), pp.300-323. doi: 10.1080/19439342.2014.911770.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L. and Fishbein, B. (2020). 'TIMSS 2019 International Results in Mathematics and Science'. Available at: https://timssandpirls.bc.edu/timss2019/international-results/ (Accessed: 26 September 2022).

Neyman, J. (1923). 'On the application of probability theory to agricultural experiments. Essay on principles. Section 9', *Statistical Science,* 5, pp.465–472.

Nollenberger, N., Rodríguez-Planas, N. and Sevilla, A. (2016). 'The Math Gender Gap: The Role of Culture', *American Economic Review,* 106 (5), pp.257-61. doi: 10.1257/aer.p20161121.

Nye, B., Hedges, L.V. and Konstantopoulos, S. (1999). 'The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment', *Educational Evaluation and Policy Analysis,* 21(2), pp.127–142. doi: 10.2307/1164296.

Nye, B., Hedges, L.V. and Konstantopoulos, S. (2000). 'The effects of small classes on academic achievement: The results of the Tennessee class size experiment', *American Educational Research Journal,* 37(1), pp.123–151. doi: 10.3102/00028312037001123.

Nye, B., Konstantopoulos, S. and Hedges, L.V. (2004). 'How large are teacher effects?', *Educational Evaluation and Policy Analysis,* 26(3), pp.237–257. doi: 10.3102/01623737026003237.

OECD (2010). *Mathematics Teaching and Learning Strategies in PISA*. Paris: OECD Publishing. Available at: https://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/mathematicsteachingandlearningstrategiesinpisa.htm (Accessed: 26 September 2022).

OECD (2012). 'Does money buy strong performance in PISA?', *PISA IN FOCUS 13*. Available at: https://doi.org/10.1787/5k9fhmfzc4xx-en (Accessed: 26 September 2022).

OECD (2013). *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV).* Paris: OECD Publishing. Available at: https://doi.org/10.1787/9789264201156-en (Accessed: 26 September 2022).

OECD (2015). *Students, Computers and Learning: Making the Connection*. Paris: OECD Publishing. Available at:
http://dx.doi.org/10.1787/9789264239555-en (Accessed: 26 September 2022).

OECD (2016a). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*. Paris: OECD Publishing. Available at: http://dx.doi.org/10.1787/9789264255425-en (Accessed: 26 September 2022).

OECD (2016b). *PISA 2015 Results (Volume I): PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing. Available at: http://dx.doi.org/10.1787/9789264266490-en (Accessed: 26 September 2022).

OECD (2016c). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. Paris: OECD Publishing. Available at: http://dx.doi.org/10.1787/9789264267510-en (Accessed: 26 September 2022).

OECD (2019). *Education at a Glance 2019: OECD Indicators*. Paris: OECD Publishing. Available at: https://doi.org/10.1787/f8d7880d-en (Accessed: 26 September 2022).

Orlando, J. (2014). 'Veteran teachers and technology: change fatigue and knowledge insecurity influence practice', *Teachers and Teaching*, 20(4), pp.427-439. doi: 10.1080/13540602.2014.881644.

Papanastasiou, E., Zembylas, M., and Vrasidas, C. (2003). 'When computer use is associated with negative science achievement', *Journal of Science Education and Technology*, 12(3), pp.325-332.

Papke, L. (2005). 'The effects of spending on test pass rates: evidence from Michigan', *Journal of Public Economics* 89(5), pp. 821–839. doi: 10.1016/j.jpubeco.2004.05.008.

Pfeffer, F.T. (2008). 'Persistent Inequality in Educational Attainment and its Institutional Context', *European Sociological Review* 24(5), pp.543–65. doi: 10.1093/esr/jcn026.

Pierce, T. (2009). 'Social anxiety and technology: face-to-face communication versus technological communication among teens', *Computers in Human Behavior*, 25(6), pp. 1367–1372. doi: 10.1016/j.chb.2009.06.003.

Polson, N. G. and Scott, J. G. (2012). 'On the half-Cauchy prior for a global scale parameter', *Bayesian Analysis*, 7(4), pp.887–902. doi: 10.1214/12-BA730.

R Core Team (2022). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/ (Accessed: 26 September 2022).

Rivkin, S.G., Hanushek, E.A. and Kain, J.F. (2005). 'Teachers, schools, and academic achievement', *Econometrica* 73(2), pp.417–458. doi: 10.3982/ECTA12211.

Rockoff, J. E. (2004). 'The impact of individual teachers on student achievement: Evidence from panel data', *The American Economic Review,* 94(2), pp.247–252. doi: 10.1257/0002828041302244.

Rouse, C.E., and Krueger, A.B. (2004). 'Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program', *Economics of Education Review,* 23(4), pp.323–338. doi: 10.1016/j.econedurev.2003.10.005.

Rubin, D.B. (1974). 'Estimating causal effects of treatments in randomized and nonrandomized studies', *Journal od Educational Psychology*, 66(5), pp.688–701.

Savina, E., Mills, J.L., Atwood, K., and Cha, J. (2017). 'Digital media and youth: A primer for school psychologists', *Contemporary School Psychology*, 21(1), pp.80-91. doi: 10.1007/s40688-017-0119-0.

Schuetz, G., Ursprung, H.W. and Woessmann, L. (2008). 'Education Policy and Equality of Opportunity', *Kyklos*, 61(2), pp.279–308. doi: 10.2139/ssrn.876444.

Sekhon, J.S. (2011). 'Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R', *Journal of Statistical Software*, 42(7), pp. 1-52. doi: 10.18637/jss.v042.i07.

Shapley, K., Sheehan, D., Maloney, C. and Caranikas-Walker, F. (2011). 'Effects of Technology Immersion on Middle School Students' Learning Opportunities and Achievement', *The Journal of Educational Research,* 104(5), pp.299-315. doi: 10.1080/00220671003767615.

Silvernail, D. L. and Gritter, A. K. (2007). *Maine's middle school laptop program: Creating better writers*. Gorham. [online]. ME: Maine Education Policy Research Institute. Available at: https://www.maine.gov/mlti/resources/Impact_on_Student_Writing_Brief.pdf (Accessed: 26 September 2022).

Sivaganesan, S., Müller, P., and Huang, B. (2017). 'Subgroup finding via Bayesian additive regression trees', *Statistics in Medicine*, 36(1), pp. 2391-2403. doi: https://doi.org/10.1002/sim.7276.

Skwarchuk, S.L., Sowinski, C. and LeFevre, J.A. (2014). 'Formal and informal home learning activities in relation to children's early numeracy and literacy skills: The development of a home numeracy model', *Journal of experimental child psychology*, 121(1), pp.63-84. doi: 10.1016/j.jecp.2013.11.006.

Staiger, D.O. and Rockoff, J.E. (2010). 'Searching for effective teachers with imperfect information', *Journal of Economic Perspectives,* 24(3), pp.97–117. doi: 10.1257/jep.24.3.97.

Starling, J.E., Murray, J.S., Lohr, P.A., Aiken, A.R.A., Carvalho, C.M. and Scott, J.G. (2020). *Targeted Smooth Bayesian Causal Forests: An analysis of heterogeneous treatment effects for simultaneous versus interval medical abortion regimens over gestation*. doi: 10.48550/arXiv.1905.09405.

Suhr, K., Hernandez, D., Grimes, D. and Warschauer, M. (2010). 'Laptops and fourth-grade literacy: assisting the jump over the fourth-grade slump', *The Journal of Technology, Learning and Assessment*, 9(5), pp.1–45.

Therneau, T.M., Atkinson, E.J. and Ripley, B. (2019a). *Package 'rpart'*, R package version 4.1-15. Available at: https://cran.r-project.org/web/packages/rpart/rpart.pdf (Accessed: 26 September 2022).

Therneau, T.M., Atkinson, E.J. and Mayo Foundation (2019b). *An Introduction to Recursive Partitioning Using the RPART Routines.* [online]. Available at: https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf (Accessed: 26 September 2022).

Uhls, Y. T., Michikyan, M., Morris, J., Garcia, D., Small, G. W., Zgourou, E. and Greenfield, P. M. (2014). 'Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion cues', *Computers in Human Behavior*, 39, pp.387-392. doi: 10.1016/j.chb.2014.05.036.

Valencia, K., Rusu, C., Quiñones, D. and Jamet, E. (2019). 'The impact of technology on people with autism spectrum disorder: a systematic literature review', *Sensors*, 19(20), pp.4485. doi: 10.3390/s19204485.

Van der Klaauw, W. (2008). 'Breaking the link between poverty and low student achievement: An evaluation of title I', *Journal of Econometrics,* 142(2), pp.731–756. doi: 10.1016/j.jeconom.2007.05.007.

Vats, D. and Knudson, C. (2018). 'Revisiting the Gelman-Rubin diagnostic'. To be published in *Statisical Science* [Preprint]. Available at: https://arxiv.org/abs/1812.09384 (Accessed: 26 September 2022).

Wendling, T., Jung, K., Callahan, A., Schuler, A., Shah, N. and Gallego, B. (2018). 'Comparing methods for estimation of heterogeneous treatment effects using observational data from health care databases', *Statistics in medicine*, 37(23), pp.3309–3324. doi: 10.1002/sim.7820.

Willms, J.D. and Somers, M. (2001). 'Family, Classrooms, and School Effects on Children's Educational Outcomes in Latin America', *School Effectiveness and School Improvement,* 12(4), pp.409-445. doi: 10.1076/sesi.12.4.409.3445.

Woessmann, L. (2004). *How Equal Are Educational Opportunities? Family Background and Student Achievement in Europe and the United States*. IZA discussion paper No. 1284. Bonn: Institute for the Study of Labor. Available at: https://docs.iza.org/dp1284.pdf (Accessed: 26 September 2022).

Woessmann, L. (2009). 'International Evidence on School Tracking: A Review CESifo DICE Report', *Journal for Institutional Comparisons,* 7(1), pp.26–34.

Wong, V.C., Cook, T.D., Barnett, W.S. and Jung, K. (2008). 'An Effectiveness-Based Evaluation of Five State Pre-kindergarten Programs', *Journal of Policy Analysis and Management,* 27 (1), pp.122–54. doi: 10.1002/9781444307399.ch24.

Yeager, D.S., Hanselman, P., Walton, G.M., Murray, J.S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B.I. et al. (2019). 'A national experiment reveals where a growth mindset improves achievement', *Nature*, 573(7774), pp.364–369. doi: 10.1038/s41586-019-1466-y.

Appendix 1 Covariate Balance Tables

Table A1.1 Covariate balance table (computer availability in mathematics lessons)

	Diff.Un	M.Threshold.Un	Diff.Adj	M.Threshold.Un
resources	0.1796	**>0.05**	-0.0004	<0.05
preschool	0.0149	<0.05	0.0000	<0.05
experience	0.1468	**>0.05**	-0.0002	<0.05
affluent	-0.0635	**>0.05**	-0.0001	<0.05
average	0.0961	>0.05	0.0001	<0.05
disadvantaged	-0.0326	<0.05	-0.0000	<0.05
shortages	-0.2038	**>0.05**	-0.0004	<0.05
prop.score			0.0207	<0.05

Note: Diff.Un refers to the difference in means between the two groups prior to adjusting. The standardized mean difference is displayed for continuous variables and the raw mean difference (difference in proportion) is displayed for binary variables. Diff.Adj refers to the (standardized) difference in means between the two groups after adjusting. M.Threshold indicates whether or not the calculated mean difference after adjusting exceeds or is within the threshold given by threshold = c(m = .05)). Variables are imbalanced if the standardized mean differences (for continuous variables) and differences in proportion (for binary variables) are greater than .05.

Table A1.2 Covariate balance table (computer availability in science lessons)

	Diff.Un	M.Threshold.Un	Diff.Adj	M.Threshold.Un
lang_never	-0.0274	<0.05	0.0000	<0.05
resources	0.1513	**>0.05**	-0.0002	<0.05
preschool	0.0090	<0.05	0.0000	<0.05
experience	0.2177	**>0.05**	0.0007	<0.05
affluent	0.0742	**>0.05**	0.0002	<0.05
disadvantaged	-0.0639	**>0.05**	-0.0001	<0.05
shortages	-0.0028	<0.05	0.0003	<0.05
prop.score			-0.0075	<0.05

Note: see Table A1.1.

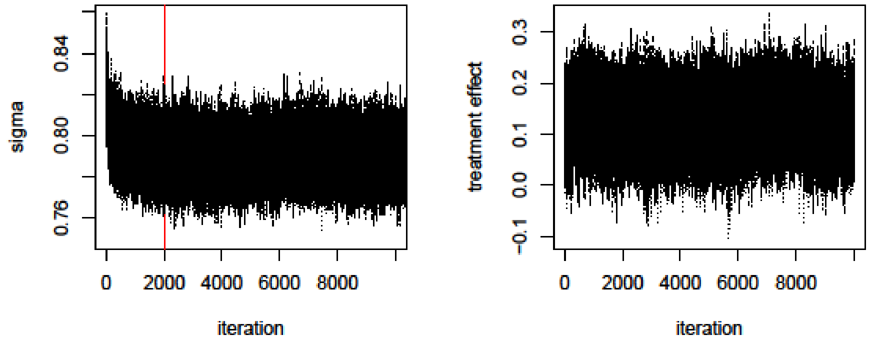
Table A1.3 Covariate balance table (computer availability in reading lessons)

	Diff.Un	M.Threshold.Un	Diff.Adj	M.Threshold.Un
preschool	-0.0050	<0.05	0.0001	<0.05
lang_sometimes	-0.0248	<0.05	-0.0002	<0.05
resources	0.1308	**>0.05**	0.0005	<0.05
experience	0.1268	**>0.05**	0.0004	<0.05
free_lunch_all	0.0028	<0.05	-0.0000	<0.05
shortages	0.3175	**>0.05**	0.0010	<0.05
comp_use_home	0.0274	<0.05	0.0001	<0.05
comp_use_other	0.0032	<0.05	0.0001	<0.05
prop.score			0.0437	<0.05

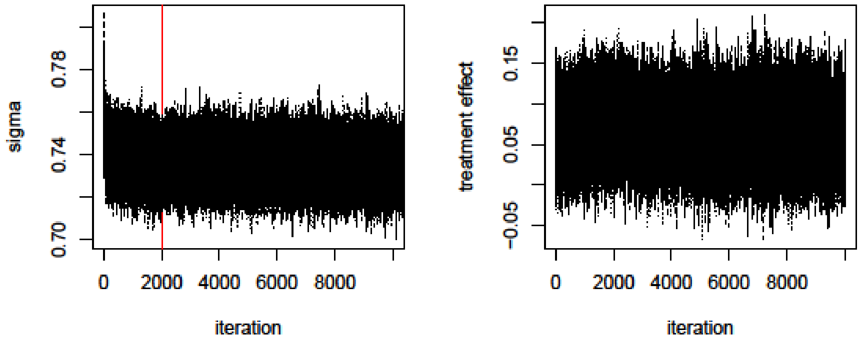
Note: see Table A1.1.

Appendix 2 MCMC convergence diagnostics

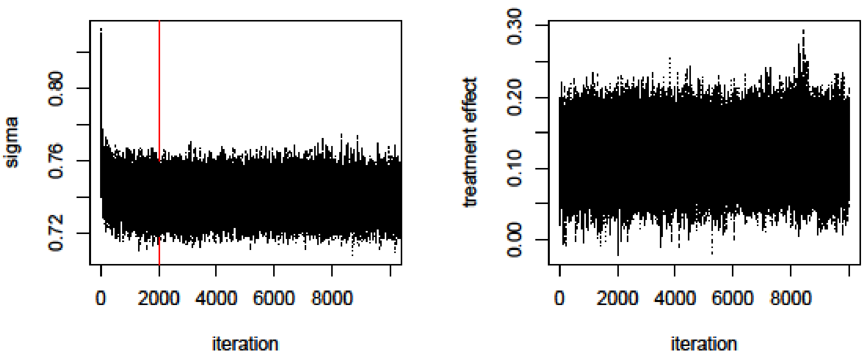
Figure A2.1 Convergence diagnostic



(a) mathematics



(b) science



(c) reading

Note: Left panel: Sigma by MCMC iteration. Samples to the left of the vertical red line are burn-in from 10 chains generated by MCMC algorithm. Samples to the right are the post-burn-in iterations. Right panel: Traceplot of the treatment effect by post-burn-in iteration. Chains appear to have mixed well and converged to a stationary distribution.

Appendix 3 Descriptive statistics

Table 3.1 Descriptive statistics (computer availability in mathematics lessons)

	Computer availability = 0 n = 3 653		Computer availability = 1 n = 515	
	mean	st.dev	mean	st.dev
Math Achievement Score	-0.02	1.00	0.16	0.96
Student characteristics				
female	0.50	0.50	0.48	0.50
preschool	0.95	0.21	0.97	0.17
resources	10.41	1.47	10.68	1.48
early_litnum	10.91	1.68	11.05	1.66
comp_use	0.60	0.49	0.67	0.47
lang_always	0.88	0.33	0.90	0.30
lang_almost_always	0.07	0.25	0.07	0.25
lang_sometimes	0.05	0.21	0.02	0.15
lang_never	0.00	0.07	0.01	0.09
Teacher characteristics				
teacher_experience	23.11	10.48	24.46	9.17
School characteristics				
shortages	9.46	1.52	9.14	1.58
affluent	0.50	0.50	0.44	0.50
average	0.39	0.49	0.48	0.50
disadvantaged	0.11	0.32	0.08	0.27

Table 3.2 Descriptive statistics (computer availability in science lessons)

	Computer availability = 0 n = 3 163		Computer availability = 1 n = 995	
	mean	st.dev	mean	st.dev
Science Achievement Score	-0.05	1.02	0.15	0.91
Student characteristics				
female	0.49	0.50	0.50	0.50
preschool	0.95	0.21	0.96	0.19
resources	10.39	1.49	10.60	1.42
early_litnum	10.91	1.67	10.98	1.70
comp_use	0.62	0.49	0.63	0.48
lang_always	0.69	0.46	0.71	0.46
lang_almost_always	0.17	0.38	0.19	0.40
lang_sometimes	0.11	0.31	0.09	0.29
lang_never	0.03	0.18	0.01	0.08
Teacher characteristics				
teacher_experience	22.37	11.49	24.32	8.93
School characteristics				
shortages	9.41	1.47	9.40	1.66
affluent	0.48	0.50	0.55	0.50
average	0.40	0.49	0.38	0.49
disadvantaged	0.12	0.33	0.06	0.24

Table 3.3 Descriptive statistics (computer availability in reading lessons)

	Computer availability = 0 n = 4 038		Computer availability = 1 n = 970	
	mean	st.dev	mean	st.dev
Reading Achievement Score	-0.03	1.01	0.14	0.95
Student characteristics				
female	0.49	0.50	0.49	0.50
preschool	0.97	0.18	0.96	0.19
resources	10.21	1.42	10.41	1.49
early_litnum	10.93	1.71	11.08	1.75
comp_use_home	0.82	0.39	0.85	0.36
comp_use_school	0.51	0.50	0.61	0.49
comp_use_other	0.43	0.49	0.43	0.50
lang_always	0.86	0.35	0.87	0.33
lang_almost_always	0.08	0.27	0.09	0.28
lang_sometimes	0.06	0.23	0.03	0.17
lang_never	0.01	0.08	0.01	0.09
Teacher characteristics				
teacher_experience	21.25	10.86	22.65	11.05
School characteristics				
shortages	10.24	1.10	10.61	1.16
free_lunch_all	0.01	0.10	0.01	0.12
free_lunch_some	0.30	0.46	0.23	0.42
free_lunch_none	0.69	0.46	0.75	0.43

Appendix 4 Data sources and variable construction

Table A4.1 Variable definition (TIMSS) - mathematics

	TIMSS Code	Description
Mathematics Achievement	ASMMAT01	First plausible value in math achievement (with PSI)
Treatment Status	ATBM04A	Computer/tablet availability during math (available = 1)
Student characteristics		
female	ASBG01	A binary variable for gender (female = 1)
preschool	ASDHAPS	A binary variable for preschool attendance (attended = 1)
resources	ASBGHRL	Home resources for learning
early_litnum	ASBHELN	Early literacy and numeracy activities before school
comp_use	ASBE02B	Used computer/tablet for mathematics schoolwork (used = 1)
lang_always	ASBH14	Child always speaks language of test at home.
lang_almost_always	ASBH14	Child almost always speaks language of test at home.
lang_sometimes	ASBH14	Child sometimes speaks language of test at home.
lang_never	ASBH14	Child never speaks language of test at home.
Teacher characteristics		
teacher_experience	ATBG01	Years of teaching
School characteristics		
shortages	ACBGMRS	Instruction affected by math resource shortages
affluent	ACDGSBC	School composition by socioec. background/More Affluent
average	ACDGSBC	School comp./neither more affluent nor more disadvantaged
disadvantaged	ACDGSBC	School composition/more disadvantaged

Source: IEA (2019)

Table A4.2 Variable definition (TIMSS) – science

	TIMSS Code	Description
Science Achievement	ASSSCI01	First plausible value in science achievement (with PSI)
Treatment Status	ATBS03A	Computer/tablet availability during science (available = 1)
Student characteristics		
female	ASBG01	A binary variable for gender (female = 1)
preschool	ASDHAPS	A binary variable for preschool attendance (attended = 1)
resources	ASBGHRL	Home resources for learning
early_litnum	ASBHELN	Early literacy and numeracy activities before school
comp_use	ASBE02C	Used computer/tablet for science schoolwork (used = 1)
lang_always	ASBH14	Child always speaks language of test at home.
lang_almost_always	ASBH14	Child almost always speaks language of test at home.
lang_sometimes	ASBH14	Child sometimes speaks language of test at home.
lang_never	ASBH14	Child never speaks language of test at home.
Teacher characteristics		
teacher_experience	ATBG01	Years of teaching
School characteristics		
shortages	ACBGSRS	Instruction affected by science shortages
affluent	ACDGSBC	School composition by socioec. background/More Affluent
average	ACDGSBC	School comp./neither more affluent nor more disadvantaged
disadvantaged	ACDGSBC	School composition/more disadvantaged

Source: IEA (2019)

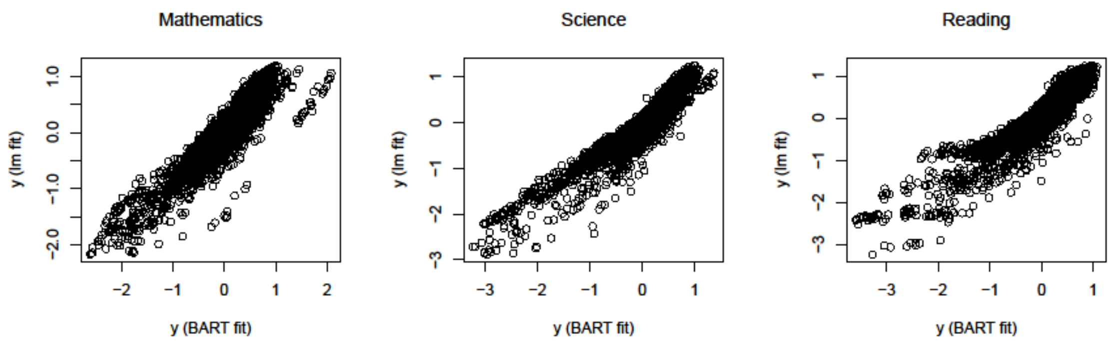
Table A4.3 Variable definition (PIRLS) - reading

	PIRLS Code	Description
Reading Achievement	ASRREA01	First plausible value in overall reading
Treatment Status	ATBR14A	Computer/tablet availability during reading (available = 1)
Student characteristics		
Female	ASBG01	A binary variable for gender (female = 1)
preschool	ASDHAPS	A binary variable for preschool attendance (attended = 1)
resources	ASBGHRL	Home Resources for Learning
early_litnum	ASBHELN	Early literacy and numeracy activities before school
comp_home	ASBG09A	Used computer/tablet for schoolwork/home (used = 1)
comp_school	ASBG09B	Used computer/tablet for schoolwork/school (used = 1)
comp_other	ASBG09C	Used computer/tablet for schoolwork/other (used = 1)
lang_always	ASBH17	Child always speaks language of test at home.
lang_almost_always	ASBH17	Child almost always speaks language of test at home.
lang_sometimes	ASBH17	Child sometimes speaks language of test at home.
lang_never	ASBH17	Child never speaks language of test at home.
Teacher characteristics		
teacher_experience	ATBG01	Years of teaching
School characteristics		
shortages	ACBGRRS	Instruction affected by reading resorce shortages
free_lunch_all	ACBG06B	School provided free meals for all students.
free_lunch_some	ACBG06B	School provided free meals for some students.
free_lunch_none	ACBG06B	School did not provide free meals for any of its students.

Source: IEA (2016).

Appendix 5 BART fit vs. linear regression fit

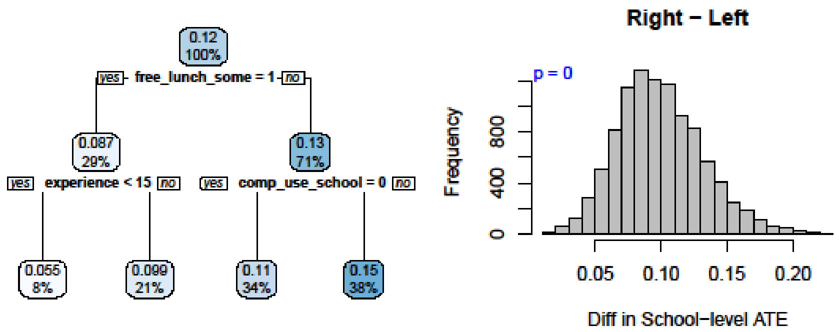
Figure 5.1 Comparison of BART fit and linear regression fit.



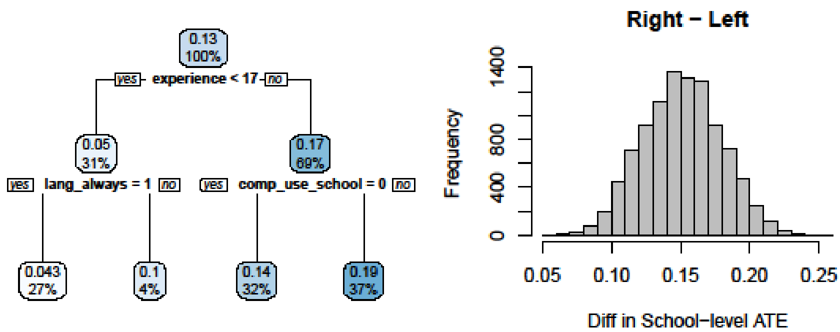
Note: For BART model, pseudo-R-squared equals to 0.40 for mathematics achievement, 0.48 for science achievement and 0.47 for reading achievement. For linear model, R-squared equals to 0.31 for mathematics achievement, 0.40 for science achievement and 0.38 for reading achievement.

Appendix 6 Subgroup analysis – overall reading score, plausible values 1 to 5.

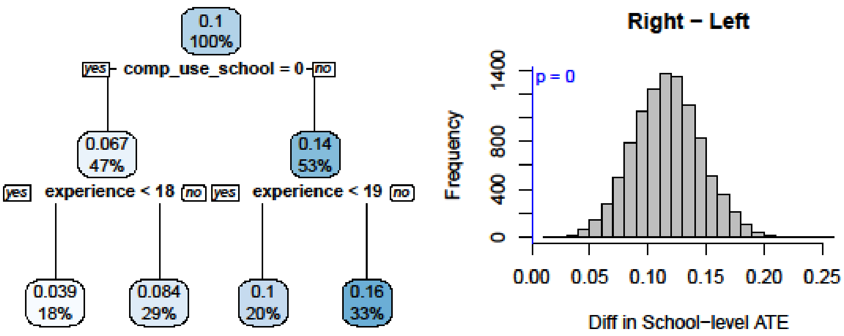
Computer in reading lessons



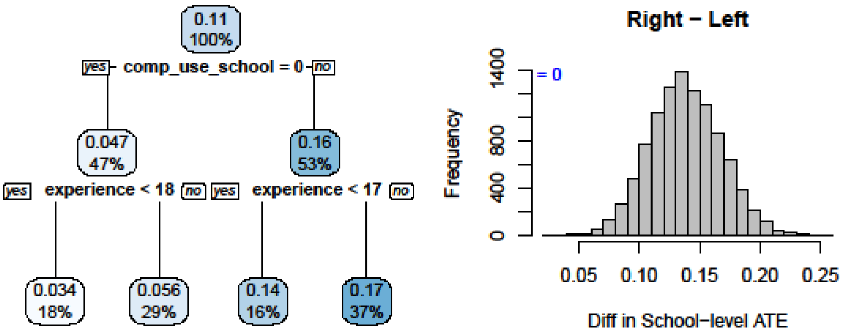
Computer in reading lessons



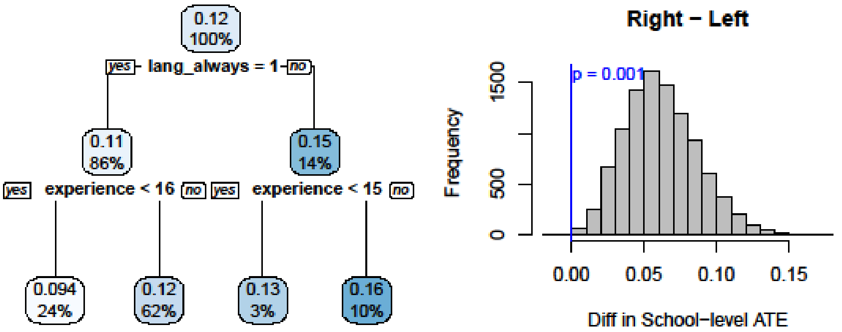
Computer in reading lessons



Computer in reading lessons

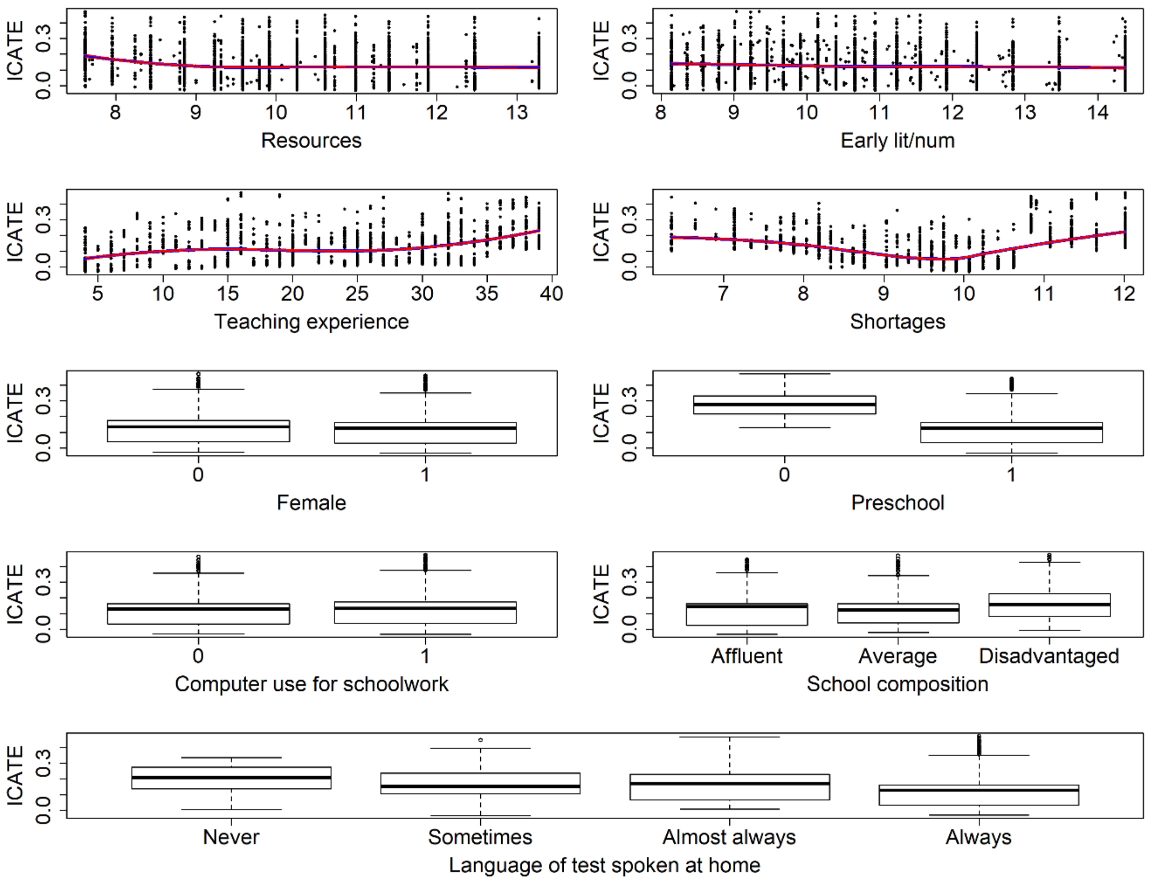


Computer in reading lessons



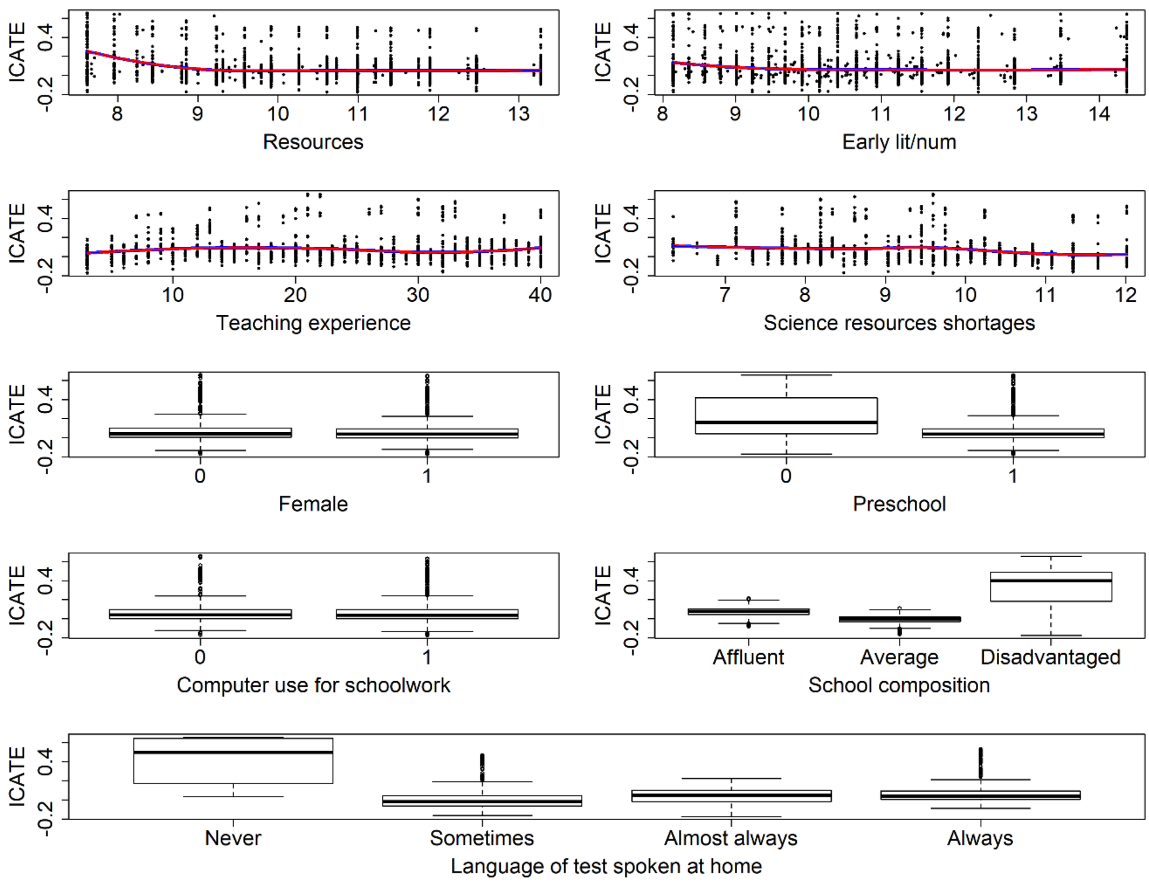
Appendix 7 Detailed heterogeneity analysis

Figure 7.1 Heterogeneity of treatment effects of computer availability in mathematics classes



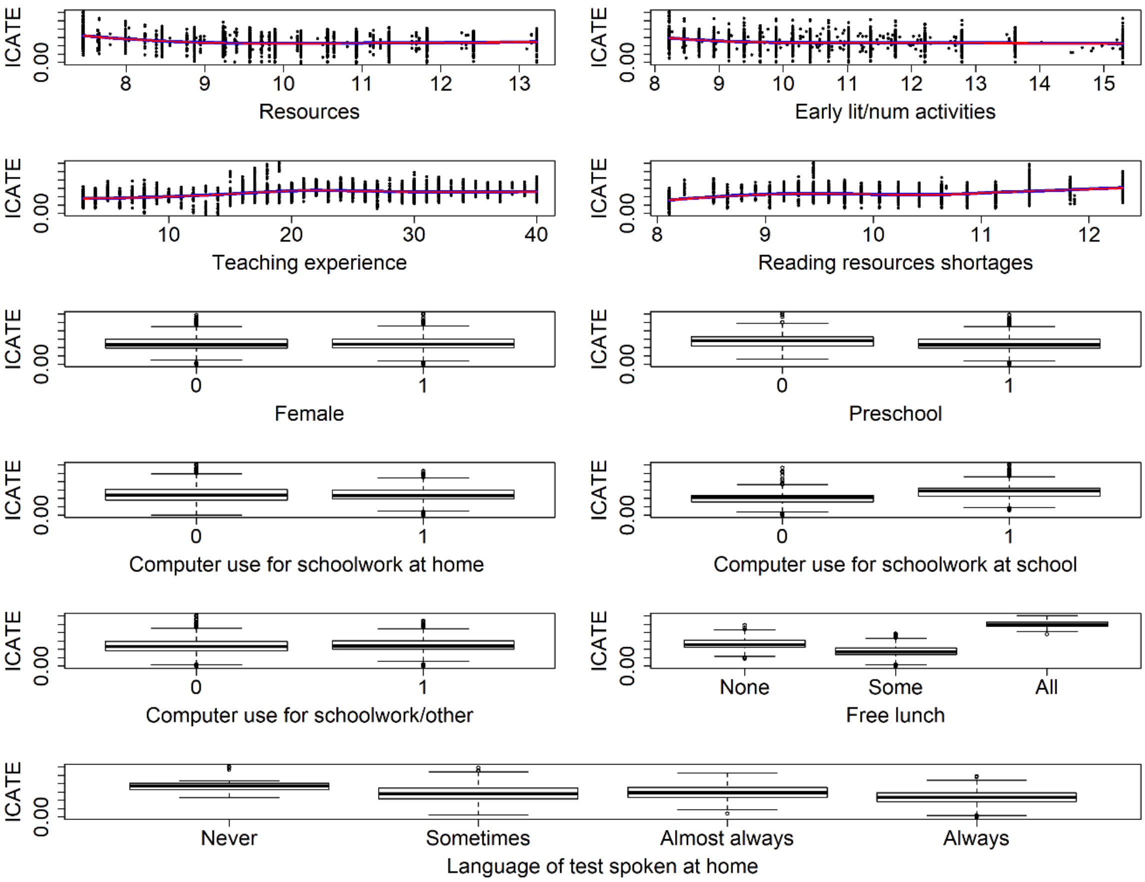
Note: Effect modification by continuous variables depicted as loess plots (blue line) with 95% confidence bands (red lines). Effect modification by categorical variables depicted as side-by-side boxplots.

Figure 7.2 Heterogeneity of treatment effects of computer availability in science classes.



Note: Effect modification by continuous variables depicted as loess plots (blue line) with 95% confidence bands (red lines). Effect modification by categorical variables depicted as side-by-side boxplots.

Figure 7.3 Heterogeneity of treatment effects of computer availability in reading classes.



Note: Effect modification by continuous variables depicted as loess plots (blue line) with 95% confidence bands (red lines). Effect modification by categorical variables depicted as side-by-side boxplots.