Academic Performance and E-learning in Italy *

Josep Amer-Mestre, Alaitz Ayarza Astigarraga
European University Institute

Marta C. Lopes
Universidad Carlos III de Madrid, IZA, GLO

Tuesday 20th July, 2021

Abstract

Using data from Google Trends in Italy, we study how different regions reacted to the implementation of online learning through national-wide E-learning platforms, due to the COVID-19 pandemic. Regions with a previously lower academic performance, actively searched more in Google about e-learning tools, surpassing regions with earlier school closures. Analysing school administrative data and students' survey data before the pandemic we find that both teachers and students in lower performing regions were already using other e-learning tools more than higher performing regions. Unlike studies in other countries, our findings suggest that the COVID-19 shock may not change the lower academic performance regions behaviour in their usage of e-learning. With the fresh release of the 2021 INVALSI results, we confirm that this is the case. We conclude that despite the larger usage of E-learning in Yellow Zones by regions that performed worse in 2018/2019, such differences did not have an impact in the grades of High School in 2021. Nevertheless, we do observe a non-negligible impact of how regions in Orange Zones used E-learning on the change between grades of 2021 and 2019. More specifically, regions that were in orange zones for a longer period and that used more E-learning during that period seemed to have, in part, mitigated the negative change in grades. Both results seem to indicate that, in case of another year of restrictions and e-learning, there should be less discretion, and more school specific policy measures to try to diminish the still evident differences between North and Southern Italy.

JEL classification: C81, I24
Keywords: E-learning, COVID-19, Google Trends, Inequality

*We thank Andrea Ichino, Thomas Crossley, Anna Lavizzari, Eleonora Erittu, Pedro Luís Silva, Pedro Freitas and Jan Stuhler for very helpful comments and suggestions. We also thank the very useful discussions at the internal seminars of European University Institute and Universidad Carlos III de Madrid, as well as comments and suggestions provided by anonymous referees. We gratefully acknowledge financial support from the European University Institute COVID-19 Knowledge Hub. Corresponding author: marta.cardoso@uc3m.es.

1 Introduction

School closures were the primary response by Governments in most affected countries by COVID-19 to control the virus' spread in Spring 2020 (ACAPS, 2020). Educational institutions were suddenly forced to undergo a quick transition from face-to-face to online teaching. Around the world, concerns were raised about potential differences in online learning usage by students with different socio-economic conditions. See Bacher-Hicks et al. (2021) and Rodríguez-Planas (2020) for analyses in the United States, Andrew et al. (2020) for England, and Engzell et al. (2020) for an evaluation in primary schools in the Netherlands.

In this paper, we look for differences in e-learning usage by academic performance. Students with better grades should get higher income returns in the future (Murnane et al., 1995). We want to test whether the sudden change in learning methods will be a contributor to potentially increase social mobility or, instead, contribute to the widening of the socio-economic inequalities.

As a case study, we focus on Italy. This country, albeit unexplored, is an interesting case for three main reasons. First, according to the COVID-19 Government Measures Dataset (ACAPS, 2020), Italy was the first country to close schools outside Asia (the epicentre of the virus outbreak). Second, Italy is a country that presents substantial regional differences in school quality (Brunello and Checchi, 2005) and academic performance (Agasisti and Vittadini, 2012). Third and finally, the centralised school management in the country put forward a website (*didattica a distanza*) established together with the school closures, to support schools in implementing online learning methods, which enable us to compare regions on the usage of nationally implemented e-learning platforms.

To measure the usage of e-learning platforms, we use real-time data on e-learning tools interest via Google Trends for each Italian region from January to June 2020 (the Spring term during which national closure of schools was implemented). To measure academic performance, we use pre-pandemic regional level of test scores in Italian and Math, performed by the National Institute for Education Evaluation and Training (INVALSI).

We find that regions that performed better before the pandemic searched relatively less for e-learning tools via Google during the first lockdown. While this finding could potentially suggest a catching-up effect from the regions with lower academic performance, we actually find that these regions were already using more e-learning before the pandemic, when we analyse older platform's usage. Exploring e-learning related questions in INVALSI and PISA (OECD Programme for International Student Assessment) at the regional level, we find that these regions were already using more Information and Technology (IT) both in class and outside school, thus suggesting that any regional differences in academic performance and e-learning usage should perpetuate in Italy. In fact when, we test for regional differences in online e-learning platforms that were popular before and after the pandemic (*Scuola.net* and *Studenti.it*) we find no statistical differences.

Our results, therefore, contrast with the findings for the United States, reported by Bacher-

Hicks et al. (2021). The authors find that areas of the country with higher income, better internet access and fewer rural schools saw substantially larger increases in search intensity. We find that the Italian regions with a comparative *disadvantage* in terms of academic performance *increased* the search intensity relatively more. In addition to the literature that studies regional academic inequalities during COVID-19 lockdowns, our paper also contributes to two other branches of the economic literature: Google Trends, and regional differences in Italy.

Most of this economic literature using Google Trends has been focusing on Macroeconomic and Financial indicators: see, for example, Vosen and Schmidt (2011) for consumption, Baker and Fradkin (2017) for unemployment insurance, Castelnuovo and Tran (2017) for uncertainty, Dergiades et al. (2015) for sovereign spreads and Hamid and Heiden (2015) for volatility in the stock market. This paper instead uses the data at the regional level, which has been recently more popular on the study of potential regional inequalities due to the COVID-19 pandemic (see Bacher-Hicks et al. (2021), for example). In this paper, we call attention for two important features of the data that are often disregarded: 1) the comparison of levels of the index between regions, and 2) the sampling feature of Google Trends. The later is particularly important in the case of smaller countries.

The economic literature that studies regional differences in Italy found evidence for differences in economic development (Peracchi, 2008), school quality (Brunello and Checchi, 2005), and academic performance (Agasisti and Vittadini, 2012), but no study, to the best of our knowledge, has documented the regional differences in the usage of technology at schools. This finding is particularly important as the economic literature is divided between none (Fairlie and Robinson, 2013; Beuermann et al., 2015; Cristia et al., 2017; Bando et al., 2017) and negative association (Brown and Liedholm, 2002; Joyce et al., 2015; Lu and Song, 2020) between e-learning and academic performance.

With the fresh release of the 2021 INVALSI results, we analyse the consequences of different number of school closures and different E-learning usage in academic performance. We conclude that, according to each colour a region was set, i.e., depending on the restrictions in terms of school closures and recommendations to use *Didattica a Distanza*, the regions used more E-learning in Orange than in Yellow Zones, and even more in Red Zones, especially when all compared with White Zones and with periods of School Holidays, between November 6 and June 18. We also conclude that despite the larger usage of E-learning in Yellow Zones by regions that performed worse in 2018/2019, such differences did not have an impact in the grades of High School in 2021. Nevertheless, we do observe a non-negligible impact of how regions in Orange Zones used E-learning on the change between grades of 2021 and 2019. More specifically, regions that were in orange zones for a longer period and that used more E-learning during that period seemed to have, in part, mitigated the negative change in grades.

Both results seem to indicate that, in case of another year of restrictions and e-learning, there should be less discretion, and more school specific policy measures to try to diminish the still

evident differences between North and Southern Italy. Therefore, our study also contributes to the public debate of how to mitigate regional differences in academic performances in Italy. As other countries in Europe also applied regional measures to combat the spread of COVID-19 in the subsequent waves during the academic year of 2020/2021, we believe our study should also serve as a good example on how discretionary policy measures might exacerbate regional differences in academic performance.

The paper is organised as follows. Section 2 provides a contextual background to how the Italian Ministry of Education implemented e-learning in the country during the first COVID-19 lockdown. Section 3 describes the main data used in this study. Section 4 explains the methodology that originates the results in Section 5. Section 6 provides additional empirical evidence on the differences in e-learning for two platforms that were widely used before the pandemic. Section 7 analyses the impact of the school closures and different E-learning behaviour on the academic performance results of 2021. Finally, section 8 concludes.

2 Contextual Background

Italy was the first European country to be hit by the COVID-19 in 2020. The first case of the virus in Italy was confirmed by January 31, but both intensity and speed of new cases were unequal across the country, thus leading to a highly regionalised impact, as reported by Giuliani et al. (2020). The first schools to close were the ones in the two most affected regions, Lombardy and Veneto (*zona rossa*), on February 23.

On March 4, Italy ordered the closure of *all* schools and universities. Five days later, on March 9, the president declared a national lockdown. On March 11, all commercial activity except for supermarkets and pharmacies were closed, and on March 21, the Italian government closed all non-essential businesses and industries and restricted movement of people.

The program *Didattica a Distanza* (distance learning, in English) was firstly announced on the radio by the Minister of Education, Lucia Azzolina, at the end of February. The details were not provided immediately but on March 4, e-learning became mandatory for all schools in Italy. Since that date and throughout the Spring term, the Ministry of Education website made available a new tab with dedicated training webinars and available information on different platforms that were constantly updated.

Eventually the website stabilised in three platforms: G Suite, provided by Google, (which includes Google Classroom and Google Meetings), Microsoft Office 365 (which includes Word, PowerPoint, Excel, Outlook and Teams), provided by Microsoft, and WeSchool, provided by the Italian main communication company. While all these platforms already existed before the pandemic, their usage was relatively scarce, but quickly became popular with the program of *Didattica a Distanza*, as we show in the next section.

3 Data

3.1 Google Trends

To measure the engagement with online learning platforms in each of the Italian regions during the first COVID-19 lockdown in Italy, we rely on Google Trends. Google Trends calculates the fraction of a given geographic area's Google searches devoted to a given term relative to the overall Google searches in that geographic area- otherwise, places with the most search volume would always be ranked highest-. Further, it scales the resulting number on a range of 0 to 100, assigning 100 to the point in time and geographic area with the highest fraction value. This scaling is done *within* the data set that Google Trends allows to download at once: 1) a single series per term, region and time window; 2) a maximum of 5 different series, corresponding to 5 different geographic areas over a common time window and keyword; or 3) a maximum of 5 different series, corresponding to 5 different keywords over a common time window and geographic area. This means that only those series that are scaled up together have values readily comparable -in levels- one to another. Given that only changes across regions and over time matter for identification in our empirical strategy , we follow approach 1) to build our main data set. That is, we download a single series per keyword j region r and time window T, where the index $I_{j,r,t}$, constructed by Google Trends, is the ratio between the popularity of term j relative to the maximum popularity of that term over the time period T in keyword j, measured on a 0 to 100 scale. It is calculated as follows:

$$I_{j,r,t} = 100 \frac{S_{j,r,t} / \sum_{i \in I} S_{i,r,t}}{\max_{t \in T} (S_{j,r,t} / \sum_{i \in I} S_{i,r,t})}$$

The numerator is measured as the ratio between the number of searches of term j in region r at point t ($S_{j,r,t}$) and the sum of searches for all terms $i \in I$ in that region and point in time ($\sum_{i \in I} S_{i,r,t}$). The denominator is the maximum of these ratios over the time period T for keyword j and region r. The index $I_{j,r,t}$ is the outcome variable of our regressions. Following Bacher-Hicks et al. (2021), we use the logarithm of Google Trends' index to interpret estimates as percent changes and assume that increased search intensity for a term corresponds to increased raw search volume, given evidence that overall Google search volumes did not change substantially during the pandemic.

When extracting Google Trends data, one should note two characteristics of Google Trends. First, Google Trends uses a representative sample -not the population- of all Google searches. This is important, in particular, when extracting data for small geographic areas with low search volume. Given that Italy has some small regions, we take this point seriously and download 20 different series for each keyword-region-time-window combination to make sure that our results are

not driven by any particular sample that Google Trends makes available at a particular moment[1]. Then, we average the results by term, region and point in time over all the 20 samples. Thus, the upper bound of our index is not necessarily 100 -can be smaller-. Second, Google Trends provides different frequency data, depending on the time span of the series that one wants to download. For series spanning 9 months or less, it makes daily frequency data available. For series spanning more than 9 months, it only makes weekly frequency data available. For our main empirical specification we select a (common) period for all regions spanning from September 4 2016 to June 7 2020 - corresponding to the end of the academic year- that allows to control for month and year fixed effects. Given that the selected period comprises more than 9 months, the series obtained has weekly frequency. In the analysis that we carry out in section 7 the series have daily frequency data.

To provide the reader with clear evidence on why the data set from Google Trends is well suited for our study, after explaining the selection of keywords in the next section we show descriptive evidence where the raw series, as downloaded from Trends, are comparable in levels.

3.1.1 Selection of Keywords

To avoid confounding teleworking and e-learning, a key point in our study is to choose platforms that are exclusively designed for e-learning. For example, while Google Drive can be used by teachers to upload study material, it is also a commonly used application by firms. Thus, its increase in popularity during the pandemic would be attributed to a compound effect of the increase in teleworking and e-learning that our data does not allow to disentangle. Taking this into consideration, we restricted our keyword list to 5 different platforms exclusively dedicated to e-learning: *Studenti.it*, *Scuola.net*, *Edmodo*, *Google Classroom* and *WeSchool*. Note that *(Google) Classroom* and *WeSchool* feature as the third and fourth most searched words in the 10 trending words' list of Italy during the year 2020, only after *Coronavirus* and *Elezione USA* (USA elections) words, which take the first and second places respectively.

Studenti.it is an Italian website for studying support, managed by the Italian schooling books publisher Mondalori Media S.p.A.. It is one of the most visited websites in Italy by high school students, university students and young people looking for training and employment. The website is constantly updated, and it provides students with the subjects of study of the current school year, study material to prepare for the exams (such as the middle school exam -Esame 3ª media, taken at the age of 13- and graduation exams - Maturità, at the age of 19), as well as plenty of practical information, including news from the Ministry of Education.

Scuola.net is a project of La Fabbrica. La Fabbrica is a training institution for teaching staff of the Italian school accredited by MIUR - Ministry of Education, University and Research. It is

[1]Google Trends renews the publicly available sample every given hour. Given that all the downloads were automated, we used slightly different time windows - couple of days apart- by region and keyword in each of the 20 downloads. This trick ensures that each downloaded series belongs, as we want, to a different sample.

a website dedicated to teachers of the various school grades. An space where they can participate in free educational initiatives and access solutions for digital teaching.

Edmodo, Google Classroom and WeSchool are 3 e-learning platforms that provide similar services, allowing teachers to set assignments, to have work submitted by students, to mark, to return graded papers, to distribute quizzes, surveys etc.

3.1.2 Descriptive Evidence on Google Trends' Series

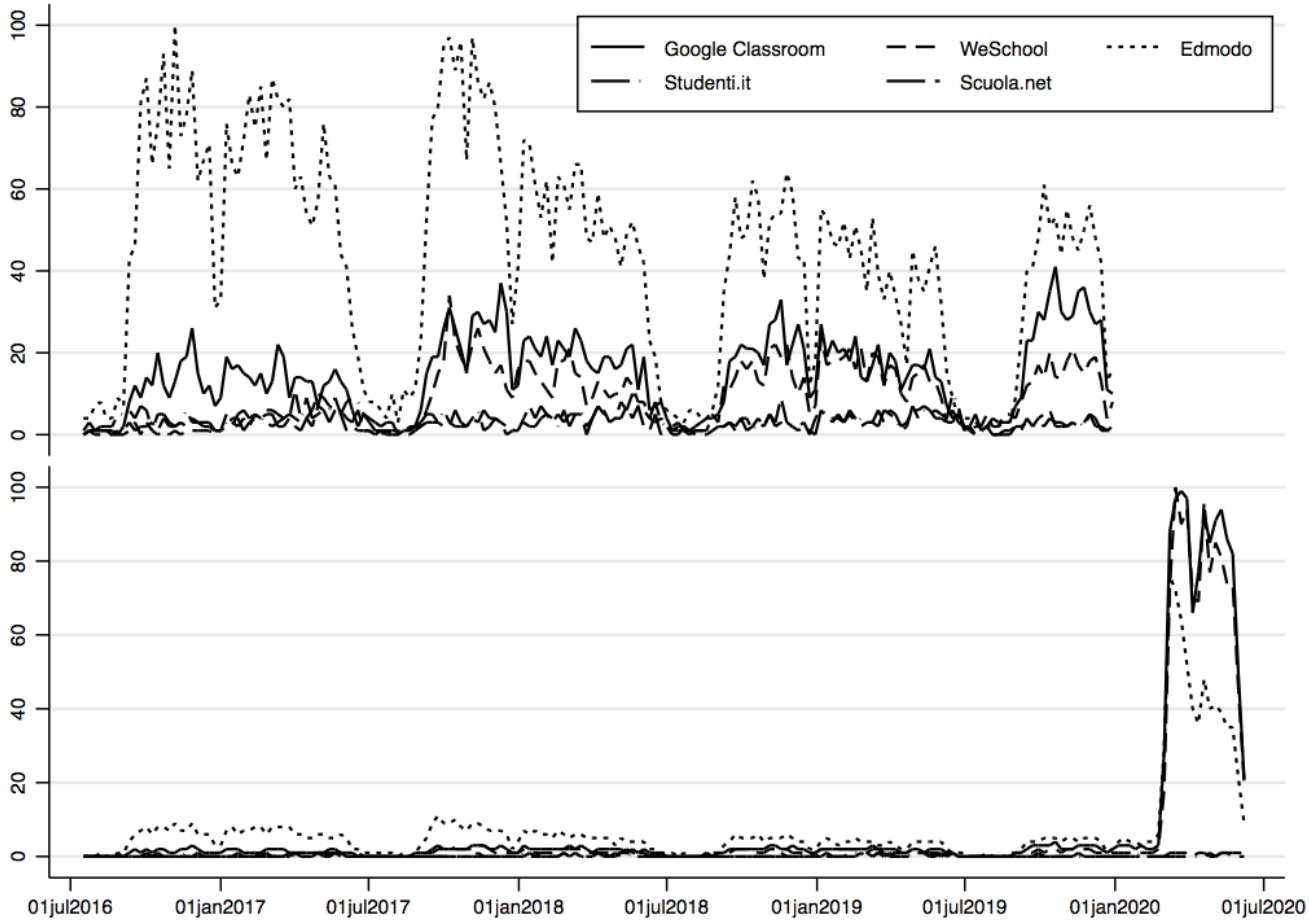
As explained above, in this section we show the raw data, as downloaded from Google Trends. We use the section for two purposes. First, to justify our choice about how to construct the data set. Second. to validate the quality of the data publicly available from Google Trends.

As it suffices for the purpose and to avoid having a large number of series that would make the graph fuzzy, we choose the country of Italy as the geographic area and show one series per keyword over the given time window. Because we have exactly 5 keywords, we download this data set *at once* following approach 3). As explained in the previous section, the values of the index $I_{j,r,t}$ are now comparable *across* - not only within- the series. This is because, now, Google Trends does the scaling up of $I_{j,r,t}$ as follows: assigns the value 100 to the date and keyword that reaches the maximum fraction value across all the 5 series, and then scales every point of each of the 5 series relative to this maximum point. Figure 1 below shows these 5 series. The top and bottom figures in Figure 1 correspond to two different downloads, that differ only the time window that was chosen for the download. The bottom figure downloads the series for both the before and after the pandemic period (from June 2016 to June 2020). The top figure downloads the series for before the pandemic period only, from June 2016 to December 2019.

The bottom figure 1 shows clear evidence that the search of e-learning platforms in Italy increased dramatically right after the Covid-19 outbreak. This increase was lead by Google Classroom, which reached the highest value across all the keyword series on the week between the 22-28th of March 2020, and thus gets the value 100 in the graph. That same week, WeSchool was searched 91% and Edmodo was searched 60% as much as Google Classroom. Studenti.it and Scuola.net show a constant search behavior over the entire time window *relative* to the other 3 platforms. On that same week, they were each searched 1% as much as Google Classroom. It is important to stress that this does not mean that these two platforms did not increase after the Covid-19 outbreak. It means that *if* they changed, they changed so much less compared to the other three, that the difference can not be noticed. Because all the values in the series are relative values, the extreme increases of the other 3 platforms - specially that of Google Classroom- push the values of *studenti.it* and *scuola.net* down[2]. This graph helps to see visually the nature

[2]Due to the different nature of these two search terms their different response to the Covid-19 outbreak does not come as a surprise*studenti.it* and *scuola.net* are websites where one can get informed and access material uploaded by other website users, while the other 3 are proper e-learning platforms, fully set up by teachers either as a complement - before the pandemic- or a substitute -during the pandemic- to on site teaching

Figure 1: Google Searches in Italy for 5 selected keywords



Note: This figure downloads the data from Google Trends for the keywords Google Classroom, WeSchool, Edmodo, Studenti.it and Scuola.net setting the country of Italy as the geographic area. We download two bundles of 5 series each. The first bundle - top graph- contains series spanning from September 2016 to June 2020. The second bundle - bottom graph- spans from September 2016 to December 2019. Given that the series are downloaded in bundles, the series in each graph are comparable within and across across themselves.

of Google Trends' data and justifies our choices on how to use it: We download an independent series per region and keyword - without relying on the comparison feature of approach 2)[3]-, because otherwise, few regions with very dramatic changes would push the series of all the other regions towards zero, despite experiencing an increase themselves too. If we proceeded this way, we would loss much of the necessary variability across regions in our sample[4]. Note that this is just a consequence of the choice on how to download the data, it is not the value of the data itself: If the fraction of searches for a given search term is effectively zero, the series would have the zero value regardless of how we download the data. The top figure illustrates this point. As it is clear there, *Google Classroom*, *WeSchool*, and in particular, *Edmodo*, have now values far from zero in the pre-pandemic times. The increase due to the Covid-19 outbreak was so dramatic, that *relative* to post-pandemic numbers, Italians were barely searching for e-learning platforms -even though their search was not zero in absolute terms.

Unavoidably, because to perform the before and after analysis one needs to download the entire time window, and because the increase on the google searches after the Covid-19 outbreak is very big in most regions for the 3 online learning platforms, our single keyword-region series have plenty of small values along the pre-pandemic period.

Finally, we also use the top figure in Figure 1 as supporting evidence to validate the use of Google Trends data to understand the engagement of Italian students on online learning over time. The figure clearly shows that the fraction of searches experiences a big fall during the summer break and that also falls, to a lesser extent, during Christmas break and Easter holidays. While the level is highest for *Edmodo*, showing that it was the most searched e-learning platform in Italy before the pandemic, the same pattern is clearly followed by Google Classroom and WeSchool too.[5].

3.2 INVALSI

To measure academic performance at the regional level, we use data from INVALSI, the National Institute for the Valuation of the Education and Training System.[6] This institute organizes yearly standardised tests to assess students' performance at primary school (2^{nd} and 5^{th} grades),

[3]Approach 2) implies downloading the data set in bundles of 5 regions, keeping the keyword and time window constant.

[4]As noted before, we stress that the choice of downloading independent time series, one for each region and keyword, means that only changes -and not values- can be interpreted.

[5]To further validate the data, not only for Italy as a whole but across regions, we looked for the search of the keywords "WeSchool registrazione", "WeSchool accedi", "Google Classroom registrazione" and "Google Classroom accedi" in the two most populated regions in Italy, Lombardy and Campania. As before, we look at the two series in a single bundle, so that the levels are comparable across the two regions. Lombardy belongs to the set of northern regions and Campania to the set of southern regions. Google Trends responds that there is not enough data to show for "WeSchool registrazione". For the rest of the keywords, we find that the following: 1) *relative* to the after the pandemic period, before the pandemic all the searches are close to zero, 2) after the pandemic, the searches for all the keywords spike, 3) after the pandemic, the searches for all the keywords are higher in Campania than in Lombardy.

[6]INVALSI Data is available at http://invalsi-serviziostatistico.cineca.it/ upon request. Methodological information is available at: http://invalsi-areaprove.cineca.it.

at lower secondary school (8^{th} grade), and at higher secondary school (10^{th} and 13^{th} grades).

For the purpose of this paper, we focus on the students evaluated at the 10^{th} grade. First of all, in Italy, there are more students enrolled in secondary schools than in primary school. Second, as students go up on the education system, many of them have extra motivation to study to get access to university. Third and finally, we give preference to the 10^{th} rather than the 13^{th} grade, as these are the students that are about to complete mandatory education.[7]

At the 10^{th} grade, two tests are administered to all students in the presence of an external examiner: Italian and Mathematics. In Table 1 we present the regional rankings in the academic year of 2018/2019.

Table 1: Regional Average Grades in INVALSI 10^{th} Grade Tests

Region	Average Italian	Ranking Italian	Average Math	Ranking Math
Lombardy	213	1	215	2
Veneto	213	2	216	1
Friuli-Venezia Giulia	209	3	214	3
Valle d'Aosta	208	5	205	10
Trentino-Alto Adige	208	4	211	4
Emilia-Romagna	207	6	210	5
Piedmont	206	7	207	7
Umbria	205	9	207	8
Liguria	205	8	206	9
Marche	204	10	208	6
Tuscany	200	11	203	11
Abruzzo	199	12	200	12
Lazio	198	13	196	14
Basilicata	196	14	196	13
Molise	194	15	195	15
Apulia	193	16	191	16
Campania	189	17	186	17
Sicily	187	18	184	18
Sardinia	183	19	178	19
Calabria	181	20	176	20

This table reports the regional average grades for the academic year 2018/2019. The dashed line divides the regions that are above and below the median across regional average grades.

INVALSI grades are reported according to the WLE (Weighted likelihood estimates) of individual parameters of the Rasch model (Rasch, 1993) where 200 matches the national average. We observe that regions in the South of the country tend to perform worse in both subjects but the ranking in Italian is not the same as in Mathematics. However, even if regions change position on each subject's ranking, none of them changes substantially relative to the median.

[7]For the purpose of categorising regions according to the academic performance it makes no statistical difference between using the 8^{th}, 10^{th}, or 13^{th} grades as their correlation at the regional level is approximately 85%.

3.3 COVID-19 Cases and Other Regional Data

Before providing additional details on the empirical strategy we describe here the source of our three crucial control variables. First, we control for the total number of COVID-19 cases reported daily for each region, provided by the Italian Ministry of Health on their website. Given that the North of the country was firstly and more severely hit by COVID-19, we want to clean our analysis from different trends in the virus spread that would induce different searches in e-learning platforms.

Second, we control for the share of home internet usage by region in 2019, obtained from ISTAT (National Statistics Institute), and collected by the Annual Questionnaire of Multiscopes for households in Italy. Although virtually all Italian households live in areas that are covered by broadband internet, not all households consume this service.[8] Within those that use internet we do not need to further control for internet speed as we can see in Figure A2 that all territories have access to similar levels of average download speed levels.

Finally, we include a northern dummy which follows the ISTAT terminology for statistical purposes. This dummy takes the value one for Emilia-Romagna, Friuli-Venezia Giulia, Lombardy, Piedmont, Trentino-Alto Adige, Valle d'Aosta, and Veneto This control is particularly useful as we exclude from our analysis other cultural characteristics that might drive differences in academic performance and e-learning usage.

4 Empirical Strategy

The time window for all the specifications in our main analysis is between the January 6 and June 7, 2020. To estimate the average effect of COVID-19 on the access to e-learning platforms across regions we perform a simple before and after analysis relative to the date of schools closure:

$$\begin{aligned}\ln(1+G.T.Index_{j,r,d}) = \alpha_0 + \alpha_1 \mathbb{1}AfterSchoolClosure_d + \\ + \gamma \ln(1+TotalCases_{rd}) + X'\delta + \lambda_j + \epsilon_{j,r,d}\end{aligned} \tag{1}$$

$\ln(1+G.T.Index_{j,r,d})$ is the log of Google Trends index for term j in region r in day d. Note that because the index includes zeros we shift it by one unit so that the dependent variable is defined. $\mathbb{1}AfterSchool\ Closure_d$ is an indicator variable that takes the value 1 from the day schools closed in region r and 0 otherwise.[9] $\ln(TotalCases_{rd})$ is the total number of COVID-19 cases for region r in day d, to capture the potential increase in the need to use more e-learning

[8] In 2017 the European Commission estimated that 99% of all Italian households lived in areas covered by fixed broadband, Commission (2017).

[9] Schools closures were implemented in Piedmonte, Emilia-Romagna, Lombardy and Veneto on February 23th 2020, in Marche and the province of Trento on February 24th, in Liguria on March 1st, and on March 4th for the rest.

rather than alternative live sources. X is a matrix of regional characteristics which includes the percentage of students in the population, to capture the potential relative interest for searching these words in Google; the share of the population using internet, to capture amount of terms searched in that region; and a dummy for whether the region is located in the South of the country to capture other culture characteristics that are common across regions, as well as the fact that the North of the country was firstly hit by the virus. δ is the vector of coefficients associated with the regional variables. λ_j are searched terms fixed effects. Finally, $\epsilon_{j,r,d}$ is the error term. The main coefficient of interest in this regression is given by α_1, which indicates the average increase in the search intensity across all the regions in the period after the school closures.

To study whether there were regional differences on the search intensity change after the school closures we split the regression above by academic performance, with the following difference-in-differences specification:

$$\begin{aligned}\ln\left(1+G.T.Index_{j,r,d}\right) = \alpha_0 &+ \alpha_1 \mathbb{1}AfterSchoolClosure_d + \beta_2 \ln(INVALSIScore_r) + \\ &+ \beta_3 \mathbb{1}AfterSchoolClosure_d \times \ln(INVALSIScore_r) + \\ &+ \gamma \ln(1+TotalCases_{rd}) + X'\delta + \lambda_j + \epsilon_{j,r,d} \end{aligned} \tag{2}$$

$\ln(INVALSIScore_r)$ is the log of the average regional score in the Italian language test performed by INVALSI. Our coefficient of interest is β_3, which measures the differential effect of having higher grades in INVALSI test on the search intensity after the school closures, relative to the period before the schools closure.

Finally, to show how the search intensity evolved over time before and after the date of schools closure, we re-estimate the same regression on a week-by-week basis. Formally, we mean:

$$\begin{aligned}\ln(1+G.T.Index_{j,r,w}) &= \textstyle\sum_{i=1}^{8+13}(\alpha_i \mathbb{1}Week_w) + \beta_1 \ln(INVALSIScore_r) + \\ &+ \textstyle\sum_{i=}^{8+13}(\beta_{i+1} \mathbb{1}Week_w \times \ln(INVALSIScore_r)) + \\ &+ \gamma \ln(1+TotalCases_{rd}) + X'\delta + \lambda_j + \epsilon_{j,r,d} \end{aligned} \tag{3}$$

$\mathbb{1}Week_w$ is an indicator variable which takes the value 1 on the respective calendar week w. Our coefficients of interest in this regression are all the ones associated to the weekly interaction terms.

5 Results

As explained above, we are first interested in quantifying the increase in e-learning platforms usage due to physical school closures. To do so we estimate regression 1 and present the results in Table 2. In the first column we pooled all search data across the main five e-learning platforms

in Italy, while results for each of them are shown in columns 2 to 6. All regressions are weighted by the 2019 regional populations to be nationally representative. Moreover, we include platform, year, and week of the year fixed effects.

In the first column we observe that, on average, regions increased the search of the e-learning terms by 185%, relative to the period before school closures. When we split the first regression by platform we observe that the increase in search intensity was stronger for Google Classroom than for WeSchool, and that the latter was particularly less used in North. While the share of internet usage positively contributes to more searches for the Google Classroom, this variable is not statistically significant for the WeSchool platform.

The results for the estimated coefficients of regression 2 are presented in Table 3. As before, the first column presents the main coefficients of interest in a regression with platform fixed effects but no controls. The second column includes the regional controls, as well as the total number of daily cases by region. In columns 3 and 4 we repeat the full specification for each platform separately.

Table 2: Before-After Regression on Google Search Index

	(1)	(2)	(3)	(4)	(5)	(6)
	All	GC	WS	Ed	Sc	St
After Regional Schools Closure	1.853***	3.023***	3.215***	2.152***	0.440***	0.418***
	(0.061)	(0.065)	(0.065)	(0.062)	(0.160)	(0.130)
Constant	1.361***	0.769***	0.569***	1.633***	1.753***	2.087***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.028)	(0.022)
Observations	19,776	4,120	4,120	4,120	3,708	3,708
Adjusted R-squared	0.476	0.864	0.866	0.795	0.213	0.231
Platform FEs	Yes	-	-	-	-	-
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week of year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the results from estimating equation 1 by ordinary least squares during the period of January 6th to June 7th 2020. The dependent variable is the logarithm of the Google Search Index for *Google Classroom* and *WeSchool*. *After Schools Closure* takes value 1 when schools closed in each region and 0 before. Columns 1 and 2 show the results using the full sample of observations (without and with controls), while columns 3 and 4 report the each regression for each searched term. The regression coefficients are weighted by each region's population and include fixed effects for week of year and year. Heteroskedasticity robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Difference-in-Difference Results

	(1) All	(2) GC	(3) WS	(4) Ed	(5) Sc	(6) St
INVALSI Score	-0.049	0.077	0.060	0.092	-0.198	-0.266
	(0.174)	(0.099)	(0.058)	(0.096)	(0.437)	(0.388)
After School Closure	1.030***	2.450***	2.167***	1.638***	-1.504***	-1.586***
	(0.250)	(0.202)	(0.292)	(0.209)	(0.469)	(0.435)
After School Closure × INVALSI Score	-0.188***	-0.291***	-0.212***	-0.142***	-0.183**	-0.291***
	(0.045)	(0.044)	(0.064)	(0.052)	(0.079)	(0.079)
North	0.078	0.061	0.005	-0.108	0.128	0.275
	(0.214)	(0.153)	(0.107)	(0.228)	(0.513)	(0.489)
ln(Covid-19 Cases)	0.113***	0.080***	0.129***	0.063**	0.264***	0.254***
	(0.029)	(0.024)	(0.036)	(0.030)	(0.058)	(0.051)
Share of Internet Access	0.020	0.023	-0.005	0.020	0.025	0.035
	(0.030)	(0.017)	(0.009)	(0.019)	(0.077)	(0.071)
Constant	-0.066	-1.375	0.421	0.235	-0.772	-0.911
	(2.310)	(1.278)	(0.701)	(1.375)	(5.857)	(5.407)
Observations	19,776	4,120	4,120	4,120	3,708	3,708
Adjusted R-squared	0.387	0.819	0.800	0.612	0.144	0.163
Platform FEs	Yes	-	-	-	-	-
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week of year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the results from estimating equation 1 by ordinary least squares during the period of June 27th 2016 to June 7th 2020. The dependent variable is the logarithm of the Google Search Index for selected E-learning platforms. *After Schools Closure* takes value 1 when schools closed in each region and 0 before. *INVALSI Score* represents the average score obtained in 2018 in the INVALSI test for Italian language. This variable has been standarized (demeaned and divided by its standard deviations) hence its units are standard deviations. *North* takes value 1 for Emilia-Romagna and all regions above it, and 0 otherwise. *Share of Internet Usage* contains the share of households in each region that used internet in 2019. *ln(COVID-19 Cases)* contains the total number of COVID-19 cases reported in each region and day. All regressions include fixed effects for week of year and year. GC stands for Google Classroom, WS for WeSchool, Ed for Edmodo, Sc for Scuola.net and St for Studenti.it. Bootstrapped standard errors are clustered by region and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

In the first column we observe that the different behaviour after the date of schools closure differed by regions with different INVALSI scores. Namely, regions with better academic performance have searched relatively less than regions with lower academic performance after that date. To provide a better visualisation of this difference we plot in Figure A3 the average search indices for regions above the median INVALSI score and for regions below the median INVALSI score. The figure clearly illustrates that while academically high and low performing regions have a similar pattern both before and after school's closure, the jump in the search intensity is substantially

different. Unlike Bacher-Hicks et al. (2021), who find that areas of the United States with higher income (revealed to be areas with average lower SAT scores, by Chetty et al. (2020)) demonstrated substantially larger increases in search intensity, our graph shows the opposite for the case of Italy. It was the academically low performing regions who experienced a larger increase in the search intensity.

Adding controls to the regression now increases the magnitude of the main coefficient both because, as shown in Table 1, the INVALSI scores are higher in the North of the country, and also because the number of daily cases was higher in the regions with better INVALSI scores. The signs of these two coefficients are the same as in the first regression.

Analysing the difference-in-differences results by platform we conclude that not only the jump in Google searches was larger for Google Classroom but that this one was relatively smaller for regions with better INVALSI scores. The differential effect is similar in the case of the WeSchool platform. While the North dummy and the Share of Internet Access are statistically significant for the search of WeSchool the same is not observed in the case of Google Classroom.

Finally, to observe how the difference in search intensity evolved along time, within each period (before and after the schools closure) we plot in Figure A4 the estimated effects on the weekly version of the difference-in-differences specification in regression 3. While there are no statistically significant differences in the period before relatively to the week of school closures, in the period after there is an average decrease of 8% towards regions with 1% better INVALSI scores. In the panel below in the same figure we also replicate the same exercise across Mathematics grades and we find that even though the magnitude of the effect is smaller the results are not qualitatively different.

6 Students and Teachers E-learning Before the Pandemic

6.1 Students - PISA

To analyse the usage of e-learning technologies by students in regions with different academic performance, we use data from PISA (Programme for International Student Assessment). PISA is an international standardised survey to 15-year-old students that is comprised of a cognitive test on reading, mathematics and science, and complementary questionnaires to assess students' attitudes and motivations. In this section we focus on the ICT Familiarity Questionnaire and the Educational Career Questionnaire.

While these questionnaires include a very rich set of questions, the caveat is that not all the regions participate in every wave. PISA 2015 provides data for Bolzano, Campania, Lombardy and Trento, while PISA 2018 provides data for Bolzano, Toscana, Sardegna and Trento. Note that both Bolzano and Trento (which form Trentino-Alto Adige) do not have publicly managed schools and therefore might be using e-learning differently than schools managed by the State. Excluding these

two regions, PISA 2018 does not include any other region from the "above median performance" group we consider in our main analysis. Therefore we use PISA 2015 and compare Lombardy with Campania. The two regions are among the most populated regions in Italy and have already been used as representative cases of the north-south divide in Italy in other studies (Acconcia and Graziano, 2017). In this analysis we use Lombardy as an example of the academically higher performing regions of the North and Campania as an example of the lower performing regions of the South.

From the various questions available, we focus on three that assess the ICT usage and availability outside school, as the availability and usage at school will be discussed in the data reported from teachers to INVALSI, in the next subsection. Panels A and B in Table 4 report differences in the usage of ICT resources for schoolwork, and additional instructions, respectively.

Table 4: ICT usage

Variable: Proportion of students	Campania (1)	Lombardy (2)	Difference (3)	Italy (4)
Panel A				
Outside school, at least once a week				
- for schoolwork	0.626	0.567	0.060***	0.591
	(0.013)	(0.013)	[0.001]	(0.009)
- to follow up school lessons	0.602	0.415	0.187***	0.504
	(0.014)	(0.013)	[0.000]	(0.009)
- for doing homework on computer	0.423	0.343	0.080***	0.362
	(0.014)	(0.012)	[0.000]	(0.009)
- for doing homework on mobile	0.416	0.266	0.150***	0.322
	(0.014)	(0.012)	[0.000]	(0.009)
Panel B				
Additional Math Instructions				
- Internet tutoring by a person or app	0.235	0.162	0.073***	0.185
	(0.017)	(0.016)	[0.002]	(0.011)
- Video recorded	0.168	0.069	0.099***	0.111
	(0.015)	(0.011)	[0.000]	(0.009)
Additional Italian Instructions				
- Internet tutoring by a person or app	0.275	0.226	0.049	0.263
	(0.020)	(0.023)	[0.112]	(0.016)
- Video recorded	0.155	0.103	0.052**	0.130
	(0.017)	(0.016)	[0.027]	(0.012)

The data reported in Panels A and B come from PISA 2015 ICT Familiarity Questionnaire and Educational Career Questionnaire respectively. Columns 1,2, and 4 report the proportion of students that answered positively to each of the metrics. Standard errors are reported in parenthesis. Column 3 reports the difference between Campania and Lombardy. The stars ,***,**,*, in this column indicate whether the difference is statistically significant at 1%,5%, and 10%, respectively. The p-values associated with the differences tests are reported in brackets. All averages are weighted by the PISA final trimmed non-response adjusted student weights.

Panel A in Table 4 shows clear evidence that already in the year 2015 students in Campania were using e-learning technologies for schoolwork outside school more than students in Lombardy. Students in Campania were 10.4% more likely to use internet for schoolwork, 45.1% more likely to use internet to follow-up school lessons, 23.3% more likely to do their homework using a computer and 56.4% more likely to do them using a mobile phone. As reported in the third column of Table 4, all these differences are statistically significant at a 1% level.

Panel B shows that students in Campania in 2015 were also more likely to use ICT in their additional instructions (not part of the student's mandatory school schedule) in both Mathematics and Italian. In both disciplines and regions, the internet tutoring was more common than the video recorded instructions but the differences across regions in Mathematics were statistically significant for both types of ICT. As for Italian, only video recorded instructions revealed a statistically significantly higher proportion of students using it in Campania, compared to Lombardy.

To evaluate whether the higher usage of ICT from students in Campania compared to Lombardy is driven by the ICT availability, in Table A3 we report the ICT access at home in the year 2015, where access is defined by having the digital device available and have used it at least once. Although students in Campania report higher availability of desktops, the proportion of students with laptops and tablets in Lombardy is statistically higher. When we look at internet, the proportion of students with internet connection is statistically smaller than that in Lombardy, but not when considered mobile internet. Therefore we conclude that the results above are not driven by students in Campania having higher access to ICT.

6.2 Teachers - INVALSI

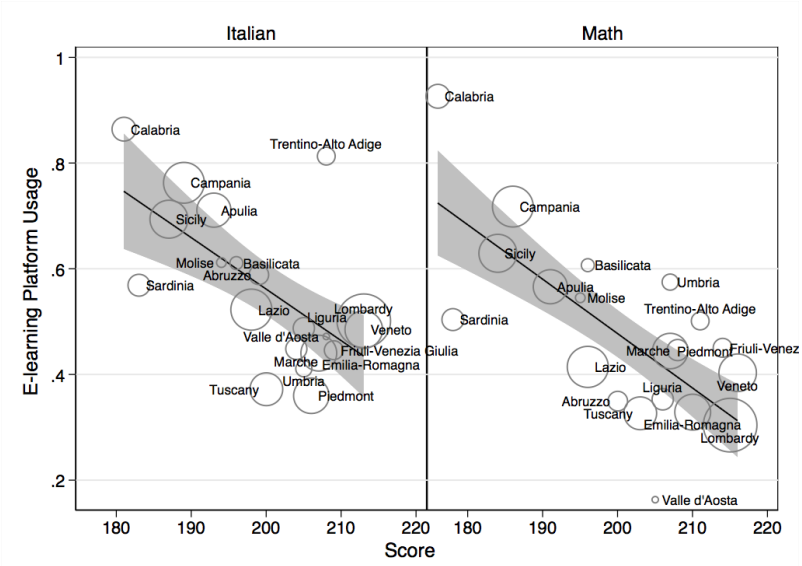
Together with the tests described in Section 3.2, INVALSI carries out surveys to students, teachers and school principals, allowing for a comprehensive evaluation of the Italian education system. Surveys to school teachers provide us with valuable evidence of the usage of e-learning technology in their didactic activity before the pandemic. In this section we analyse the answers provided to the question: "Thinking about the didactic activity you carried out this year, please indicate how often you carried out the following activities: e) use of e-learning platforms.", with the following response options being 0 =Never or almost never; 1 = Some times; 2 = Often; 3 = Always or almost always.

Figure 2, plots the average regional response to this question for teachers in Italian and Mathematics at Grade 10, across the average regional score in the same subject, according to the INVALSI tests in the academic year of 2019/2020. In both subjects there exists a negative relationship between the usage that teachers make of e-learning platforms when conducting their lessons and the regional mean academic performance of their students.

Comparing this with the academic year of 2017/18, where the question was also available, we observe a similar pattern. In Figure A6, we plot the average e-learning platform usage reported by

Italian and Mathematics' teachers in the two consecutive academic years preceding the pandemic, splitting regions by whether they are above or below the median regional scores for Grade 10. In regions with an average score below the median, teachers were more likely to report a higher usage of e-learning platforms. Although the differences are similar for both subjects, Italian language teachers responded with slightly higher frequent usage in both academic years.

Figure 2: Teachers' E-learning Platform Usage by Students' Academic Performance



Note: This figures shows the correlation between the reported usage of e-learning platforms by teachers when conducting their didactic activity in each region with the average results for the 2018/2019 INVALSI tests in Italian and Mathematics at Grade 10. The usage values for e-learning platforms are taken from the responses to question: *Thinking about the didactic activity you carried out this year, please indicate how often you carried out the following activities: e) use of e-learning platforms.* With the following response options: *0 = Never or almost never; 1 = Some times; 2 = Often; 3 = Always or almost always.* Sizes of circles correspond to the population share of each region, in 2019. The solid line corresponds to a linear fit weighted by the population share of each region. The shaded area corresponds to a 95% confidence interval of the linear fit.

Although there are substantial differences in the usage of e-learning, we do not observe the same when it comes to computer availability and usage in schools. In Figure A7 we plot the share of teachers that reported having access to a computer and its usage during their lessons between 2013 and 2019. Panel a) suggests that, until the academic year of 2017/2018, more teachers in higher academically performing regions had access to a computer to conduct their lessons, but in the last two academic years the two rates converged, especially in Italian. Moreover, if we analyse computer usage by conditioning on those teachers reporting having access to a computer during their lessons we observe no differences between regions with different academic performances. In Panel b), we observe that, once again, Italian language teachers report a higher usage of computers in classes but no difference between regions that are above and below the median scores. We take all this evidence of teachers in lower academically performing regions displaying a higher usage

of e-learning platforms during the years preceding the pandemic, as suggestive evidence of these teachers being more prepared to swiftly shift their lessons to online learning after the national school closure.

7 Academic Year of 2020/2021

From November 6th of 2020, the Italian Government divided the regions in three colours: yellow, orange and red, according to the different measures to contain the spread of COVID-19 in the second wave. The rules of e-learning (*Didattica a Distanza*) changed, affecting mostly Secondaria di Secondo Grado (from grade 10th onward) in red zones. From January 2021, a new colour was introduced - white, where most of the measures would not be in place.

Table 5: Descriptive Statistics on the Colour System during Schooling Days of 2020/2021

	First Date	Last Date	Nr of Times	Nr. Days			
Region	Red Zone	Red Zone	Red Zone	Red	Orange	Yellow	White
Abruzzo	15nov2020	23dec2020	2	33	97	63	12
Apulia	06nov2020	25apr2021	4	41	72	87	5
Basilicata	21dec2020	21mar2021	2	24	90	91	0
Calabria	08nov2020	12apr2021	4	33	88	84	0
Campania	06nov2020	18apr2021	5	62	31	112	0
Emilia-Romagna	21dec2020	12apr2021	3	26	73	101	5
Friuli-Venezia Giulia	21dec2020	12apr2021	3	26	60	100	19
Lazio	21dec2020	28mar2021	2	17	41	142	5
Liguria	21dec2020	28mar2021	2	10	78	105	12
Lombardy	08nov2020	12apr2021	5	54	67	79	5
Marche	21dec2020	31mar2021	2	20	73	112	0
Molise	21dec2020	21mar2021	2	24	33	129	19
Piedmont	08nov2020	12apr2021	5	40	75	85	5
Sardinia	06nov2020	02may2021	3	25	47	93	40
Sicily	21dec2020	31jan2021	2	18	101	86	0
Trentino-Alto Adige	08nov2020	31mar2021	6	69	74	57	5
Tuscany	15nov2020	12apr2021	4	33	82	90	0
Umbria	21dec2020	23dec2020	1	3	125	72	5
Valle d'Aosta	06nov2020	09may2021	5	62	67	76	0
Veneto	21dec2020	31mar2021	2	20	51	122	12

This table reports the descriptive statistics of the colour system in Italian regions between November 6, 2020 and June 18, 2021. Trentino-Alto Adige aggregates Bolzano and Trento in order to make it compatible with the Google Trends data, provided for 20 regions. Therefore, the colour in this region is assumed to be the highest, in the colour scale, between the two.

Table 5 shows the descriptive statistics of the different colours across regions between the 6th of November (beginning of the new colour system) and the 18th of June (end of the academic year).

The first two columns show a large variation in the dates each region entered first and exited last from red zones. The earliest regions were Apulia, Campania, Sardinia, and Valle d'Aosta, and most of them were also the latest to leave this colour. In fact, Campania and Valle d'Aosta are among the top three regions, together with Trentino-Alto Adige, with more times (at least 5) and more days in red zone (at least 62). On the other end Umbria was only once in Red zone during three days, having the record of days in orange zone.

The variation in time and region allows us to test whether different regions used E-learning differently despite of being under the same rules. To do so, we run the following specifications:

$$\begin{aligned}\ln\left(1+G.T.Index_{r,d}\right) = \alpha_0 + \alpha_1 \mathbb{1}RedZone_{r,d} + \alpha_2 \mathbb{1}OrangeZone_{r,d} + \alpha_3 \mathbb{1}YellowZone_{r,d} + \\ + \gamma_1 \ln(1+TotalCases_{rd}) + \gamma_2 \mathbb{1}Weekend_d X'\delta + \epsilon_{r,d} \qquad (4)\end{aligned}$$

$$\begin{aligned}\ln\left(1+G.T.Index_{r,d}\right) = \alpha_0 + \sum_c \alpha_c \mathbb{1}ZoneC_{r,d} + \beta_2 INVALSIScore_r + \\ + \sum_c \delta_c \mathbb{1}ZoneC_{r,d} \times INVALSIScore_r + \\ +_1 \ln(1+TotalCases_{rd}) +_2 \mathbb{1}Weekend_d + X'\gamma + \epsilon_{r,d} \qquad (5)\end{aligned}$$

Unlike in the main analysis, here we only use searched for "Google Classroom" term as this was the platform that was widely applied in schools during the academic year of 2020/2021. In (4), $\mathbb{1}RedZone_{r,d}$ takes the value 1 if the region r is a red zone in day d, and zero otherwise, $\mathbb{1}OrangeZone_{r,d}$ takes the value 1 if the region r is a orange zone in day d, and zero otherwise, and $\mathbb{1}YellowZone_{r,d}$ takes the value 1 if the region r is a yellow zone in day d, and zero otherwise. The base group of the colour dummies aggregates both White Zone and common holidays across regions, i.e., Christmas and Easter, which are expected to be periods with low E-learning usage. Additionally to the main analysis, and because here we have daily data on the Google Trends index, we also add the dummy $\mathbb{1}Weekend_d$, which takes the value 1 every time the calendar day is on the weekend. The X matrix includes both the north dummy and the share of internet usage.

In (4) we assume all regions under the same colour have the same usage of "Google Classroom". In (5) we test if regions in similar lockdown rules present a different e-learning usage according to their average grade in Italian in INVALSI tests of 2018/2019. To do so, we interact the dummies associated to the colour of the region ($\mathbb{1}ZoneC_{r,d}$ equals one if region r is in color c, being c either red, orange, or yellow, in day d) with the standardized INVALSI score.

The first column in Table 6 reports the estimates from (4). As expected, the usage of E-learning, captured by the Google searches of the term "Google Classroom", is more intense with the intensity of the colour applied to each region, in comparison with periods in which the region

Table 6: Difference-in-Differences Results - Academic Year 2020/2021

Variables	(1) Google Classroom	(2) Google Classroom
𝟙RedZone	1.531***	2.585
	(0.146)	(2.980)
𝟙OrangeZone	1.394***	2.444*
	(0.045)	(1.429)
𝟙YellowZone	0.899***	2.247***
	(0.064)	(0.576)
INVALSI Score		-0.002
		(0.017)
𝟙RedZone × INVALSI Score		-0.006
		(0.017)
𝟙OrangeZone × INVALSI Score		-0.006
		(0.008)
𝟙YellowZone × INVALSI Score		-0.008**
		(0.003)
Share of Internet Usage	0.002	0.010
	(0.021)	(0.032)
ln(COVID-19 Cases)	0.380**	0.377**
	(0.183)	(0.188)
𝟙North	-0.545*	-0.467
	(0.303)	(0.386)
𝟙 weekend	-0.719***	-0.719***
	(0.056)	(0.057)
Constant	-2.435**	-2.666
	(1.145)	(2.779)
Observations	4,500	4,500
Adjusted R-squared	0.355	0.356

Note: This table reports the results from estimating equation 5. The dependent variable is the logarithm of the Google Search Index for *Google Classroom*. *RedZone*, *OrangeZone*, and *YellowZone* take value 1 when a region is, respectively, red, orange or yellow zone in a certain day. *INVALSI Score* contains the regional average score of the 2018/2019 INVALSI test in Italian. *North* takes value 1 for Emilia-Romagna and all regions above, and 0 otherwise. *Share of Internet Usage* contains the share of households in each region that used internet in 2019. *COVID-19 Cases* contains the total number of COVID-19 cases reported in each region and day. *Weekend* takes the value 1 on weekends, and 0 otherwise. Bootstrapped standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

is either White or in School Holidays (the base group). The second column reports the estimates from (5). Although the gradient of magnitudes is still observed in the group of colour dummies, the estimate associated to the Orange zone and the estimate associated with the Red Zone is not significant, which means that in Red Zones, the regions at the average of the INVALSI score (when the standardized variable takes the value zero) do not present different E-learning usage in comparison with White and School Holidays periods.

The significant estimates in both the base Yellow Zone dummy and its interaction with the INVALSI score indicate that there are relevant differences in the usage of E-learning in the Yellow Zones, where there should be a higher variability across schools, as the COVID-19 measures are less strict. The results show that regions at the average of the INVALSI score in the yellow zones tend to use more E-learning than when in White or in School Holidays, but this effect is mitigated as the INVALSI Score increases. This outcome related to the Yellow Zone corroborates what we have seen in the main analysis, i.e., regions with higher INVALSI score use less E-learning.

In order to test whether the different usage of E-learning during this academic year resulted in differences in academic performance, we do a last analysis in this section using the recently released grades of the academic year of 2020/2021.[10] To do so, we estimate the following regressions:

$$\begin{aligned}\Delta \ln INVALSIScore2021_r = \alpha_0 + \alpha_c \%DaysinZoneC_r + \beta_c AverageGTIndexinZoneC_r + \\ + \delta_c \%DaysinZoneC_r \times AverageGTIndexinZoneC_r + \\ + \gamma_1 \mathbb{1}North_r + \gamma_2 ShareInternet_r + \epsilon_r \end{aligned} \tag{6}$$

Equation (6) evaluates whether the percentage of days in each of the coloured areas, and the respective E-learning usage during those days, had an impact on the change between the grade in INVALSI tests in 2021 and the grade in the tests taken before the pandemic, in 2019. Due to the number of observations, and because the percentage of days are not mutually exclusive, we run (6) for each of the three colours, red, orange, and yellow. $\%DaysinZoneC_r$ indicates the percentage of days (between November 6 and June 18) in the zone of colour C. $AverageGTIndexinZoneC_r$ is the average of the Google Trends Index of the word "Google Classroom" during the days the region r was in the zone of colour C. Like in previous regressions, we also control for wether a given region is located in the north of the country, and also for the share of Internet usage in that region.

Table 7 reports the estimates from (6), applied to each of the three different colours. While both red and yellow regions do not present significant results, there seems to be a small (significant at 10% level) impact of the percentage of days a region was in Orange zone, and how the region used

[10] At the time of the writing of this version, only the regional averages were available. When we get the update of the microdata from INVALSI with student individual grades we should be able to also evaluate the dispersion within region.

Table 7: Impact of Closures and E-learning in INVALSI Scores - Academic Year 2020/2021

Variables	(1) ΔLn(INVALSI)	(2) ΔLn(INVALSI)	(3) ΔLn(INVALSI)
% Days in RedZone	0.079		
	(0.957)		
Average GTIndex in RedZone	-0.004		
	(0.026)		
– × % Days in RedZone	-0.025		
	(0.263)		
% Days in OrangeZone		-1.457*	
		(0.866)	
Average GTIndex in OrangeZone		-0.153*	
		(0.092)	
– × % Days in OrangeZone		0.479*	
		(0.277)	
% Days in YellowZone			0.222
			(1.010)
Average GTIndex in YellowZone			0.022
			(0.178)
– × % Days in YellowZone			-0.084
			(0.414)
𝟙North	0.022**	0.018**	0.019
	(0.010)	(0.009)	(0.014)
Share of Internet Usage	0.000	0.002	0.001
	(0.002)	(0.002)	(0.001)
Constant	-0.062	0.295	-0.194
	(0.123)	(0.352)	(0.451)
Adjusted R-squared	0.017	0.386	0.000
Observations	20	20	20

Note: This table reports the results from estimating equation 6. The dependent variable is the logarithm of the Google Search Index for *Google Classroom*. % Days in RedZone, OrangeZone and YellowZone are the percentages of days between November 6 2020 and June 18 2021 that a region was in, respectively, red, orange, or yellow zone. takes value 1 when a region is red zone in a certain day. Average GTIndex in RedZone, OrangeZone and YellowZone is the average Google Trends index of the word "Google Classroom" during the days the region was either Red, Orange, or Yellow, respectively. *North* takes value 1 for Emilia-Romagna and all regions above, and 0 otherwise. *Share of Internet Usage* contains the share of households in each region that used internet in 2019. *** p<0.01, ** p<0.05, * p<0.1.

E-learning, on the change in the average grade of that region between 2019 and 2021. Namely, the more the region was in orange zone, the worse were the grades in 2021, when compared to 2019. However, regions that were in orange zones for a longer period and that used more E-learning during that period seemed to have, in part, mitigated the negative change in grades. Finally, it should also not be neglected that despite the differences in percentage of days in each of the colours, and the variation in the usage of E-learning, the differences between North and Southern Italy persisted in 2021.

8 Conclusion

With the enforcement of lockdown and school closures, Italy, like many other countries worldwide in 2020, was forced to implement e-learning tools in an unprecedented way. While some online resources such as *Studenti.it Scuola.net*, and *Edmodo* were already widely used by students and teachers before the pandemic to support education outside class, the usage of Google Classroom and WeSchool, implemented with the government program *Didattica a Distanza* was virtually zero before the academic year of 2019/2020.

As Italy is a country with large regional differences in both school quality, academic performance, and economic development, we study whether the usage of newly implemented e-learning tools was statistically different by regions with different academic performance. In the main analysis of this paper, we employ a similar methodology as Bacher-Hicks et al. (2021) that studies regional differences by socio-economic status, recurring to the Google Trends. Surprisingly we find the opposite effects than those found for the United States, i.e., regions with lower academic performance (and lower average socio-economic status) had higher increases in the search of these tools.

We further investigate the origins of these differences, by using questionnaires to students (PISA) and teachers (INVALSI) on the usage of computer, internet, and e-learning tools before the pandemic. The analysis of these results show that both students and teachers in regions with lower academic performance were already using relatively more information and technology, and e-learning platforms before the national schools closure.

In the last part of the paper, we analyse consequences to the 2020/2021 Academic Year. We conclude that, according to each colour a region was set, i.e., depending on the restrictions in terms of school closures and recommendations to use *Didattica a Distanza*, the regions used more E-learning in Orange than in Yellow Zones, and even more in Red Zones, especially when all compared with White Zones and with periods of School Holidays, between November 6 and June 18. We also conclude that despite the larger usage of E-learning in Yellow Zones by regions that performed worse in 2018/2019, such differences did not have an impact in the grades of High School in 2021. Nevertheless, we do observe a non-negligible impact of how regions in Orange Zones used E-learning on the change between grades of 2021 and 2019. More specifically, regions that were in

orange zones for a longer period and that used more E-learning during that period seemed to have, in part, mitigated the negative change in grades. Both results seem to indicate that, in case of another year of restrictions and e-learning, there should be less discretion, and more school specific policy measures to try to diminish the still evident differences between North and Southern Italy.

References

ACAPS (2020). Covid-19 government measures dataset. data available at https://www.acaps.org/special-report/covid-19-government-measures-update-0.

Acconcia, G. and Graziano, P. R. (2017). *The Youth Guarantee: The Cases of Milan (Lombardy) and Naples (Campania)*, pages 339–348. Palgrave Macmillan UK.

Agasisti, T. and Vittadini, G. (2012). Regional economic disparities as determinants of student's achievement in italy. *Research in Applied Economics*, 4(2):33.

Andrew, A., Cattan, S., Dias, M. C., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., and Sevilla, A. (2020). Inequalities in children's experiences of home learning during the covid-19 lockdown in england. *IFS???*

Bacher-Hicks, A., Goodman, J., and Mulhern, C. (2021). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, 193:104345.

Baker, S. R. and Fradkin, A. (2017). The impact of unemployment insurance on job search: Evidence from google search data. *Review of Economics and Statistics*, 99(5):756–768.

Bando, R., Gallego, F., Gertler, P., and Fonseca, D. R. (2017). Books or laptops? the effect of shifting from printed to digital delivery of educational content on learning. *Economics of Education Review*, 61:162–173.

Beuermann, D. W., Cristia, J., Cueto, S., Malamud, O., and Cruz-Aguayo, Y. (2015). One laptop per child at home: Short-term impacts from a randomized experiment in peru. *American Economic Journal: Applied Economics*, 7(2):53–80.

Brown, B. W. and Liedholm, C. E. (2002). Can web courses replace the classroom in principles of microeconomics? *American Economic Review*, 92(2):444–448.

Brunello, G. and Checchi, D. (2005). School quality and family background in italy. *Economics of Education Review*, 24(5):563–577.

Castelnuovo, E. and Tran, T. D. (2017). Google it up! a google trends-based uncertainty index for the united states and australia. *Economics Letters*, 161:149–153.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2020). Income Segregation and Intergenerational Mobility Across Colleges in the United States*. *The Quarterly Journal of Economics*, 135(3):1567–1633.

Commission, E. (2017). Europe's digital progress report 2017: Connectivity. page 41.

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., and Severín, E. (2017). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics*, 9(3):295–320.

Dergiades, T., Milas, C., and Panagiotidis, T. (2015). Tweets, google trends, and sovereign spreads in the giips. *Oxford Economic Papers*, 67(2):406–432.

Engzell, P., Frey, A., and Verhagen, M. (2020). Pre-analysis plan for: Learning inequality during the covid-19 pandemic. *mimeo???*

Fairlie, R. W. and Robinson, J. (2013). Experimental evidence on the effects of home computers on academic achievement among schoolchildren. *American Economic Journal: Applied Economics*, 5(3):211–40.

Giuliani, D., Dickson, M. M., Espa, G., and Santi, F. (2020). Modelling and predicting the spatio-temporal spread of coronavirus disease 2019 (covid-19) in italy. *Available at SSRN 3559569*.

Hamid, A. and Heiden, M. (2015). Forecasting volatility with empirical similarity and google trends. *Journal of Economic Behavior & Organization*, 117:62–81.

Joyce, T., Crockett, S., Jaeger, D. A., Altindag, O., and O'Connell, S. D. (2015). Does classroom time matter? *Economics of Education Review*, 46:64–77.

Lu, Y. and Song, H. (2020). The effect of educational technology on college students' labor market performance. *Journal of Population Economics*, 33(3):1101–1126.

Murnane, R. J., Willett, J. B., and Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics*, 2(77):251–266.

Peracchi, F. (2008). Height and economic development in italy, 1730-1980. *American Economic Review*, 98(2):475–81.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests.* ERIC.

Rodríguez-Planas, N. (2020). Hitting where it hurts most: Covid-19 and low-income urban college students. Technical report, IZA Discussion Paper.

Vosen, S. and Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. google trends. *Journal of Forecasting*, 30(6):565–578.

Appendix

Table A1: Before-After Regression on Google Search Index

VARIABLES	(1) All	(2) All	(3) GC	(4) GC	(5) WS	(6) WS	(7) Ed	(8) Ed	(9) Sc	(10) Sc	(11) St	(12) St
After 4 March	1.820***		2.964***		3.157***		2.084***		0.449***		0.424***	
	(0.062)		(0.070)		(0.071)		(0.067)		(0.160)		(0.130)	
Before 15 Feb. after 15 Mar.		1.933***		3.195***		3.426***		2.265***		0.327*		0.431***
		(0.065)		(0.055)		(0.053)		(0.056)		(0.178)		(0.148)
Constant	1.367***	1.348***	0.779***	0.745***	0.580***	0.544***	1.642***	1.615***	1.754***	1.755***	2.087***	2.092***
	(0.008)	(0.008)	(0.009)	(0.008)	(0.010)	(0.008)	(0.009)	(0.008)	(0.028)	(0.028)	(0.022)	(0.022)
Observations	19,776	19,392	4,120	4,040	4,120	4,040	4,120	4,040	3,708	3,636	3,708	3,636
Adjusted R-squared	0.475	0.479	0.858	0.863	0.860	0.866	0.790	0.793	0.213	0.215	0.231	0.237
Platform FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week of year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the results from estimating equation 1 by ordinary least squares during the period of January 6th to June 7th 2020. The dependent variable is the logarithm of the Google Search Index for *Google Classroom* and *WeSchool*. *After March 4* takes value 1 after March 4 2020 and 0 before. *Before 15 Feb. After 15 March* takes value 1 after March 15 2020 and 0 before 15 February. Columns 1 and 2 show the results using the full sample of observations (without and with controls), while columns 3 and 4 report the each regression for each searched term. The regression coefficients are weighted by each region's population and include fixed effects for week of year and year. Heteroskedasticity robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Difference-in-Difference Results

VARIABLES	(1) All	(2) All	(3) GC	(4) GC	(5) WS	(6) WS	(7) Ed	(8) Ed	(9) Sc	(10) Sc	(11) St	(12) St
INVALSI Score	-0.048	-0.049	0.079	0.074	0.063	0.055	0.094	0.095	-0.198	-0.184	-0.266	-0.266
	(0.174)	(0.175)	(0.099)	(0.102)	(0.058)	(0.057)	(0.096)	(0.097)	(0.437)	(0.433)	(0.389)	(0.389)
North	0.078	0.082	0.062	0.066	0.006	0.013	-0.108	-0.112	0.124	0.115	0.271	0.278
	(0.214)	(0.215)	(0.155)	(0.157)	(0.107)	(0.108)	(0.229)	(0.232)	(0.512)	(0.508)	(0.489)	(0.486)
ln(Covid-19 Cases)	0.124***	0.096*	0.097***	0.024	0.150***	0.077	0.089**	0.082	0.272***	0.411***	0.258***	0.376***
	(0.032)	(0.053)	(0.028)	(0.046)	(0.041)	(0.058)	(0.035)	(0.073)	(0.064)	(0.107)	(0.059)	(0.088)
Share of Internet Access	0.020	0.020	0.023	0.024	-0.005	-0.004	0.020	0.020	0.025	0.024	0.035	0.035
	(0.030)	(0.031)	(0.017)	(0.017)	(0.009)	(0.010)	(0.019)	(0.019)	(0.077)	(0.077)	(0.071)	(0.071)
Constant	-0.060	-0.103	-1.365	-1.431	0.433	0.322	0.250	0.207	-0.769	-0.656	-0.909	-0.927
	(2.307)	(2.331)	(1.276)	(1.308)	(0.702)	(0.718)	(1.379)	(1.399)	(5.861)	(5.814)	(5.410)	(5.408)
Observations	19,776	19,392	4,120	4,040	4,120	4,040	4,120	4,040	3,708	3,636	3,708	3,636
Adjusted R-squared	0.386	0.386	0.817	0.816	0.798	0.798	0.610	0.605	0.144	0.146	0.163	0.167
Platform FEs	Yes	Yes	-	-	-	-	-	-	-	-	-	-
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week of year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

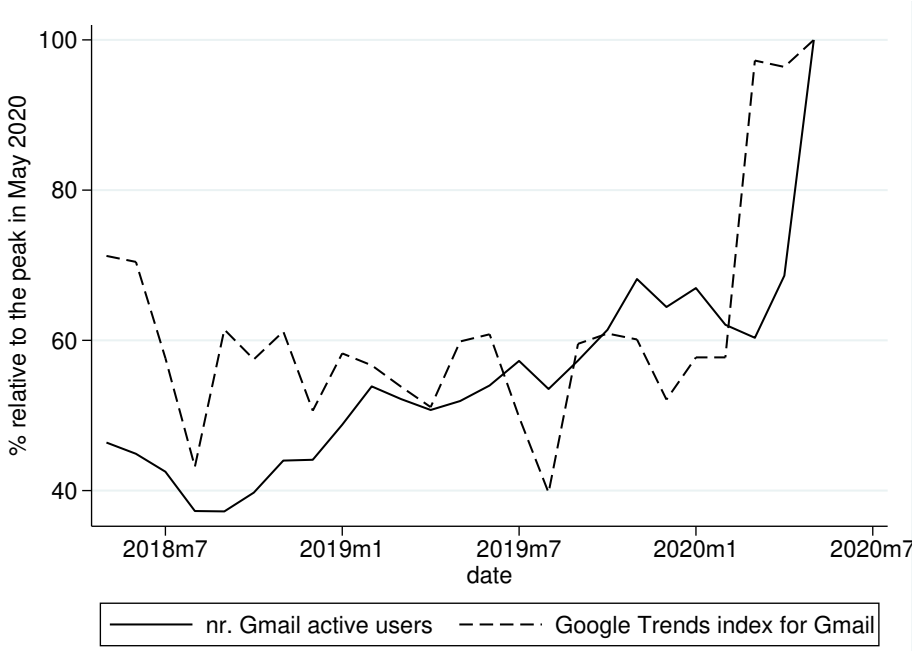
Notes: This table reports the results from estimating equation 1 by ordinary least squares during the period of June 27th 2016 to June 7th 2020. The dependent variable is the logarithm of the Google Search Index for selected E-learning platforms. *After March 4* takes value 1 after March 4 2020 and 0 before. *Before 15 Feb. After 15 March* takes value 1 after March 15 2020 and 0 before 15 February. *INVALSI Score* represents the average score obtained in 2018 in the INVALSI test for Italian language. This variable has been standarized (demeaned and divided by its standard deviations) hence its units are standard deviations. *North* takes value 1 for Emilia-Romagna and all regions above it, and 0 otherwise. *Share of Internet Usage* contains the share of households in each region that used internet in 2019. *ln(COVID-19 Cases)* contains the total number of COVID-19 cases reported in each region and day. All regressions include fixed effects for week of year and year. GC stands for Google Classroom, WS for WeSchool, Ed for Edmodo, Sc for Scuola.net and St for Studenti.it. Bootstrapped standard errors are clustered by region and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: ICT access at home

Variable: Proportion of students	Campania	Lombardy	Difference	Italy
	(1)	(2)	(3)	(4)
Type of ICT				
- desktop/tablet/laptop	0.919	0.948	-0.028***	0.933
	(0.007)	(0.006)	[0.003]	(0.005)
- desktop computer	0.640	0.561	0.080***	0.587
	(0.013)	(0.013)	[0.000]	(0.009)
- portable laptop	0.733	0.773	-0.041**	0.763
	(0.012)	(0.011)	[0.011]	(0.008)
- tablet computer	0.578	0.619	-0.041**	0.562
	(0.013)	(0.012)	[0.022]	(0.009)
- internet connection	0.945	0.966	-0.020***	0.950
	(0.006)	(0.005)	[0.010)	(0.004)
- cell phone with internet	0.945	0.954	-0.009	0.945
	(0.006)	(0.006)	[0.313]	(0.004)

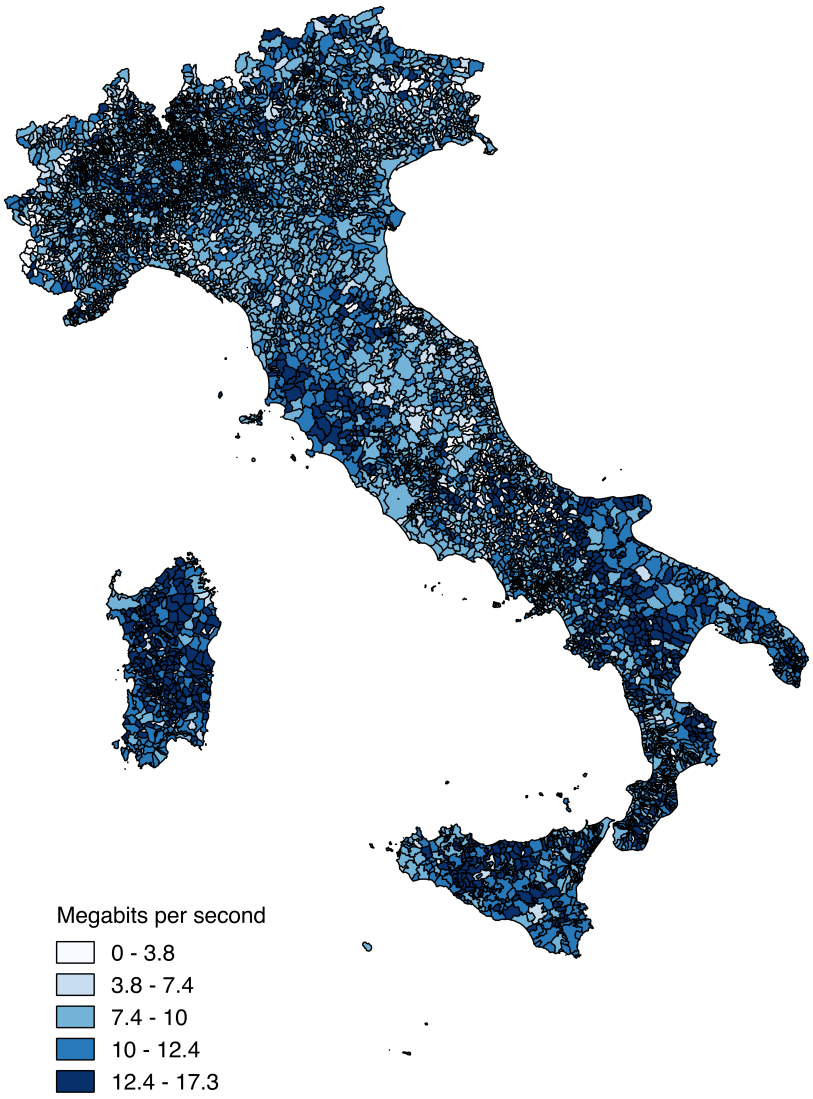
The data comes from PISA 2015 ICT Familiarity Questionnaire. Columns 1,2, and 4 report the proportion of students that answered that the reported device was available for them *and they used it*. The metric is equal to 0 if the device was available at home for them but they did not use it or if it was not available. The standard errors are reported in parenthesis. Column 3 reports the difference between Campania and Lombardy. The stars ,***,**,*, in this column indicate whether the difference is statistically significant at 1%,5%, and 10%, respectively. The p-values associated with the differences tests are reported in brackets. All averages are weighted by the PISA final trimmed non-response adjusted student weights.

Figure A1: Comparison between number of active Gmail users and Google Trends Index for Gmail



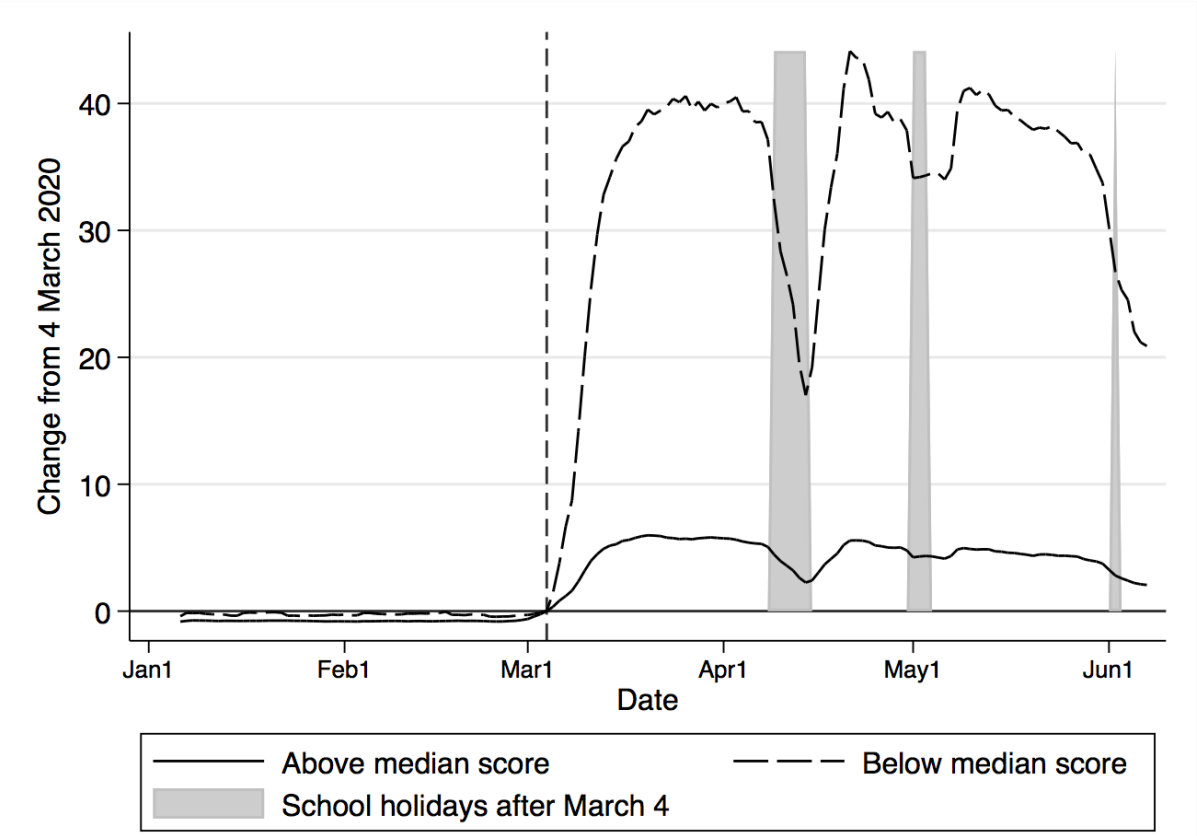
Note: This figure plots the average monthly number of active users of Gmail, provided by AirnowData, and the average monthly Google Trends index for Gmail, between May 2018 and May 2020. Both series are rescaled relative to the peak in May 2020.

Figure A2: Geographic Distribution of ADSL Download Speed in 2018



Note: This figure plots the average ADSL download speed in each Italian municipality in December 2018. Lighter colors indicate no data or low downlad speeds while darker colors represent higher average download speeds. Source: Autorità per le Garanzie nelle Comunicazioni (AGCOM).

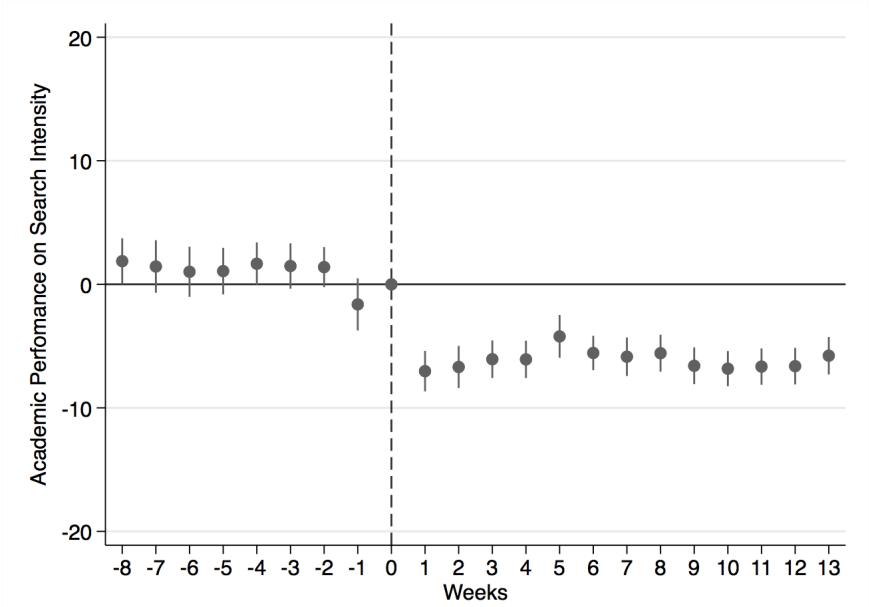
Figure A3: Google Trends Search Index for Google Classroom
by Academic Performance



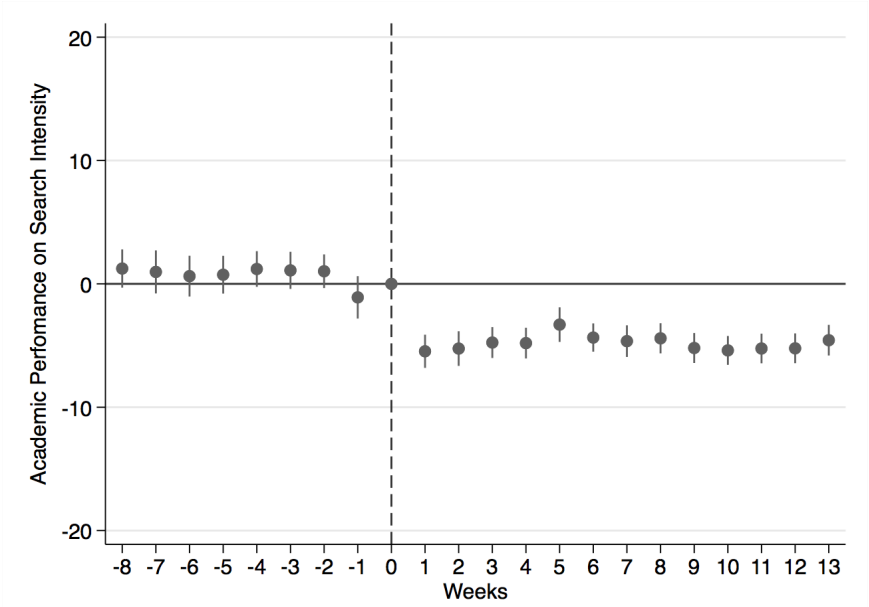
Note: This figure plots smooth daily changes of the Google Trends search index for the term *Google Classroom* in two groups of regions relative to March 4, 2020. The smoothing technique corresponds to the 7 day moving average computed by taking the average of each daily value together with the values from the preceding 6 days. Search index represented under below (above) median score contain the population weighted mean of the search index for the regions with a mean score in Italian below (above) the national median. Regional mean scores in Italian are extracted from the 2018 INVALSI report corresponding to Grade 10 students. Regional population shares used for the weights correspond to 2019 and are extracted from ISTAT. Shaded areas mark the official school holidays including the adjacent weekends that happened after the national closing of schools on March 4, date represented by the dashed line. See the footnote of Figure **??** for the precise days tha have been considered as school holidays.

Figure A4: Difference in E-learning Adoption by Academic Performance

(a) Italian



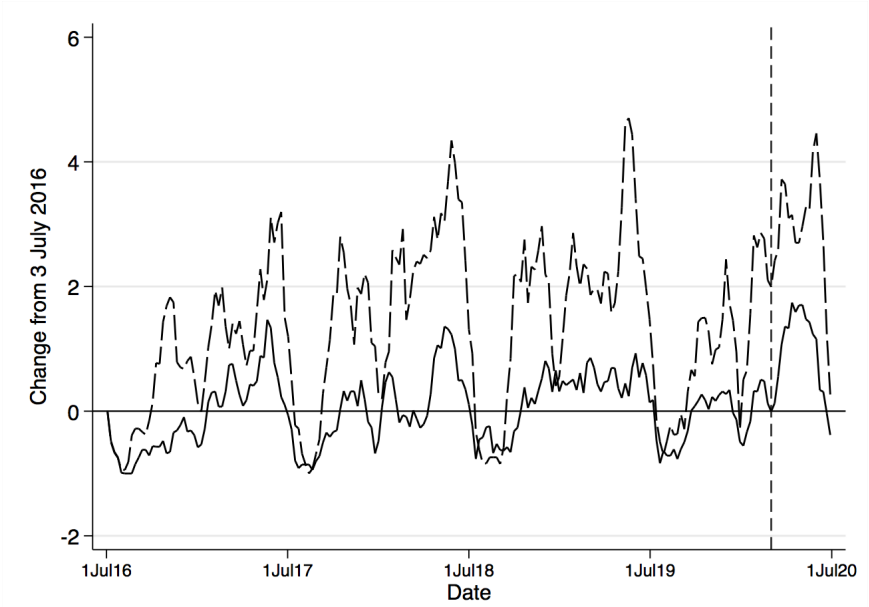
(b) Mathematics



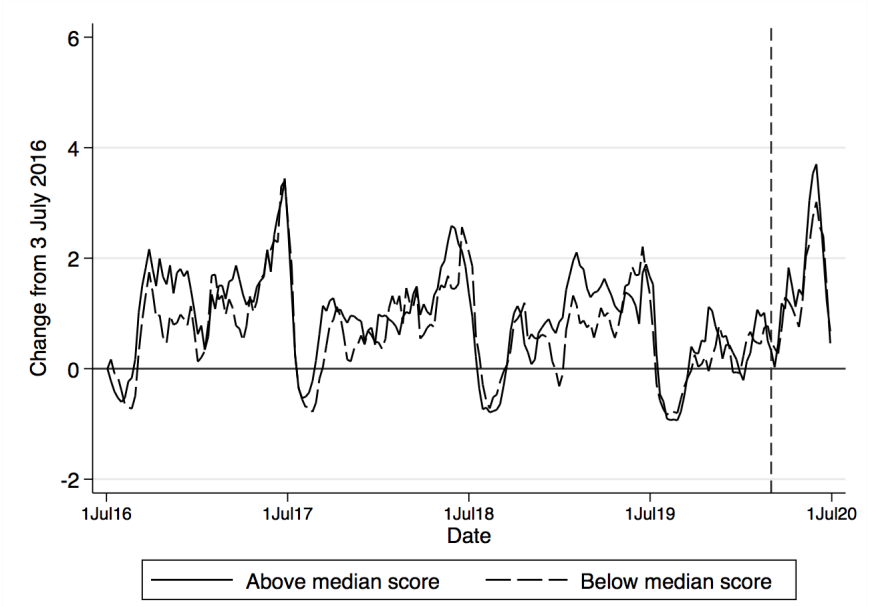
Note: This figure plots selected coefficients resulting from estimating by ordinary least squares equation 3. Panel (a) reads as one week after schools closed, Italian regions with a 1% higher average academic performance in Italian language are associated with a 8% on average lower level of internet searches of the two selected online learning platforms. The coefficient plotted are those in front of the interaction term between the natural logarithm of *INVALSI Score* and *AfterSchoolClosure*. Included controls and fixed effects are the same one as those detailed in column 2 of Table 3. The sample is comprised of weekly values on the Search Intensity Index from January 6 to June 7 2020 for three selected terms "*Google Classroom*" and "*WeSchool*". Vertical solid lines represent 95% confidence intervals computed using heteroskedasticity robust standard errors. Vertical dashed line marks the week of March 2 to 8 when all schools had to close. The regression is weighted by each region's population.

Figure A5: Google Trends Search Index for two E-learning Platforms by Academic Performance

(a) Scuola.net



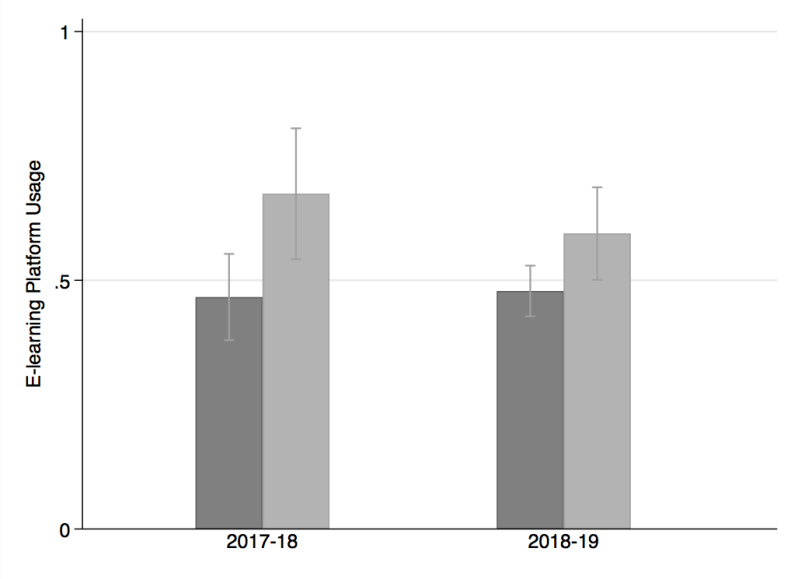
(b) Studenti.it



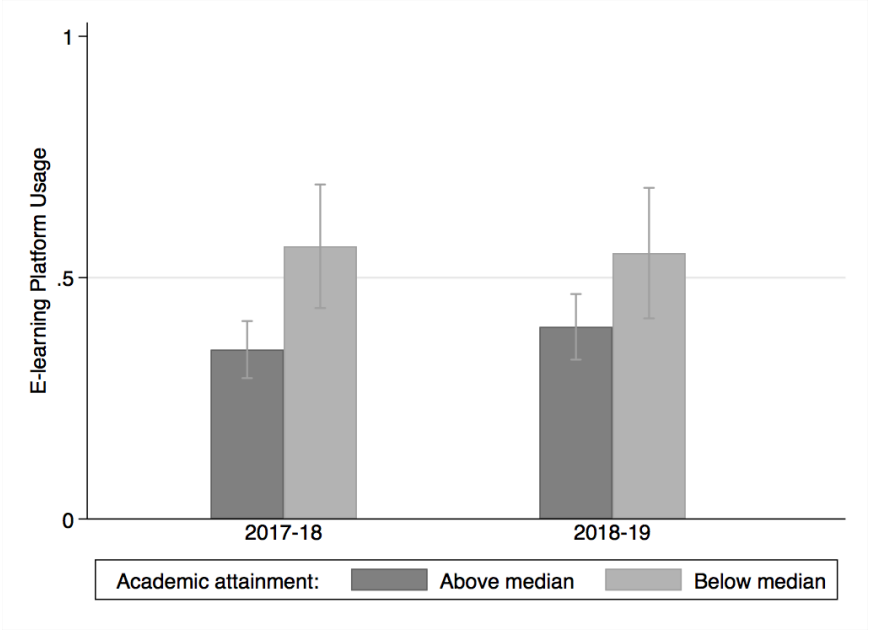
Note: This figure plots smooth weekly values of the Google Trends search index for the terms *Scuola.net* and *Studenti.it* in two groups of regions relative to July 3, 2016. The smoothing technique corresponds to the 4 week moving average computed by taking the average of each weekly value together with the values from the preceding 3 weeks. Search index represented under below (above) median score contain the population weighted mean of the search index for the regions with a mean score in Italian below (above) the national median. Regional mean scores in Italian are extracted from the 2018 INVALSI report corresponding to Grade 10 students. Regional population shares used for the weights correspond to 2019 and are extracted from ISTAT.

Figure A6: Teachers' E-Learning Platform Usage by Academic Performance

(a) Italian



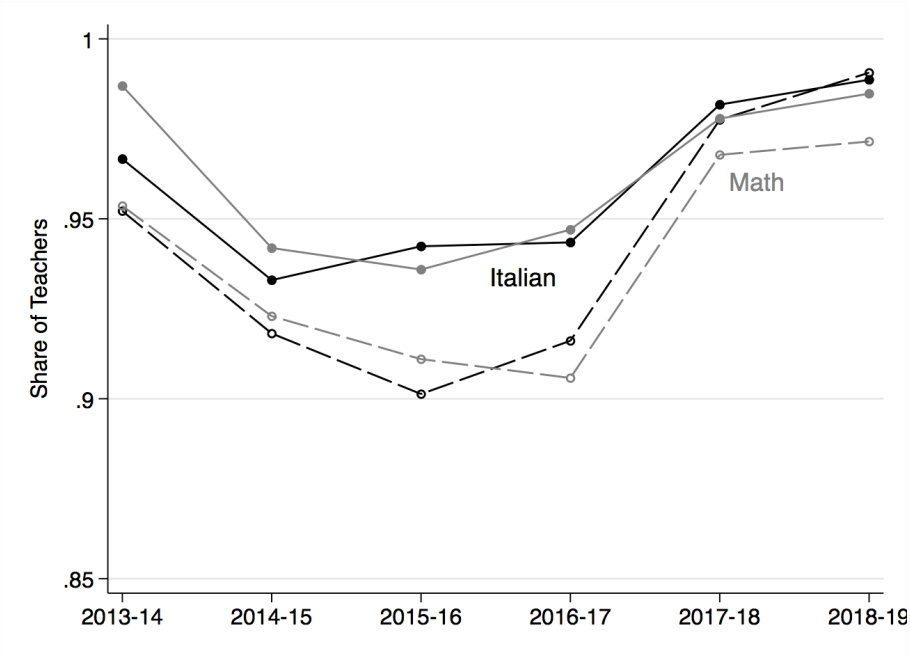
(b) Mathematics



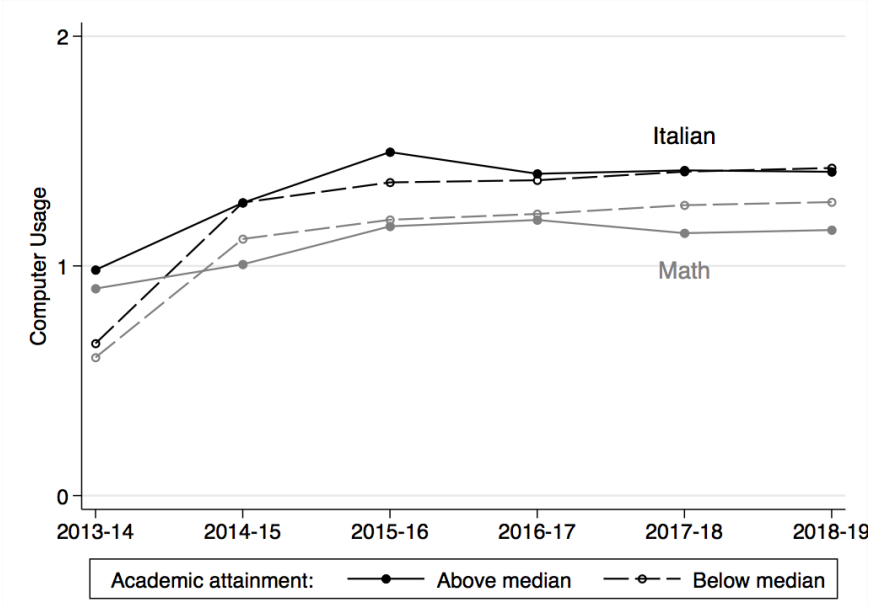
Note: This figure plots the reported usage of e-learning platforms by teachers when conducting their didactic activity in Italian and mathematics, respectively. The values are taken from the responses to question D2e administered by INVALSI to Grade 10 teachers both subjects during 2017-2018 and 2018-2019. The question states: *Thinking about the didactic activity you carried out this year, please indicate how often you carried out the following activities: e) use of e-learning platforms.* With the following response options: *0 = Never or almost never; 1 = Some times; 2 = Often; 3 = Always or almost always.* Below (above) the median contain the population weighted mean of the responses in the regions with a mean score in each subject below (above) the national median, respectively. Regional mean scores in both subjects are extracted from the 2018 INVALSI report corresponding to Grade 10 students. Regional population shares used for the weights correspond to 2019 and are extracted from ISTAT. Vertical lines represent the 95% confidence interval of the sample mean.

Figure A7: Computer Availability and Usage by Teachers in Class by Academic Performance

(a) Computer Availability



(b) Computer Usage



Note: The figure plots the proportion of Italian and mathematics teachers reporting having access to a computer, panel (a), and their usage, panel (b), in class during their lessons. Values are taken from a specific responses to question D6a administered by INVALSI to Grade 10 teachers of both subjects from 2013 to 2019, which states: *How much did you use the computer in lessons with the students of your class in the last school year?* Panel (a) plots one minus the share of teachers who responded *Not present in School*. Panel (b) plots the group average of the following response options: *0 = I don't use it; 1 = Occasional use; 2 = Regular use*. The answers used are those from teachers responding that there was a computer available in class. Below (above) the median contain the population weighted mean of the responses in the regions with a mean score in each subject below (above) the national median, respectively.