UNIVERSITÀ CATTOLICA DEL SACRO CUORE
Dipartimento di Economia e Finanza

Working Paper Series

Let's tweet again? The impact of social networks on literature achievement in high school students: Evidence from a randomized controlled trial

Gian Paolo Barbetta, Paolo Canino, Stefano Cima

Working Paper n. 81

May 2019

Let's tweet again? The impact of social networks on literature achievement in high school students: Evidence from a randomized controlled trial

Gian Paolo Barbetta
Università Cattolica del Sacro Cuore, Fondazione Cariplo

Paolo Canino
Fondazione Cariplo

Stefano Cima
Fondazione Cariplo

Working Paper n. 81
May 2019

Dipartimento di Economia e Finanza
Università Cattolica del Sacro Cuore
Largo Gemelli 1 - 20123 Milano – Italy
tel: +39.02.7234.2976 - fax: +39.02.7234.2781
e-mail: dip.economiaefinanza@unicatt.it

Let's tweet again?
The impact of social networks on literature achievement in high school students: Evidence from a randomized controlled trial

Gian Paolo Barbetta,[§] Paolo Canino,[#] and Stefano Cima[#]

May 2019

Abstract

The availability of cheap wi-fi internet connections has stimulated schools to adopt Web 2.0 platforms for teaching. Using social networks and micro-blogs, teachers aim to stimulate students' participation in school activities and their achievement. Although anecdotal evidence shows a high level of teacher satisfaction with these platforms, only a small number of studies has produced rigorous estimates of their effects on students' achievement.

We contribute to the knowledge in this field by analyzing the impact of using micro-blogs as a teaching tool on the reading and comprehension skills of students. Thanks to a large-scale randomized controlled trial, we find that using Twitter to teach literature has an overall negative effect on students' average achievement, reducing performance on a standardized test score by about 25 to 40% of a standard deviation. The negative effect is heterogeneous with respect to some students' characteristics. More specifically, the use of this Web 2.0 application appears to have a stronger detrimental effect on students who usually perform better.

JEL classification: I21

Keywords: ICT, education, literature performance, RCT.

[§] Corresponding author: gianpaolo.barbetta@unicatt.it.
Università Cattolica del Sacro Cuore (Department of economics and finance) and Fondazione Cariplo.
ORCiD:0000-0001-5865-3766
[#] Fondazione Cariplo.

1. Introduction

Over the last 30 years, information and communication technologies (ICT) have reshaped many aspects of our personal and professional lives. The education sector has not been immune from this trend. In fact, the gradual drop of ICT prices has greatly increased their availability for educational institutions. Thus, electronic devices (i.e., personal computers, multimedia projector screens, and e-book readers) and their applications are now common teaching tools in many schools, from primary through tertiary education. Moreover, over the last several years, almost all schools in developed countries—and many in developing ones—have gained full access to the internet, most of them through wi-fi connections. Thanks to the wide availability of these connections, the use of Web 2.0 applications, which include social networks such as Facebook and micro-blogging platforms such as Twitter,[1] in schools has also grown. Given that these platforms allow for interaction and collaboration, a growing number of educators have adopted them in the hopes of getting students more involved in school activities.

Since the early stage of their diffusion in schools, scholars have been puzzled about the possible effect of using ICT on students' learning performance. In fact, while these instruments are enthusiastically supported by educators, their impact on students' achievement is still controversial.

In this paper, we empirically assess the effect of using a Web 2.0 platform (Twitter) for teaching literature to high school students. Schools frequently adopt social networks and micro-blogging to facilitate communication between teachers and students and among schoolchildren themselves. Their use as a learning tool is also increasing. For example, in the field of literature, teachers use Twitter to comment about, discuss, and rewrite books with their students, with the aim stimulating students' active participation in the learning process and deepening their understanding of the material.

We use a large-scale randomized controlled trial (RCT) to analyze the impact of using Twitter as a literature teaching tool on the reading and comprehension skills of students. Our sample is about 1,500 students attending 70 Italian high schools in 2016/17. Schools that entered the experiment were randomly assigned into a treatment or a control group. The treatment group adopted Twitter to read, discuss, and study the book "*Il fu Mattia Pascal*," written by Italian Nobel Laureate Luigi Pirandello. The control group used traditional classroom teaching methods. Our outcome variable is represented by the performance of students on a standardized follow-up test measuring understanding, comprehension, and memorization of the book. To the best of our knowledge, this is the largest and most rigorous study of the effect of using micro-blogging as a teaching tool on

[1] While some scholars refer to Twitter as a micro-blogging service (Ebner, Lienhardt, Rohs, & Meyer, 2010), others refer to it as a "micro-blogging with social network features" (Veletsianos, 2012).

students' achievements. Therefore, these results may be useful to address the question about the effect of using Web 2.0 applications in school.

On the one hand, adopting social networks to teach literature has potential positive effects on students' achievement thanks to the high interaction between peers and teachers that the media guarantee as compared to traditional teaching methods. Moreover, young people are used to expressing their own opinions on social networks. This habit could increase their propensity to play an active role on a micro-blog when discussing the topics that emerge from reading, when commenting on and analyzing a book, and when confronting issues raised by teachers on the same media. Furthermore, teachers and peers can solicit students to take part in literary discussions out of school, therefore increasing the total time they devote to the topic. On the other hand, reading and discussing literary works on social media (using smartphones or other ICT devices) could lessen the attention students give to the text, inducing *skim reading*, and therefore hampering a full understanding of what they read. Likewise, social networks could represent a potential distraction for students, reducing the time they devote to reading and studying. Whether positive or negative effects prevail is a matter of measurement.

We find that using a micro-blog to teach literature has an overall detrimental effect on students' average achievement, reducing performance on a standardized test score by about 25–40% of a standard deviation. Our results represent a relevant and innovative contribution to the literature by analyzing the effect of ICT and Web 2.0 platforms on students' performance.

Moreover, these results have interesting policy implications. In fact, we find that the negative effect is heterogeneous with respect to students' characteristics. Students who generally perform better (female students, students born in Italy, students attending lyceums, and students with higher scores on baseline test) are harmed more compared to their counterparts. These results hint that Web 2.0 platforms decrease students' performance.

The remainder of the paper is structured as follows. Section 2 concentrates on recent literature dealing with the use of ICT in schools. Section 3 provides essential background information on the Italian system of upper-secondary education. Section 4 describes the teaching method whose effectiveness we test. Section 5 presents the experiment design and the empirical framework of our analysis. Section 6 presents our results and investigates the possible mechanisms underlying the effect. Section 7 concludes.

2. Research background

Despite huge interest, the impact of increased availability of ICT infrastructures on students' learning is still uncertain. Some observational studies find that the mere increase in ICT availability has little or no effect on students' achievement (Barrera-Osorio & Linden, 2009; Checchi, Rettore, & Girardi, 2015; Cristia, Ibarrarán, Cueto, & Santiago, 2012; Goolsbee & Guryan, 2006; Gui, Micheli, & Comi, 2018; Leuven, Lindahl, Oosterbeek, & Webbink, 2007), while one finds partial positive effects (Machin, McNally, & Silva, 2007). Recent research based on RCTs underlines that the pure availability of ICT may have detrimental effects on student performance. For example, an RCT involving (a quite small number of) primary school pupils in Norway found that students who read texts in print scored better (on a reading comprehension test) than students who read the texts digitally (Mangen, Walgermo, & Brønnick, 2013). Furthermore, an RCT involving under-graduate students in a US college found that exam scores of students assigned to classrooms that allowed computers and internet connections were lower than those of students in classrooms where computers were prohibited (Payne Carter, Greenberg, & Walker, 2017).

The negative results showed by these studies could be explained by the potential distraction of chatting or web-surfing while in the classroom (Grace-Martin & Gay, 2001; Hembrooke & Gay, 2003; Fried, 2008; Sana, Weston, & Cepeda, 2013), which is made possible by internet connections that are now widely available at schools. A different possible explanation refers to *skim reading*—defined as word spotting and browsing through the text (Ziming, 2005)—which is believed to be the norm in digital reading. When we read on digital supports, our brain reduces time allocated to deep reading resulting in an inability to grasp complexity and greatly reducing our understanding of what we read (Wolf, 2018a, 2018b).

Therefore, the sheer availability of ICT is probably not enough to positively affect students' learning. The role played by teachers, and their skills in integrating ICT into the teaching process, are crucial to understanding whether these technologies have a positive impact on student achievement (UNESCO, 2000). As shown by Comi, Argentin, Gui, Origo, and Pagani (2017), in Italy, teaching practices based on ICT have quite heterogeneous effects on student achievements. In fact, ICT increase pupils' results when they are used to communicate with students, their families, and colleagues, or when they facilitate teaching how to surf social media and blogs. On the contrary, ICT have a negative impact on students' results when they are used to increase students' involvement in classroom.

These results are particularly interesting given the advent of Web 2.0 applications (social networks and micro-blogging), which are characterized by user interaction, collaboration, pervasive

network connectivity, and improved communication channels. These features were not available with the old world wide web (www) technology, when data was posted on web sites, and users could only view or download the content, therefore playing a passive role. Because of the interaction they stimulate, Web 2.0 applications have received wide attention in schools. Teachers particularly use Twitter to create groups of peers that exchange ideas, advice, and resources, and many teachers encourage students to create their own networks so that they can extend their learning experience beyond the classroom (Messner, 2009). Anecdotal evidence shows that educators who use Twitter as a learning tool are enthusiastic. They adopt the platform from the beginning of their students' school career (first- and second-grade students) by keeping class-blogs with children, with the goals of teaching how to write text in real life, how to choose words and punctuation, and how to adapt texts to a space constraint, which forces students to focus ideas. According to its fans, using Twitter as a school device encourages students to consider "literacy (as) the foundation to everything else (we) do in school" (Kurtz, 2009). But the use of Twitter is not limited to primary school. On the contrary, Carpenter and Krutka (2014) hint that Twitter as a learning tool is more common among educators in higher grades.

Teachers use Twitter at school toward two main goals: advancing communication and improving teaching. In their literature review, Carpenter and Krutka (2014) show that according to teachers' opinions, Twitter has the capacity to facilitate communication between students and instructors, thereby enabling relationships that would otherwise be quite difficult (Dunlap & Lowenthal, 2009). For example, students benefit from reminders about class assignments and deadlines (Domizi, 2013) or from in-class instant feedback (Young, 2009; Kassens-Noor, 2012). Furthermore, Twitter is often credited with increasing peer-to-peer interactions and communication (Badge, Johnson, Moseley, & Cann, 2012). However, many scholars emphasize that educators must encourage such a behavior to make the tool effective (Chen & Chen, 2012; Junco, Elavsky, & Heiberger, 2012; Lowe & Laffey, 2011).

When used as a teaching and learning tool, Twitter is credited with increasing students' involvement and satisfaction with courses (Rinaldo, Tapp, & Leverie, 2011; Kruptka, 2014), stimulating engagement and raising grades (Junco, Heiberger, & Loken, 2011), increasing informal learning outside the classroom (Kassens-Noor, 2012), and encouraging concise writing (Dunlap & Lowenthal, 2009).

Twitter has often been used to improve linguistic competences of native- and foreign-language learners (Acar & Kimura, 2012; Antenos-Conforti, 2009; Blattner, Dalola, & Lomicka, 2016; Borau, Ullrich, Feng, & Shen, 2009) and to increase the literary competences of secondary school students (Black, 2009, Bledsoe, 2009; Castrillo de Larreta-Azelain, 2013; Fewell, 2014; Lomicka

and Lord, 2011; Sefton-Green, 2003, 2004). A specific protocol has been developed for using Twitter as a collaborative writing tool for literary analysis in secondary English classroom (McWilliams, Hickey, Hines, Conner, & Bishop, 2011).[2] In the words of the protocol developers, working with Twitter would "emphasize to students that literacy practices for the 21st century involve the production and interpretation of an array of texts, including print and non-print materials . . . , would engage students in digital media practices and would invite their critical, collaborative participation in the process of transforming a textual world into a dynamic online community. In so doing . . . we believed that students would engage in . . . active inquiry and critical thinking[.]" (McWilliams, Hickey, Hines, Conner, & Bishop, 2011, p. 239).

Although teachers' opinion about the use of Twitter to increase students' literacy skills are often enthusiastic (Messner, 2009; Jones, 2011; Miles, 2012), robust evidence about its impact is still lacking. In fact, only a few studies have analyzed the effect of this tool. For example, Stollak, Vandemberg, Burklund, and Weiss (2011) examine grade differences among students at college based on social media usage, and find no robust evidence of a negative relationship between the two variables, but they do not directly focus on the use of Twitter as a teaching tool. Hattem and Lomicka (2016) develop a "meta-analysis" of 17 studies related to language learning, but only provided qualitative answers to the question about the educational benefits of using Twitter. Kassens-Noor (2012) investigates the effect of using Twitter on how students apply, create, and retain knowledge on a sample of 15 students in a midwestern university class, but the study is methodologically flawed given that students self-selected into the treatment or the control group. Vazquez-Cano (2012) investigates the effect of using Twitter on the reading skill of a sample of 280 students attending three high schools in Spain and finds that "Twitter can improve the results in many indicators of linguistic competence, which are fundamental for adequate language performance inside networking and Internet sites, with very good results in normal writing and reading outside Web 2.0 applications" (Vazquez-Cano, 2012, p. 145). Nonetheless, the positive effect is the result of a major flaw in methods given that no treatment and control groups were considered, but rather, only pre- and post-treatment tests.

Junco, Heiberger, and Loken (2011) perform a more rigorous investigation. The authors use an RCT to investigate the effect of using Twitter on engagement and grades of a sample of 125 students taking a first-year seminar course for pre-health professionals. Results "showed that the experimental group had a significantly greater increase in engagement than the control group, as well as higher semester grade point averages" (Junco, Heiberger, & Loken, 2011, p. 119). Although

[2] This protocol is very similar to the one adopted in Italy, whose effectiveness we test with our RCT, and is aimed at producing most of the learning effects mentioned earlier in this paper.

promising, this result is based on a small and extremely specific sample so that its external validity is limited; moreover, its focus is not on language and literature learning.

Given the large and enthusiastic adoption of Twitter for teaching language and literature and the limited number of studies investigating the effects of this approach on high school students, our study makes a new and relevant contribution to the existing literature.

3. Upper secondary education in Italy

In Italy, upper secondary education (ISCED 3) begins when students are about 14 years old and covers grades nine through thirteen. Upper secondary education is free in public schools. These schools enrolled about 96% of the more than 2.6 million students, in 2016/17; private schools play only a marginal role. Italy adopts an "early-track" system of upper secondary education. At the age of fourteen, students are required to choose between three different kinds of high schools: lyceums (*licei*), technical schools (*istituti tecnici*), or vocational schools (*istituti professionali*). Each type of school has a different emphasis. Lyceums emphasize sound training in basic subjects, such as language, math, physics, biology, philosophy and history for students planning to attend tertiary education after grade thirteen. Technical schools focus more on technical subjects (i.e. accounting, electronics, chemistry, etc.) that may help in technical professions and tertiary technical education. Vocational schools forgo some general education subjects to focus on practical skills that can be rapidly sold in the labor market. In 2016/17, about 48.6% of Italian students attended a lyceum, 31.3% attended a technical school, and 20.1% of students attended a vocational school (MIUR, 2018).

In principle, one could move from one kind of school to another but, in practice, this happens only for students who fail to pass a grade at a very early stage of their career and wind up starting from scratch in a less challenging high school. All upper secondary schools are automatically admitted to higher education and university (ISCED 4 and 5). However, most students attending vocational schools (as well as some attending technical ones) do not go on to higher education and enter the labor market right after finishing high school, or even before.

In Italy, attending school is compulsory only until the age of sixteen, when students should have finished grade ten. Consequently, some students drop out before completing high school (grade thirteen). This is a problem in Italy, where, in 2016, 36% of the population aged 25 to 44 has, at the most, a lower secondary education degree (ISCED 2), as compared to 20.5% in the European Union (Eurostat, 2017).

Beside leaving school quite early, on average, Italian students perform quite poorly in international standardized tests. For example, in 2015, 15-year-old Italian students performed worse

than average in reading on the Pisa test (a score of 485, against 493 for the OECD average) (INVALSI, 2015). Although students attending lyceums performed much better (average score of 526) that the OECD average, students in other schools did much worse (473 for technical schools and 410 for vocational schools). Moreover, these results improved only a little over the last 15 years

Given this background, it is no surprise that Italian policymakers have identified a country "reading emergency." The situation spurred the Italian Ministry of Culture to launch a "National plan to promote reading," in 2014, and the Ministry of Education started to emphasize "new digital technologies" to narrow the country's educational gap. In this regard, in 2015, the Ministry launched the National Plan for a Digital School (*Piano nazionale scuola digitale*) and gave financial support to schools to improve their internet connections, provide wi-fi to students and teachers, buy computers and digital interactive whiteboards, and train teachers to use digital devices and technologies. This plan also adopted a "Bring your own device" approach, allowing and encouraging the use of students' personal smartphones and tablets in schools for study and research.

4. The TwLetteratura teaching method

While the public financial investment described in the previous paragraph intended to strengthen the technical infrastructure of Italian schools, less effort has been placed on developing new teaching methods that make the most of new technologies. However, the protocol proposed by the Italian nonprofit organization TwLetteratura[3] (TwL)—already adopted by more than 250 Italian schools—represents a laudable exception and one that, most likely, will soon influence the teaching practices of Italian high school literature teachers. Because it is a standardized protocol, analyzing its effect is both quite simple and very important.

The origin of this protocol goes back to 2012, when TwL started convening—in a virtual environment—hundreds of passionate book readers. Over the years, people used Twitter[4] to read and comment on certain literary masterpieces, such as Macbeth, by William Shakespeare, *Le città invisibili* (Invisible Cities), by Italo Calvino, and *I promessi sposi* (The Betrothed), by Alessandro Manzoni, giving life to the TwL method. This method was introduced in Italian high schools in 2013/14 and, since then, it has been adopted by more than 250 schools and about 14.000 students. The method has several goals: to boost students' interest in reading, to make them more acquainted with literary works, and to improve their school performance in literature courses. In the opinion of its promoters, the TwL method should help students increase their literary skills and, at the same

[3] See www.twletteratura.org for more information.
[4] Starting in 2016, the organization adopted its own app for social reading, called "Betwyll," that can be downloaded from the Apple and Google stores.

time, "instruct an informed practice of new technologies and online social networks." The idea is that using a common micro-blogging platform, such as Twitter, could make the reading of literature masterpieces more attractive for young people, who spend a lot of time surfing social networks but not much time reading books.

The TwL method is simple. The organization TwL (central unit of the project) chooses a book, creates a specific hashtag (such as #Macbetto), and establishes a reading schedule (a chapter a week, for example). The participating schools agree to adopt the same reading schedule. Once a chapter has been read (individually or in class), students can start tweeting their preferred quotes, their comments, or their rewritings of sentences in the chapter, within the usual 140 characters limit.[5] Students use the common hashtag to tweet. They can post as many tweets as they like and can comment on other people's tweets. Teachers, as well as members of the central unit of TwL, take part in these activities to stimulate the whole community of writers.

According to its proponents, the TwL method integrates and enriches the traditional teaching methods used in schools, but is not an alternative to them. The "theory of change" behind this approach predicts that the use of social networks will encourage students to pay more attention to the reading of literary masterpieces when compared to traditional individual reading and studying. Moreover, commenting on or rewriting parts of these literary works in no more than 140 characters should involve an in-depth analysis of the text, thus promoting a greater understanding of its meaning. This practice should also produce a prolific osmosis between spoken and written language. Furthermore, the use of a micro-blogging platform should reduce the stress produced by the traditional study of literature, therefore encouraging the participation of all students (particularly the less diligent ones) and fostering an aptitude for collective problem solving. The TwL method closely resembles the one described by McWilliams, Hickey, Hines, Conner, and Bishop (2011).

The community of teachers adopting the method reached 250 schools in just a few years. Many of these teachers used the TwL approach more than once, in different classes and schools, so that its use quickly spread and will likely continue to spread. Therefore, the TwL approach can be considered as a consolidated method rather than a pilot intervention, even more since it has been mentioned by the European Union report "*Promoting reading in the digital environment*" (EU, 2016).

Although the TwL method is widely accepted by the teachers' community and its use is rapidly increasing, its benefits are not based on solid evidence. Therefore, and thanks to the cooperation of TwL, we independently designed and conducted an RCT to measure the benefits of the method. The

[5] Now extended to 280 characters.

experiment was meant to estimate the impact of the TwL method on students' ability to read, memorize, and understand a literary work, as measured by a standardized test. As mentioned in the Introduction, the literature has examined the use of ICT at school referring almost exclusively to the availability of technological devices, such as personal computers or smartphones, both as part of classroom instruction and by individual students. To our knowledge, this is the first large-scale rigorous test of the effect of using Twitter as a learning tool on students' achievement.

5. Empirical framework

5.1. Design of the experiment

In the spring of 2016, TwL issued a call on its web site for high school teachers to enroll in a new year of activities. The call proposed the adoption of the TwL method to study of "*Il fu Mattia Pascal*" (The Late Mattia Pascal) in 2016/2017. This novel is often read in Italian high schools because its author—Luigi Pirandello, Nobel Laureate in literature in 1934—is considered one of the most important writers in Italian modern literature. A wide group of school teachers from seventy-four different schools spread over the country answered the call and voluntarily enrolled their classes in the TwL activities for the following year.

To test the effectiveness of the method, we designed and performed a delayed treatment cluster-randomized controlled trial (C-RCT). The seventy-four schools were randomly assigned either to the treatment group, which was asked to adopt the TwL method at the beginning of the first term of 2016/2017, or to the control group, which was asked to implement the TwL method starting in the second term of the same school year. The randomization procedure was undertaken by the evaluation team, was concealed from and independent of the project performers, and entailed two steps. First, schools were classified into different strata using three variables: a) a possible predictor of their students' knowledge of "*Il fu Mattia Pascal*" (average school result at a baseline test administered in October 2016; in quartiles), b) the geographic location of the school (North, Center, South and Islands), and c) the number of students involved in the experiment (one class versus more than one class per school). Five of the twenty-four (4 x 3 x 2) resulting strata included no schools; six strata included an odd number of schools (with two strata having one school, and four having three or more schools), while thirteen strata included an even number of schools (see Appendix Table A.1). In each of the four strata including three or a larger odd number of schools, one school was randomly chosen and set apart so the strata would include an even number of organizations. Consequently, we obtained seventeen strata, which included an even number of schools (thirteen plus the new four strata, for a total of sixty-eight schools). In each of these strata, schools were

randomly assigned to the treatment (50%) or to the control group (50%). Hence, each group included thirty-four schools. The six residual schools (the two schools belonging to strata with only one school and the four schools randomly extracted from the strata including an odd number of schools) were grouped into four strata according to their average result at baseline test (see Appendix Table A.2). In each non-empty stratum, schools were randomly assigned to the treatment group or the control group, with the same probability of being assigned to either.

Overall, 37 schools were randomly assigned to the treatment group and invited to adopt the TwL protocol from the beginning of 2016/17, while 37 schools were randomly assigned to the control group and were asked to adopt their traditional approach in reading and studying "*Il fu Mattia Pascal*" during first term. Both groups were requested to read the first ten chapters (out of 20) of the book in November and December 2016.

5.2. Outcome variable and measurement

The outcome variable we use to measure the effectiveness of the TwL method is the number of correct answers to a standardized follow-up test taken by students in late December 2016 and aimed at measuring the command, knowledge, and understanding of chapters one through ten of the book "*Il fu Mattia Pascal*". To get more precise estimates, we use—as the main covariate in our regression models—the individual score at a baseline test administered in October 2016 and aimed at measuring students' command of the Italian language and their understanding of written texts. Both variables have been standardized to have mean of zero and a variance of one so that we can directly show the effect size of our estimates.

A field expert[6] designed the baseline and follow-up tests.[7] The baseline test was designed to measure the literacy level of students, i.e. their capacity to understand what they had read and command of the Italian language. The test asked students to read two short passages (one to two pages long) and to answer 57 multiple choice questions in 45 minutes. The follow-up test was specifically designed to measure the students' ability to memorize, understand, and interpret chapters one through ten of "*Il fu Mattia Pascal"*. It included 106 questions to be answered in 50 minutes.

Both tests were administered online (through the Questbase platform) in the computer room of each school. Students were uniquely identified with an individual anonymous code. Although

[6] The expert is a team member of the *Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione* (INVALSI - National Institute for the Evaluation of the System of Education and Instruction), which is in charge of designing the standardized tests used every year by the Italian Ministry of Education to assess the skills of Italian students of different grades (the INVALSI test).
[7] Both tests are available upon request.

teachers were in the computer room with their students, they did not know in advance either test questions or their answers. This, together with the short time allotted to complete tests, should rule out possible cheating. Moreover, all tests were automatically graded to avoid any impact of an individual teacher's severity (or leniency) in grading. Teachers were provided with individual scores so that they could be used to grade their students. Figure 1 shows the Consort-like diagram of the C-RCT.

As shown in Figure 1, four schools dropped out of the project before the follow-up test was administered. Moreover, some students of the 70 schools that remained in the project did not take both tests. As a result, our analysis includes a total of 70 schools and 1,465 students that took both the baseline and the follow-up test. As far as the schools are concerned, both overall and differential attrition rates are slightly higher than 5%. With respect to students, the overall attrition rate is close to 30%, but differential attrition is close to null. Therefore, attrition is within the safe area suggested by the What Works Clearinghouse (WWC) standards (version 4.0) for RCTs.

5.3. Data and descriptive statistics

To estimate the effect of the TwL teaching method on students' achievement, we use the sample of 1,465 students that took both the baseline and the follow-up test in the experiment (Figure 1).

Considering the quality of the randomization process, Table 1 (Panel A) shows that the treatment and control group schools are well balanced and no difference between the two groups—in the test score at the baseline and in the most relevant covariates—is statistically significant. As to individual students, the two groups show some differences. The treatment group has a larger proportion of students attending schools in the Northern part of the country (as opposed to Central Italy), attending technical and vocational schools (as opposed to lyceums), in 9th or 10th grade (as opposed to being in 11th to 13th grade), and have at least one parent in a high economic position (Table 1, Panel B). These imbalances can be explained by the technical decision to randomize schools rather than classes or students. The characteristics of the TwL method do not allow for randomization of individual students. Moreover, randomization of classes within schools was problematic, given that the same teacher could lecture to more than one class. Nonetheless, the absence of statistically significant differences in the average score of students at the baseline (our most relevant covariate) is encouraging. Moreover, our regression models control for observable covariates, therefore reducing the possible effect of baseline imbalances.

Although our sample is relatively large and incorporates different kinds of schools located in various areas of the country, it cannot be considered as a random sample of the population of Italian schools and students. Consequently, the external validity of the results of this RCT could be

questioned and needs further testing[8]. Nonetheless, it should be considered that although teachers and schools voluntarily enrolled in the project, meaning they self-selected into the RCT, students did not self-select. In fact, if their teachers decided to enter the project, students could not drop out of the experiment.

5.4. Models

We estimate the effect of the TwL method measuring the difference in the average level of the outcome variable for the treated and control students with an OLS regression model with clustered robust standard errors (data grouped at the school level):

$$Y_{ij} = \alpha + \beta T_{ij} + \varepsilon_{ij} \qquad \text{(Eq. 1)}$$

where Y_{ij} represents the score on the follow-up test of student *i* attending school *j*, T_{ij} is a dummy variable that equals 1 for student *i* attending school *j* randomized to the treatment arm of the experiment, and ε_{ij} is the error term. As schools were randomly assigned to the treatment and control groups, estimates of the β coefficient capture the causal effect of the TwL method on the outcome variable.

Furthermore, to test the robustness of our results, we also estimate an OLS regression model that controls for a set of covariates, with clustered robust standard errors (data grouped at the school level):

$$Y_{ij} = \alpha + \beta T_{ij} + \gamma X_{ij} + \delta Z_{ij} + \varepsilon_{ij} \qquad \text{(Eq. 2)}$$

where X_{ij} represents a vector of students' individual characteristics[9], Z_{ij} is a vector of dummy variables measuring school location (North, South, Center) and type (lyceum, technical or vocational).

The models of Eqs. (1) and (2) compare individuals *assigned* to the treatment or the control group. This is known as the *intention-to-treat* (ITT) effect of the intervention (Angrist & Piskhe, 2009). In our case, all students attending a school randomized to the treatment group were assigned to the TwL method and they were not free to make a different choice. Nonetheless, not all of them participated in the experiment with the same intensity given that students could not be forced to tweet. Overall, about 61% of the students posted at least one tweet while the experiment was

[8] See Appendix Table A.3 for a comparison of our sample to the overall population of Italian high schools. The question of external validity of RCTs has been extensively discussed in the literature. See, for example Angrist (2004), Angrist and Pischke (2010), Deaton (2010a, 2010b), Heckman (1992), Heckman and Vytlacil (2007a, 2007b), Hotz, Imbens, and Mortimer (2005), Imbens (2010), and Rothwell (2005).

[9] The vector includes individual performance on the baseline test and a set of dummy variables including: gender, place of birth (Italy versus other countries), grade (9th and 10th grades versus more than 10th grade), parents' place of birth (Italy versus other countries), parents' education (at least one with degree versus no one with degree), parents' profession (at least one with a high-level profession versus no one with a high-level profession), and families' attitude toward books and reading (less than two shelves of books at home versus more than two shelves of books at home).

running, whereas only 27% posted at least five.[10] This behavior originates a one-side-noncompliance situation, which is quite common of RCTs in the social and educational fields because individuals taking advantage of the treatment represent a non-random sample of those assigned to the treatment itself. In this condition, the ITT estimate of the effect of the TwL method could underestimate the value of receiving the treatment. In order to consider this problem, we also estimate the average-treatment-effect (ATT) of the TwL method using an instrumental variable (IV) 2SLS regression model,[11] where random treatment assignment is used as an instrument for actual compliance with the protocol (measured alternatively as posting at least one or at least five tweets) (Eqs. 3 and 4).

6. Results

6.1. Overall effect of TwL

The comparison of students in the treatment arm of the RCT (those attending schools that adopted the TwL method) to students in the control arm of the study (students attending schools that did not adopt the protocol) not only fails to find any positive and statistically significant effect of the method on the outcome variable, but shows negative effects of the intervention on students' achievement, some of which are significant at the usual statistical levels.

Table 2 shows the effect of the TwL method. The ITT estimates of the β coefficients of the models of Eqs. (1) and (2) are always negative, their magnitude is between -0.25 and -0.29 standard deviations, and they are statistically significant at the usual levels. Given the evidence, on average, the TwL method has a detrimental effect on the performance of students assigned to the experiment, compared with the control group. The IV estimates of the β coefficients reported in columns (3) and (4) of Table 2 are coherent with results given in the first two columns.

In Eq. (2), we introduced a series of dummy covariates to increase the precision of the estimates. To check the robustness of our results, we also tested those equations with covariates expressed as categorical or continuous variables. Results are in line with previous estimates: in Eq. (2), the β coefficient moves from -0.25 ($p = 0.034$) to -0.26 ($p = 0.031$) of a standard deviation. We obtain similar results for Eqs. (3) and (4).

[10] Students were asked to tweet using the same anonymous individual code they used to take the baseline and follow-up tests.

[11] Following Angrist and Pischke (2009), we can state that "(i)nstrumental variable methods allow us to capture the causal effect of treatment on the treated in spite of the nonrandom compliance decision made by participants in experiments. . . . . Use of randomly assigned intent to treat as an instrumental variable for treatment delivered eliminates this source of selection bias." (p. 161–162).

6.2. Robustness checks and sensitivity analysis

To further check the robustness of our results, we consider schools (as opposite to individual students) as units of analysis and introduce two more models. First, an OLS regression model that controls for a set of covariates:

$$Y_j = \alpha + \beta T_j + \gamma X_j + \delta Z_j + \varepsilon_j \qquad \text{(Eq. 5)}$$

where Y_j represents the average score of the follow-up test of students attending school *j*; T_j is a dummy variable that equals 1 for schools randomized to the treatment arm of the experiment, X_j represents a vector of dummy variables measuring school location (North, South, Center) and type (lyceum, technical, or vocational), Z_j is a vector of variables with average students' characteristics measured at the school level[12], and ε_j is the error term. As schools were randomly assigned to the treatment and control groups, estimates of the β coefficient capture the causal effect of the TwL method on our outcome variable. Moreover, we estimate the same model weighting school data with the number of students involved in the experiment in each school (Eq. 6).

Results of these new estimates—in Appendix Table A.4—are coherent with those reported so far. In fact, they indicate that, on average, the performance of schools adopting the method is about 16—27% of a standard deviation lower than the performance of schools adopting the customary methods.[13] These results confirm that, on average, the TwL method has a relevant negative impact on students' knowledge and understanding of the book *"Il fu Mattia Pascal"*.

Furthermore, we also control for the sensitivity of our results to different definitions of the dependent variable. The follow-up test included two types of questions: 1) those related to book memorization (i.e. remembering the main characters, what they do and say, etc.), and 2) those related to understanding the meaning of the book. We checked the effect of the TwL method on both sets of questions. Table A.5 shows the results of our analysis using the model proposed in Eq. (1). The TwL method adversely affects both outcome variables even though the effect looks stronger on memorization than understanding.

[12] The vector includes the school average performance at the baseline test and a set of dummy variables including: gender, place of birth (Italy versus other countries), grade (9th and 10th grades versus more than 10th grade), parents' place of birth (Italy versus other countries), parents' education (at least one with degree versus no one with a degree), parents' profession (at least one with a high-level profession versus no one with a high-level profession), and families' attitude toward books and reading (less than two shelves of books at home versus more than two shelves of books).

[13] The coefficient of Eq. 5 is not statistically significant at the usual levels. Statistical insignificance is not unexpected given the relatively small number of schools involved in the experiment. Nonetheless, when using the more appropriate model of Eq. (6) (with data weighted for school size), the estimated β coefficient becomes statistically significant at the usual levels, and very close in size to the estimates of Eqs. (1) and (2).

6.3. Effect by subgroup

Table 3 explores the effect of TwL on specific subgroups of the population involved in the experiment and shows that the average negative effect described in paragraph 6.1 is somewhat stronger in some of these groups. In particular, we run heterogeneity analysis interacting the dummy variable indicating treatment with the dummy variables representing students' gender, place of birth (Italy or elsewhere), type of school attended (lyceums versus other schools), and students' skills in literature (above or below median on baseline test). All estimates are based on Eq. (2).

No subgroup shows positive effects of the TwL method. On the contrary, Table 3 suggests that adopting the TwL method is more detrimental to female students than male students and to students with higher scores on the baseline test. For these subgroups, the score on the follow-up test of students in the treatment group appears to be about one-third of a standard deviation lower than the performance of students in the control group. These differences are only significant at the 10% level, so that they could represent chance findings. Nonetheless, they go together with the further evidence (not statistically significant) showing that the TwL method is more detrimental to students born in Italy and students attending lyceums than their counterparts (students born abroad and students attending technical or vocational schools, respectively).

Together, these results suggest that TwL appears to harm the most-talented students. In fact, in Italy, females students, students born in the country, and students attending lyceums, generally perform better than their respective counterparts. This is a very relevant evidence, although a suggestive one, given that TwL claims that teaching students to use ICT devices and applications should particularly benefit the less-talented ones.

6.4. Discussion and possible mechanism

Given this evidence, we could say that, on average, our estimates show that the TwL method has a detrimental effect on the performance of students involved in the experiment. The size of the effect (in absolute value) is relatively large when compared to other interventions aimed at increasing students' skills. As mentioned earlier, previous rigorous literature on projects using micro-blogs as teaching tools is virtually non-existent. For this reason, we must refer to other interventions aimed at increasing students' skills.

In their meta-analysis of the effects of more than 120 early education interventions on cognitive development of children, Camilli, Vargas, Ryan, and Barnett (2010) find that the average effect size of "treatment" to "alternative treatment" interventions is about 7% of a standard deviation, while it rises to 23% of a standard deviation when considering a treatment to control setting. In a more recent and rigorous paper, Lortie-Forgues and Inglis (2019) analyze more than 140 educational

trials—commissioned by the Education Endowment Foundation (EEF) in the UK and by the National Center for Educational Evaluation and Regional Assistance (NCEE) in the USA—aimed at improving achievement in the K-12. The authors find a mean effect size of 6% of a standard deviation. Given these evidence, the impact of the intervention we estimate (albeit negative) is quite significant.

Our research strategy does not directly indicate any possible explanations for our results. Nonetheless, it is quite clear that participating in the experiment gave some unexpected (and adverse) incentives to students. In addition to the adverse effects already addressed by the literature, such as *skim reading* (Ziming, 2005) and reading on electronic devices (smartphones, in our case) (Mangen, Walgermo, & Brønnick, 2013), one could wonder whether a further mechanism is at work. It could be that reading several short messages (tweets) coming from peers and teachers could have given students the false impression that they had actually read and adequately studied the assigned chapters. Consequently, they may have devoted less time to careful reading and comprehension of the book (a costly activity) and more time to other preferred activities, such as surfing the net. The TwL approach could have therefore appeared as a convenient short-cut to learning about the book. A suggestive, but certainly not conclusive, evidence in this direction show that students belonging to the treatment group have a higher and statistically significantly probability (slightly more than 2 percent)[14] to use internet to perform school tasks, such as talking to peers about school assignments, looking for information on the internet, etc.. This could hint that treated students surf the net more than control ones.

This "short-cut effect" could have had no impact on the performance of less committed and engaged students (who generally devote a small amount of time to schoolwork), but it may have significantly reduced the effort of the most talented and dedicated ones, therefore greatly damaging their performance.

In principal, the adoption of the TwL method could influence also teachers' performance. In fact, knowing that students will somehow be assisted (by their own peers and by the central unit of TwL) could induce teachers to reduce their teaching effort and therefore reducing their explanations and deeper investigations of the subject during class hours. We do not have any solid evidence on this issue, but the answers given by 73 teachers (about 60% of those involved in the intervention) to a qualitative survey aimed at describing how the intervention was implemented in schools and at assessing teachers' satisfaction could give some hints. This qualitative investigation shows that

[14] This is the coefficient of a regression having as dependent variable a dummy that measures the use of internet to perform school tasks and as main independent variable the treatment dummy variable. All usual covariates are included in the regression. Information about the use of internet to perform school tasks was collected through the follow-up questionnaire.

more than 22% of teachers in the treatment group declared that they never read parts of the book in class with their students, compared to 16% of teachers in the control group. Furthermore, 56% of teachers of the treatment group never tested students on the book (if we exclude the follow-up test), compared to 35% of the control group. Even though none of the differences is statistically significant, one could get the impression that teachers may have reduced their teaching effort during the experiment.

7. Conclusions

Over the last several years, schools have greatly increased their use of technological devices, software, and connections to the internet. In particular, the widespread availability of cheap wi-fi connections to the internet has stimulated the adoption of Web 2.0 applications to the teaching environment. The Web 2.0 platform, as opposed to the old www technologies, allows intense interaction between users and between users and content producers. This feature has induced many teachers to adopt these platforms to stimulate students' participation in school activities, and anecdotical evidence reports that teachers are highly satisfied with the change. Nonetheless, only a small number of studies have produced rigorous estimates of the effects of these platforms on students' achievement.

Our paper provides new and relevant contributions to the literature in this field. We run a large-scale RCT to analyze the impact of using Twitter as a literature teaching tool on the reading and comprehension skills of students. Our sample includes 1,465 students attending 70 Italian high schools in 2016/17. The randomly chosen treatment group adopted Twitter to read, discuss, and study the book "*Il fu Mattia Pascal*" written by Italian Nobel Laureate Luigi Pirandello, while the control group read the same book using the traditional teaching approach.

We find that using Twitter to teach literature has an overall detrimental effect on students' average achievement: it reduces performance on a standardized test score by about 25–40% of a standard deviation. The negative effect is heterogeneous with respect to students' characteristics. Female students, students born in Italy, students attending lyceums, and students with higher score on baseline test are more harmed than their counterparts. Therefore, the use of this Web 2.0 application appears to have stronger detrimental effects on students that usually perform better.

Although more research is needed to test the external validity of our results, this is, to our knowledge, the widest and most rigorous study of the effect of adopting Web 2.0 platforms for teaching literature on students' performance.

Acknowledgements

We wish to thank G. Argentin, D. Besharov, G. Brunello, A. Martini, B. Romano, V. Rotondi, J. Smith, G. Sorrenti, C. Torgerson, D. Torgerson, G. Turati, the participants of the Twelfth Annual Conference on RCTs in the Social Sciences (York University), and the participants of the seminar in honor of A. Martini (Collegio Carlo Alberto, Torino) for helpful comments. Usual disclaimers apply.

Financial support from the Cariplo Foundation (Milano, Italy) is gratefully acknowledged.

References

Acar, A., & Kimura, N, (2012). Twitter as a tool for language learning: The case of Japanese learners of English. In: Charmonman, S. (ed) *The eighth international conference on e-learning for knowledge-based society*. Bangkok, Thailand: Siam Technology Press. Vol. 20, 1–14.

Angrist, J. (2004). Treatment effect heterogeneity in theory and practice. *The Economic Journal*, 114(494), C52-C83.

Angrist, J., & Pischke, J.S. (2009). *Mostly harmless econometrics. An empiricist's companion*. Princeton, NJ: Princeton University Press.

Angrist, J., & Pischke, J.S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2), 3–30.

Antenos-Conforti, E. (2009). Micro-blogging on Twitter: Social networking in intermediate Italian classes. In: Lomicka, L. & Lord, G. (eds) *The Next Generation: Social Networking and Online Collaboration in Foreign Language Learning.* San Marcos, TX: Computer Assisted Language Instruction Consortium, 59–90.

Badge, J., Johnson, S., Moseley, A., & Cann, A. (2012). Observing emerging student networks on a micro-blogging service. *Journal of Online Learning and Teaching*, 7(1), 90–98.

Barbetta, G.P., Sorrenti, G., & Turati, G. (2019). *Multigrading and child achievement*. University of Zurich, Working paper series, No. 275.

Barrera-Osorio, F., & Linden, L.L. (2009). *The use and misuse of computers in education: Evidence from a randomized experiment in Colombia*. The World Bank Policy Research Working Paper Series 4836.

Black, R.W. (2009). English-language learning, fan communities, and 21st-century skills. Journal of Adolescent & Adult Literacy, 52(8), 688–697.

Blattner, G., Dalola, A., & Lomicka, L. (2015). Tweetsmarts: A pragmatic analysis of well-known native speaker Tweeters. In: Dixon, E., & Thomas, M. (eds) *Researching language learner interactions online: From social media to MOOCs*. San Marcos, TX: Computer Assisted Language Instruction Consortium, 213–236.

Bledsoe, G. (2009). Collaborative Digital Writing: The Art of Writing Together Using Technology. In: Herrington, A., Hodgson, K., & Moran, C. (eds). *Teaching the New Writing: Technology, Change, and Assessment in the 21st-Century Classroom*. New York: Teachers College Press, 39–54.

Borau K., Ullrich, K., Feng J.,& Shen R. (2009) Micro-blogging for language learning: Using Twitter to train communicative and cultural competence, *Advances in Web Based Learning – ICWL 2009*, 78–87.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W.S. (2010). Meta-Analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, 112(3), 579–620.

Carpenter, J.P., & Krutka, D.G. (2014). How and why educators use Twitter: A survey of the field. *Journal of Research on Technology in Education*, 46(4) 414–434.

Castrillo de Larreta-Azelain, D. (2013). Learners' attitude toward collaborative writing in e-language learning classes: A Twitter project for German as a foreign language. *Revista Española de Linguistica Aplicada*, 26, 127–138.

Checchi, D., Rettore, E., & Girardi, S. (2015). ICT technology and learning: An impact evaluation of Cl@ssi2.0. IZA Discussion paper 8986.

Chen, L., & Chen, T.L. (2012). Use of Twitter for formative evaluation: Reflections on trainer and trainees' experiences. *British Journal of Educational Technology*, 43(2), E49–E52.

Comi, S., Argentin, G., Gui, M., Origo, F., & Pagani, L. (2017). Is it the way they use it? Teachers, ICT and student achievement. *Economics of Education Review*, 56, 24–39.

Cristia, J. P., Ibarrarán, P., Cueto, S., Santiago, A., & Severín, E. (2012). Technology and child development: Evidence from the one laptop per child program. IZA Discussion Papers 6401.

Deaton, A. (2010a). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48(2), 424–455.

Deaton, A. (2010b). Understanding the mechanisms of economic development. *Journal of Economic Perspectives*, 24(3), 3–16.

Domizi, D.P. (2013). Micro-blogging to foster connections and community in a weekly graduate seminar course. *TechTrends*, 57(1), 43–51.

Duncan, C.J., Morris, P.A., & Rodriguez, C. (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental Psychology*, 47, 1263–1279.

Dunlap, J.C., & Lowenthal, P.R. (2009). Tweeting the night away: Using Twitter to enhance social presence. *Journal of Information Systems Education*, 20(2) 129–136.

Ebner M., Lienhardt C., Rohs M. & Meyer I. (2010). Micro-blogs in higher education – a chance to facilitate informal and process-oriented learning? *Computers and Education* 55, 92–100.

European Union (2016). *Promoting reading in the digital environment. Report of the working group of EU member states' experts on promoting reading in the digital environment under the open method of coordination*. Luxembourg: Publications Office of the European Union.

Fewell, N. (2014). Social networking and language learning with Twitter. *Research Papers in Language Teaching and Learning*, 5, 223–234.

Fried, C.B. (2008). In-class laptop use and its effects on student learning. *Computers and Education*, 50 (3), 906–914.

Gao, F., Luo, T., & Zhang, K. (2012). Tweeting for learning: A critical analysis of research on micro-blogging in education published in 2008–2011. *British Journal of Educational Technology*, 43(5), 783–801.

Goolsbee, A., & Guryan, J. (2006). The impact of internet subsidies in public schools. *Review of Economics and Statistics*, 88(2), 336–347.

Grace-Martin, M. & Gay, G. (2001). Web browsing, mobile computing and academic performance. *Journal of Educational Technology & Society*, 4(3), 95–107

Greenhow, C., & Gleason, B. (2012). Twitteracy: Tweeting as a new literacy practice. *The Educational Forum*,76(4), 464–478.

Gui, M., Parma, A., & Comi, S. (2018). Does public investment in ICTs improve learning performance? Evidence from Italy. *Policy & Internet*, 10(2), 141–163.

Hattem, D., & Lomicka, L. (2016). What the tweets say: A critical analysis of Twitter research in language learning from 2009 to 2016, *E-Learning and Digital Media*, 13(1–2), 5–23.

Heckman, J. (1992). Randomization and social policy evaluation. In: Manski, C. & Garfinkel, I. (eds), *Evaluating welfare and training programs*. Cambridge, MA: Harvard Univ. Press, 201–230.

Heckman, J., & Vytlacil, E. (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In: Heckman, J., & Leamer, E. (eds), *Handbook of Econometrics, Vol. 6B.* Amsterdam, The Netherland: Elsevier, 4779–4874.

Heckman, J., & Vytlacil, E. (2007b. Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In: Heckman, J., & Leamer, E. (eds), *Handbook of Econometrics, Vol. 6B.* Amsterdam, The Netherland: Elsevier, 4875–5144.

Hembrooke, H. & Gay, G. (2003). The laptop and the lecture: The effects of multitasking in learning environments. *Journal of Computing in Higher Education*, 15(1), 46–64.

Hotz, J., Imbens, G., & Mortimer, J. (2005). Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics*, 125(1-2), 241–270.

Jones, A. (2011). How Twitter Saved my Literature Class: A Case Study with Discussion. In: Wankel, C. (ed.) *Teaching Arts and Science with the New Social Media*. Bingley, United Kingdom: Emerald Group Publishing, 91-106.

Junco, R., Elavsky, C.M., & Heiberger, G. (2012). Putting Twitter to the test: Assessing outcomes for student collaboration, engagement and success. *British Journal of Educational Technology*, 44(2), 273–287.

Junco, R., Heiberger, G., & Loken, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning*, 27(2), 119–132.

Kassens-Noor, E. (2012). Twitter as a teaching practice to enhance active and informal learning in higher education: The case of sustainable tweets. *Active Learning in Higher Education*, 13(1), 9–21.

Kern, R. (1995). Restructuring classroom interaction with networked computers: Effects on quantity and characteristics of language production. *The Modern Language Journal*, 79(4), 457–476.

Krutka, D. G. (2014). Social media as a catalyst for convergence culture: Immersing pre-service social studies teachers in the social media terrain. In Russell, W.B. (ed), *Digital Social Studies*. Charlotte, NC: Information Age Publishing, 271–302.

Kurz, J. (2009). Twittering about learning: Using Twitter in an elementary school classroom. *Horace Summer*, 25(1). Retrieved from http://www.essentialschools.org/cs/cespr/view/ces_res/618.

Imbens, G. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature*, 48(2), 399–423.

INVALSI, 2015. *Indagine Ocse Pisa 2015: i risultati degli studenti italiani in scienze, matematica e lettura*. Roma, Italy: INVALSI.

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *Review of Economics and Statistics*, 89(4), 721–736.

Leuven, E., & Rønning, M. (2016). Classroom grade composition and pupil achievement. *The Economic Journal*, 126(593), 1164–1192.

Lomicka, L., & Lord, G. (2011). A tale of tweets: Analyzing micro-blogging among language learners. *System*, 40(1), 48–63.

Lortie-Forgues, H., and Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158–166.

Lowe, B., & Laffey, D. (2011). Is Twitter for the birds? Using Twitter to enhance student learning in a marketing course. *Journal of Marketing Education*, 33(2), 183–192.

Machin, S., McNally, S., & Silva, O. (2007). New technology in schools: Is there a payoff? *Economic Journal*, 117(522), 1145–1167.

Mangen, A., Walgermo, B.R., & Brønnick, K. (2013). Reading linear text on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, 58, 61–68.

McWilliams, J., Hickey, D.T., Hines, M. B., Conner, J. M., & Bishop, S.C. (2011). Using collaborative writing tools for literary analysis: Twitter, fan-fiction, and The Crucible in the secondary English classroom. *Journal of Media and Literacy Education*, 2(3), 238–245.

Messner, K. (2009). Pleased to tweet you: Making a case for Twitter in the classroom. *School Library Journal*, 55(12), 44–47.

Miles, R. (2012). Teaching with Twitter: how the social network can contribute to learning. *The Guardian*, 11/6/2012.

Miur, 2018. *Portale Unico dei Dati della Scuola*. http://dati.istruzione.it/opendata/esploraidati/

Payne Carter, S., Greenberg, K., & Walker M.S. (2017). The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy, *Economics of Education Review*, 56, 118–132.

Rinaldo, S. B., Tapp, S., & Laverie, D. A. (2011). Learning by tweeting: Using Twitter as a pedagogical tool. *Journal of Marketing Education*, 33(2), 194–204.

Rothwell, P. (2005). External validity of randomised controlled trials: To whom do the results of this trial apply? *The Lancet*, 365, 82–93.

Sana, F., Weston, T., & Cepeda, N.J. (2013). Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education*, 62, 24–31.

Sefton-Green, J. (2003). Informal learning: Substance or style? *Teaching Education*, 12, 37–52.

Sefton-Green, J. (2004). *Literature Review in Informal Learning with Technology Outside School*. A NESTA Futurelab Series – report 7.

Stollak, M., Vandenberg, A., Burklund, A., & Weiss, S. (2011). Getting social: The impact of social networking usage on grades amongst college students. Proceedings from ASBBS Annual Conference, 18 859–865.

UNESCO (2000). Dakar framework for action. World Economic Forum.

Vázquez-Cano, E. (2012). Mobile learning with Twitter to improve linguistic competence at Secondary Schools. *The New Educational Review, 29*(3), 134–147.

Veletsianos, G. (2012). Higher education scholars' participation and practices on Twitter. *Journal of Computer Assisted Learning*, 28(4), 336–349.

WWC What Works Clearinghouse (n.a.). *Standards handbook (version 4.0)*. Retrieved from https://ies.ed.gov/ncee/wwc/handbooks.

Wolf, M. (2018a). *Reader, Come Home: The Reading Brain in a Digital World*, New York, NY: HarperCollins.

Wolf, M. (2018b). Skim reading is the new normal. *The Guardian*, 27/8/2018.

Young, J. (2009). Teaching with Twitter: Not for the faint of heart. *Chronicle of Higher Education*, 22/9/2009.

Ziming, L. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation*, 61(6), 700–712.

Figures and tables

Figure 1

Consort-like diagram of the C-RCT

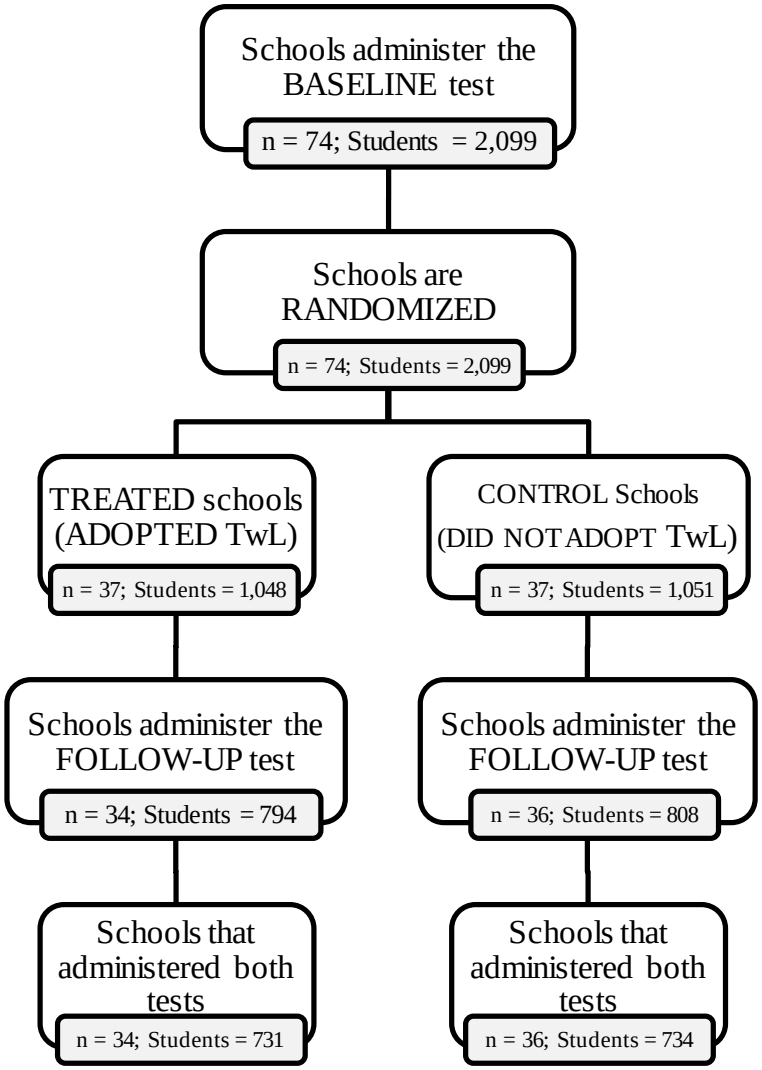

Table 1

Summary statistics and covariate balance

	Control	Treatment	Control vs. Treatment	Control	Treatment	Control vs. Treatment
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline characteristic	**PANEL A: school averages**			**PANEL B: students**		
Score on baseline test	-0.161	.025	-0.186	.017	-0.017	-0.033
	(.137)	(.109)	(.176)	(.039)	(.035)	(.052)
Female	0.604	0.601	.003	0.567	0.562	.005
			(.061)			(.026)
Italian	.947	.899	.048	.952	.947	.006
			(.035)			(.011)
At least one parent Italian	.928	.923	.004	.933	.948	-0.015
			(.026)			(.012)
North	.417	.441	-0.024	.300	.395	-0.096***
			(.120)			(.025)
Center	.139	.059	.080	.128	.071	.057***
			(.072)			(.016)
South	.444	.5	-0.056	.572	.533	.039
			(.121)			(.026)
Lyceum	.625	.647	-0.022	.764	.655	.109***
			(.116)			(.024)
Technical School	.263	.265	-0.002	.191	.265	-0.075***
			(.106)			(.022)
Vocational School	.111	.088	.023	.045	.079	-0.034***
			(.073)			(.013)
11th through 13th grade	.491	.365	.126	.471	.262	.209***
			(.116)			(0.02)
At least one parent with tertiary education	.304	.306	-0.002	.327	.367	-0.040
			(.052)			(.025)
Parents with high profession	.330	.363	-0.033	.347	.406	-0.059**
			(.043)			(.025)
Many books at home	.477	.444	.033	.480	.487	-0.007
			(.058)			(.026)
Observations	36	34		734	731	

Notes:

Columns 1–2 report, respectively, average characteristics of schools in the control and the treatment groups.

Column 3 reports coefficient estimates from a regression of the different characteristics (measured at baseline) on a dummy variable that equals one if a school is assigned to the treatment group. Standard errors are reported in brackets.

Columns 4–5 report, respectively, average characteristics of students in the control and the treatment groups.

Column 6 reports coefficient estimates from a regression of the different characteristics (measured at baseline) on a dummy variable that equals one if a student is assigned to the treatment group. Standard errors clustered at the school level are reported in brackets.

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 2

Effect of the TwL method

Dependent variable:	**Equation:**			
score on follow-up test	(1)	(2)	(3)	(4)
Students of schools adopting the TwL method	-0.287*	-0.250**	-0.414**	-1.133*
	(.163)	(.116)	(.203)	(.610)
Controls for student's characteristics		X	X	X
Controls for school's characteristics		X	X	X
Observations	1,465	1,465	1,465	1,465

Notes:

This table reports estimates from a regression of follow-up test scores on a dummy variable that equals 1 for treated students and 0 for control students. Robust standard errors are clustered at the school level and reported in brackets.

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3

Heterogenous treatment effects

Dependent variable: score on follow-up test			
Subgroup:			
Gender	Females	Males	Difference
	-0.359***	-0.107	-0.251*
	(.126)	(.142)	(.143)
Place of birth	Italy	Abroad	Difference
	-0.258**	-0.093	-0.166
	(.116)	(.218)	(.186)
Type of school	Lyceum	Other school	Difference
	-0.346***	0.016	-0.361
	(.129)	(.237)	(.270)
Score on baseline test	Above median	Below median	Difference
	-0.353***	-0.147	-0.206*
	(.123)	(.135)	(.123)
Observations		1,465	

Notes:

This table reports estimates from a regression of follow-up test scores on a dummy variable that equals 1 for treated students and 0 for control students interacted with different covariates representing subgroups. Estimates use the model identified by Eq. 5. Robust standard errors are clustered at the school level and reported in brackets.

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

APPENDIX 1

Table A.1
First stratification of schools

One class involved				
Average performance on baseline test	***North***	***Center***	***South & Islands***	***Total***
First quartile	6	2	7	**15**
Second quartile	5	-	2	**7**
Third quartile	4	-	2	**6**
Fourth quartile	7	2	4	**13**
More than one class involved				
Average performance on baseline test	***North***	***Center***	***South & Islands***	***Total***
First quartile	3	-	2	**5**
Second quartile	6	1	6	**13**
Third quartile	-	2	8	**10**
Fourth quartile	-	1	4	**5**
Total	**31**	**8**	**35**	**74**

Table A.2
Second stratification of schools

Average performance on baseline test	
First quartile	2
Second quartile	2
Third quartile	-
Fourth quartile	2
Total	**6**

Table A.3
High school students (school year 2016/2017)

	Lyceums	Technical Schools	Vocational Schools	Total (n)
		A. Italy		
North	18.6%	13.8%	8.2%	1,064,677
Center	10.6%	5.4%	3.6%	514,508
South & Islands	19.4%	12.1%	8.3%	1,040,265
Total (n)	*1,271,801*	*821,078*	*526,571*	*2,619,450*
		B. RCT sample		
North	16.2%	14.8%	3.7%	509
Center	9.4%	-	0.6%	146
South & Islands	45.4%	8.0%	1.9%	810
Total (n)	*1,040*	*334*	*91*	*1,465*

Table A.3 compares the sample of students and schools involved in the TwL experiment with the overall population of Italian high schools. Looking at Panel A and Panel B of the Table, one can see that our sample has a higher proportion (compared to the population) of students attending lyceums in schools located in the southern part of Italy. The sample has lower proportions of students attending professional schools and living in the central part of the country. This should not be a major problem, given that we do not have a clear theory about the possible differential effect of the TwL method on students living in different parts of the country or attending different kinds of schools. The only exception is that TwL expects a larger effect on students with lower school performance, those more prone to surf the internet, and those less willing to read books. In Italy. these students generally prefer vocational schools to technical schools or to lyceums. Consequently, the small number of these students in our sample may hinder the identification of heterogeneous effects.

Table A.4
Effect of the TwL method

Dependant variable:	**Equation:**	
average score on follow-up test	(5)	(6)
Schools adopting the TwL method	-0.156	-0.271**
	(.169)	(.137)
Controls for schools characteristics	X	X
Controls for average student's characteristics in the school	X	X
Weight: number of students		X
Observations	70	70

Notes:
This table reports estimates from a regression of follow-up test scores on a dummy variable that equals 1 for treated students and 0 for control students. Robust standard errors are clustered at the school level and reported in brackets.
*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.5
Sensitivity analysis

Dependent variable: score on follow-up test	**Memorizing**	**Understanding**
Students adopting TwL	-0.270**	-0.191*
	(.117)	(.114)
Controls for school characteristics	X	X
Controls for students characteristics	X	X
Observations	1,465	1,465

Notes:
Robust standard errors are clustered at the school level and reported in brackets. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Working Paper del Dipartimento di Economia e Finanza

1. L. Colombo, H. Dawid, *Strategic Location Choice under Dynamic Oligopolistic Competition and Spillovers*, novembre 2013.
2. M. Bordignon, M. Gamalerio, G. Turati, *Decentralization, Vertical Fiscal Imbalance, and Political Selection*, novembre 2013.
3. M. Guerini, *Is the Friedman Rule Stabilizing? Some Unpleasant Results in a Heterogeneous Expectations Framework*, novembre 2013.
4. E. Brenna, C. Di Novi, *Is caring for elderly parents detrimental to women's mental health? The influence of the European North-South gradient*, novembre 2013.
5. F. Sobbrio, *Citizen-Editors' Endogenous Information Acquisition and News Accuracy*, novembre 2013.
6. P. Bingley, L. Cappellari, *Correlation of Brothers Earnings and Intergenerational Transmission*, novembre 2013.
7. T. Assenza, W. A. Brock, C. H. Hommes, *Animal Spirits, Heterogeneous Expectations and the Emergence of Booms and Busts*, dicembre 2013.
8. D. Parisi, *Is There Room for 'Fear' as a Human Passion in the Work by Adam Smith?*, gennaio 2014.
9. E. Brenna, F. Spandonaro, *Does federalism induce patients' mobility across regions? Evidence from the Italian experience*, febbraio 2014.
10. A. Monticini, F. Ravazzolo, *Forecasting the intraday market price of money*, febbraio 2014.
11. Tiziana Assenza, Jakob Grazzini, Cars Hommes, Domenico Massaro, *PQ Strategies in Monopolistic Competition: Some Insights from the Lab*, marzo 2014.
12. R. Davidson, A. Monticini, *Heteroskedasticity-and-Autocorrelation-Consistent Bootstrapping*, marzo 2014.
13. C. Lucifora, S. Moriconi, *Policy Myopia and Labour Market Institutions*, giugno 2014.
14. N. Pecora, A. Spelta, *Shareholding Network in the Euro Area Banking Market*, giugno 2014.
15. G. Mazzolini, *The economic consequences of accidents at work*, giugno 2014.
16. M. Ambrosanio, P. Balduzzi, M. Bordignon, *Economic crisis and fiscal federalism in Italy*, settembre 2014.
17. P. Bingley, L. Cappellari, K. Tatsiramos, *Family, Community and Long-Term Earnings Inequality*, ottobre 2014.
18. S. Frazzoni, M. L. Mancusi, Z. Rotondi, M. Sobrero, A. Vezzulli, *Innovation and export in SMEs: the role of relationship banking*, novembre 2014.
19. H. Gnutzmann, *Price Discrimination in Asymmetric Industries: Implications for Competition and Welfare*, novembre 2014.
20. A. Baglioni, A. Boitani, M. Bordignon, *Labor mobility and fiscal policy in a currency union*, novembre 2014.
21. C. Nielsen, *Rational Overconfidence and Social Security*, dicembre 2014.
22. M. Kurz, M. Motolese, G. Piccillo, H. Wu, *Monetary Policy with Diverse Private Expectations*, febbraio 2015.
23. S. Piccolo, P. Tedeschi, G. Ursino, *How Limiting Deceptive Practices Harms Consumers*, maggio 2015.
24. A.K.S. Chand, S. Currarini, G. Ursino, *Cheap Talk with Correlated Signals*, maggio 2015.
25. S. Piccolo, P. Tedeschi, G. Ursino, *Deceptive Advertising with Rational Buyers*, giugno 2015.

26. S. Piccolo, E. Tarantino, G. Ursino, *The Value of Transparency in Multidivisional Firms*, giugno 2015.
27. G. Ursino, *Supply Chain Control: a Theory of Vertical Integration*, giugno 2015.
28. I. Aldasoro, D. Delli Gatti, E. Faia, *Bank Networks: Contagion, Systemic Risk and Prudential Policy*, luglio 2015.
29. S. Moriconi, G. Peri, *Country-Specific Preferences and Employment Rates in Europe*, settembre 2015.
30. R. Crinò, L. Ogliari, *Financial Frictions, Product Quality, and International Trade*, settembre 2015.
31. J. Grazzini, A. Spelta, *An empirical analysis of the global input-output network and its evolution*, ottobre 2015.
32. L. Cappellari, A. Di Paolo, *Bilingual Schooling and Earnings: Evidence from a Language-in-Education Reform*, novembre 2015.
33. A. Litina, S. Moriconi, S. Zanaj, *The Cultural Transmission of Environmental Preferences: Evidence from International Migration*, novembre 2015.
34. S. Moriconi, P. M. Picard, S. Zanaj, *Commodity Taxation and Regulatory Competition*, novembre 2015.
35. M. Bordignon, V. Grembi, S. Piazza, *Who do you blame in local finance? An analysis of municipal financing in Italy*, dicembre 2015.
36. A. Spelta, *A unified view of systemic risk: detecting SIFIs and forecasting the financial cycle via EWSs*, gennaio 2016.
37. N. Pecora, A. Spelta, *Discovering SIFIs in interbank communities*, febbraio 2016.
38. M. Botta, L. Colombo, *Macroeconomic and Institutional Determinants of Capital Structure Decisions,* aprile 2016.
39. A. Gamba, G. Immordino, S. Piccolo, *Organized Crime and the Bright Side of Subversion of Law*, maggio 2016.
40. L. Corno, N. Hildebrandt, A. Voena, *Weather Shocks, Age of Marriage and the Direction of Marriage Payments*, maggio 2016.
41. A. Spelta, *Stock prices prediction via tensor decomposition and links forecast*, maggio 2016.
42. T. Assenza, D. Delli Gatti, J. Grazzini, G. Ricchiuti, *Heterogeneous Firms and International Trade: The role of productivity and financial fragility*, giugno 2016.
43. S. Moriconi, *Taxation, industry integration and production efficiency*, giugno 2016.
44. L. Fiorito, C. Orsi, *Survival Value and a Robust, Practical, Joyless Individualism: Thomas Nixon Carver, Social Justice, and Eugenics*, luglio 2016.
45. E. Cottini, P. Ghinetti, *Employment insecurity and employees' health in Denmark*, settembre 2016.
46. G. Cecere, N. Corrocher, M. L. Mancusi, *Financial constraints and public funding for eco-innovation: Empirical evidence on European SMEs*, settembre 2016.
47. E. Brenna, L. Gitto, *Financing elderly care in Italy and Europe. Is there a common vision?*, settembre 2016.
48. D. G. C. Britto, *Unemployment Insurance and the Duration of Employment: Theory and Evidence from a Regression Kink Design*, settembre 2016.
49. E. Caroli, C.Lucifora, D. Vigani, *Is there a Retirement-Health Care utilization puzzle? Evidence from SHARE data in Europe*, ottobre 2016.
50. G. Femminis, *From simple growth to numerical simulations: A primer in dynamic programming*, ottobre 2016.
51. C. Lucifora, M. Tonello, *Monitoring and sanctioning cheating at school: What works? Evidence from a national evaluation program*, ottobre 2016.

52. A. Baglioni, M. Esposito, *Modigliani-Miller Doesn't Hold in a "Bailinable" World: A New Capital Structure to Reduce the Banks' Funding Cost*, novembre 2016.
53. L. Cappellari, P. Castelnovo, D. Checchi, M. Leonardi, *Skilled or educated? Educational reforms, human capital and earnings*, novembre 2016.
54. D. Britto, S. Fiorin, *Corruption and Legislature Size: Evidence from Brazil*, dicembre 2016.
55. F. Andreoli, E. Peluso, *So close yet so unequal: Reconsidering spatial inequality in U.S. cities*, febbraio 2017.
56. E. Cottini, P. Ghinetti, *Is it the way you live or the job you have? Health effects of lifestyles and working conditions*, marzo 2017.
57. A. Albanese, L. Cappellari, M. Leonardi, *The Effects of Youth Labor Market Reforms: Evidence from Italian Apprenticeships*; maggio 2017.
58. S. Perdichizzi, *Estimating Fiscal multipliers in the Eurozone. A Nonlinear Panel Data Approach*, maggio 2017.
59. S. Perdichizzi, *The impact of ECBs conventional and unconventional monetary policies on European banking indexes returns*, maggio 2017.
60. E. Brenna, *Healthcare tax credits: financial help to taxpayers or support to higher income and better educated patients? Evidence from Italy*, giugno 2017.
61. G. Gokmen, T. Nannicini, M. G. Onorato, C. Papageorgiou, *Policies in Hard Times: Assessing the Impact of Financial Crises on Structural Reforms*, settembre 2017.
62. M. Tettamanzi, *E Many Pluribus Unum: A Behavioural Macro-Economic Agent Based Model*, novembre 2017.
63. A. Boitani, C. Punzo, *Banks' leverage behaviour in a two-agent New Keynesian model*, gennaio 2018.
64. M. Bertoni, G. Brunello, L. Cappellari, *Parents, Siblings and Schoolmates. The Effects of Family-School Interactions on Educational Achievement and Long-term Labor Market Outcomes*, gennaio 2018.
65. G. P. Barbetta, G. Sorrenti, G. Turati, *Multigrading and Child Achievement*, gennaio 2018.
66. S. Gagliarducci, M. G. Onorato, F. Sobbrio, G. Tabellini, *War of the Waves: Radio and Resistance During World War II*, febbraio 2018.
67. P. Bingley, L. Cappellari, *Workers, Firms and Life-Cycle Wage Dynamics*, marzo 2018.
68. A. Boitani, S. Perdichizzi, *Public Expenditure Multipliers in recessions. Evidence from the Eurozone*, marzo 2018.
69. M. Le Moglie, G. Turati, *Electoral Cycle Bias in the Media Coverage of Corruption News*, aprile 2018.
70. R. Davidson, A. Monticini, *Improvements in Bootstrap Inference*, aprile 2018.
71. R. Crinò, G. Immordino, S. Piccolo, *Fighting Mobile Crime*, giugno 2018.
72. R. Caminal, L. Cappellari, A. Di Paolo, *Linguistic skills and the intergenerational transmission of language*, agosto 2018.
73. E. Brenna, L. Gitto, *Adult education, the use of Information and Communication Technologies and the impact on quality of life: a case study*, settembre 2018.
74. M. Bordignon, Y. Deng, J. Huang, J. Yang, *Plunging into the Sea: Ideological Change, Institutional Environments and Private Entrepreneurship in China*, settembre 2018.
75. M. Bordignon, D. Xiang, L. Zhan, *Predicting the Effects of a Sugar Sweetened Beverage Tax in a Household Production Model*, settembre 2018.
76. C. Punzo, L. Rossi, *The Redistributive Effects of a Money-Financed Fiscal Stimulus*, gennaio 2019.
77. A. Baglioni, L. Colombo, P. Rossi, *Debt restructuring with multiple bank relationships*, gennaio 2019.

78. E. Cottini, P. Ghinetti, S. Moriconi, *Higher Education Supply, Neighbourhood effects and Economic Welfare,* febbraio 2019.
79. S. Della Lena, F. Panebianco, *Cultural Transmission with Incomplete Information: Parental Perceived Efficacy and Group Misrepresentation,* marzo 2019.
80. T. Colussi, Ingo E. Isphording, Nico Pestel, *Minority Salience and Political Extremism*, marzo 2019.
81. G. P. Barbetta, P. Canino, S. Cima, *Let's tweet again? The impact of social networks on literature achievement in high school students: Evidence from a randomized controlled trial*, maggio 2019.