Journal on Mathematics Education
Volume 15, No. 1, 2024, pp. 227-252

Promoting socioeconomic equity through automatic formative assessment

Alice Barana*, Marina Marchisio Conte

Department of Molecular Biotechnology and Health Sciences, University of Turin, Torino, Italy
*Correspondence: alice.barana@unito.it

Received: 7 July 2023 | Revised: 20 October 2023 | Accepted: 28 October 2023 | Published Online: 22 November 2023

Abstract

Ensuring equity in education is a goal for sustainable development. Among the factors that hinder equity, socioeconomic status (SES) has the highest impact on learning Mathematics. This paper addresses the issue of equity at the secondary school level by proposing an approach based on adopting automatic formative assessment (AFA). Carefully designed mathematical activities with interactive feedback were experimented with a sample of 299 students of grade 8 for a school year. A control group of 257 students learned the same topics using traditional methodologies. Part of the sample belonged to low SES. The learning achievement was assessed through pre-and post-tests to understand if the adoption of AFA impacted learning and whether the results depended on the students' SES. The results show a positive effect of the experimentation (effect size: 0.42). Moreover, the effect size of the experimentation restricted to the low-SES group is high (0.77). In the treatment group, the results do not depend on SES, while in the control group, they do, suggesting that AFA is an equitable approach while traditional instruction risks perpetuating inequalities.

Keywords: Automatic Assessment, Digital Learning Environment, Equity, Formative Assessment, Socioeconomic Status

How to Cite: Barana, A., & Marchisio Conte, M. (2024). Promoting socioeconomic equity through automatic formative assessment. *Journal on Mathematics Education, 15*(1), 227-252. http://doi.org/10.22342/jme.v15i1.pp227-252

Ensuring equitable quality education at all levels is one of the 17 sustainable development goals of Agenda 2030 of the United Nations (United Nations, 2016). Equity does not mean that all students have equal results; instead, it means that differences in their backgrounds do not affect their different outcomes (OECD, 2020). The Organization for Economic Cooperation and Development defines equity in education through fairness and inclusion (Field et al., 2007). Fairness refers to ensuring that personal and socio-economic circumstances should not be obstacles to achieving education potential; inclusion refers to ensuring a minimum standard of education for all.

Although increased access to education has been registered worldwide over the years, the recent health and political crises have jeopardized the improvements toward achieving this objective. According to the Report on the Sustainable Development Goals 2022 (United Nations, 2022), the COVID pandemic has deepened the existing educational disparities, mainly affecting marginalized populations. Students coming from disadvantaged backgrounds had the most significant difficulties in accessing online education, and they are at higher risk of permanent school leaving after the closures due to the pandemic. Keeping all students, particularly those from disadvantaged families, engaged with education as long as possible is essential to promote success in life (Field et al., 2007; OECD, 2020).

From the literature, the circumstances that mostly hinder equity, and thence are mostly studied, are gender, ethnic origin, and the socioeconomic background (Gates, 2014; OECD, 2020; Rohn, 2013), the latter having the highest impact on students' results in Mathematics (OECD, 2020). There are plenty of studies showing the correlation between socioeconomic status (SES) and achievement in Mathematics at every instructional level and all over the world (Baya'a, 1990; Cascella, 2020; McConney & Perry, 2010; Osadebe & Oghomena, 2018; Valli Jayanthi et al., 2014; Wang et al., 2014). From PISA 2018 results, SES is even a stronger predictor of achievement in Mathematics than in Reading and Sciences, accounting for 13.8% of the performance (OECD, 2020). While students from wealthy families can usually rely on more solid resources to find their path, for those from lower socioeconomic classes education can make the difference and enhance their opportunities for their future (OECD, 2020). The latter group of students has often a less stimulating home environment and a reduced support in studying. As a consequence, they show a lower level of engagement towards school (Ng et al., 2018) and of academic achievement (Zhu, 2018). However, it seems that this effect can be influenced by peers' SES, both in a positive way, when peers' SES increases and when classrooms are racially integrated, and in a negative way, when students with low SES are concentrated and the few peers from more wealthy families tend to adapt their behavior to the majority (Cascella, 2020; Coleman, 1966). Therefore, the class and school environments have a key role on the development of equity. It is also suggested that underperformance of students with particular personal characteristics can be partly linked to the fact that these students tend to be concentrated in groups, as in the case of immigrants and families with low SES (Zhu, 2018). Thus, class composition has been indicated as a key strategy to act on equity from a political perspective (Field et al., 2007).

In Italy, where this study is situated, inclusion is a milestone of the national guidelines of the Ministry of Education: according to (MIUR, 2012), the inclusion of people and the integration of different cultures are two of the principles around which the educating action should be organized. The guidelines suggest activating personalized paths and specific strategies to help the weakest students to cope with their difficulties and to reduce school failure, also in collaboration with local institutions. Unlike other countries (Wright, 2016), in Italy Mathematics is not taught in ability groups, and classes are composed by mixed-ability groups. However, some differences can be noticed in the students' socioeconomic backgrounds according to the school location and, for upper secondary schools, due to the different curricula that schools offer (Cascella, 2020).

In this setting, Mathematics can have a key role. Wright (2016) develops a conceptualization of "Mathematics for social justice", where learning Mathematics should help students overcome social inequalities. In his view, Mathematics helps develop critical education and to form critical citizens, able to reason with their mind and to make autonomous choices. Mathematics can help understand the world and its scientific, economic, and social processes (Rohn, 2013). This is particularly relevant in an era where information is easily available to everyone, but social media and advertisements try to influence people's opinions, preferences, purchases, and actions. This view challenges the dominant idea of Mathematics as unreachable, far from everyday reality, useless, and disengaging. To teach Mathematics for social justice, it is necessary to move away from traditional approaches which, as a result, perpetuate social inequities (Wright, 2016). According to Gutstein (2006, p. 10) the disempowering methods for teaching Mathematics favor the capitalist economies, which are always in search of "an ever-growing army of low-skilled, compliant, docile, pleasant, obedient service workers". Boaler (2008) showed that teaching mathematics through more open-ended, collaborative, problem-solving approaches, with students in mixed-ability groups, contributes to achieving more equitable results and reducing the gender gap. According to Heritage and Wylie (2018), when the focus is on achieving equity, classrooms need to be places where students have the opportunities to change their current situation, that is, covering their next step in their learning path, from where they are now to where they can go next. Formative assessment has the potential for enabling learners to "walk on the edge" as well as, through metacognitive reflection, developing a strong sense of personal agency and identity as competent and confident doers of mathematics (Heritage & Wylie, 2018). In Italy, the National Guidelines for primary and lower secondary school suggest that assessment should have a formative value and teachers are encouraged to use it to support inclusion (MIUR, 2012). Elbers and de Haan (2005) showed that peer collaboration and working in small groups during Mathematics classes were effective for addressing and overcoming language difficulties through peer engagement. Wright (2016) proposed 5 strategies which gather all the suggestions discussed above, aimed at promoting equity in the Mathematics classroom:

1. employing collaborative, discursive, problem-solving, and problem-posing pedagogies, which promote the engagement of learners with Mathematics;
2. recognizing and drawing upon learners' real-life experiences to emphasize the cultural relevance of Mathematics;
3. promoting mathematical inquiries that enable learners to develop greater understanding of their social, cultural, political, and economic situations;
4. facilitating mathematical investigations that develop learners' agency, enabling them to take part in social action and realize their foregrounds; and
5. developing a critical understanding of the nature of Mathematics and its position and status within education and society to maintain equity in the classroom.

An additional suggestion, indicated by Nortvedt and Buchholtz (2018) but excluded by Wright's strategies is the use of the computer. This suggestion is not very common among the various strategies usually identified in the literature to address equity issues, maybe because the most disadvantaged students could have more problems in accessing technology, as reported in (United Nations, 2022). Moreover, ICT literacy is also positively correlated to SES (Scherer & Siddiq, 2019). There is, however, a consistent body of literature which shows how the use of technologies, when introduced in the classroom for a relevant period of time, can be beneficial for students of low SES in learning Mathematics at different educational levels (Araya et al., 2015; Huang et al., 2016; Page, 2002; Suppes et al., 2014). These results encourage the systematic and extensive adoption of computer-based education in primary, middle and secondary schools with students from all social and economic backgrounds to foster understanding, skills acquisition and self-esteem. In fact, there are evidences that digital technologies can offer valid feedback to push forward the learner (Barana et al., 2021; Fahlgren et al., 2021; Gaona et al., 2018; Hoogland & Tout, 2018); moreover, they can provide realistic data, display tasks and processes using different registers and media, thus making Mathematics more relevant and realistic as well as helping overcoming language barriers (Nortvedt & Buchholtz, 2018; Sangwin, 2015; Stacey & Wiliam, 2013; Yerushalmy et al., 2017). Adopting strategies to foster equity does not imply teaching in the same way to all students, since each individual has different needs (Rohn, 2013): Zhu (2018) suggests pursuing "equitable inequalities" that correspond to the individual characteristics and necessities. The risk is to maintain a school system that perpetuates the inequalities (Gates, 2014; Yang Hansen & Strietholt, 2018). How can school make a difference? Possibly, approaches based on adaptive teaching are a solution (Aleven et al., 2016). Often, what happens in the classrooms is quite different. Some authors reported the teachers' tendency to propose higher-level reasoning to the most skilled students, who are generally better behaved and more positively disposed towards learning, without moving forward from low-risk teaching and procedural understanding with disadvantaged students (Rohn, 2013; Wright, 2016). The teachers' discouragement for working on equity could be due to the exam-focused culture in schools, excessive workload, and high levels of scrutiny of teachers (Wright, 2016). To bridge the gap between research and daily teaching practices, to see some changes in equity in the classrooms, effective teacher training is essential (Nortvedt & Buchholtz, 2018).

In Mathematics Education, the research can contribute to this issue by informing us about the causes of inequalities and providing strategies to reduce the disparities (Rohn, 2013); this study fits in the second strand. Since causes of inequities and effects of strategies to overcome them widely differ from country to country, there is a need for studies involving data from different parts of the world (Lerman, 2014). Nortvedt and Buchholtz (2018) advocate more research efforts to improve equity, in particular through formative assessment, teacher training and the use of digital technologies. This paper addresses the issue of identifying innovative strategies for an equitable Mathematics Education, based on digital learning. The focus is on overcoming differences due to the students' socioeconomic status, trying to avoid disadvantaging any category of students. We frame our analysis in the Italian panorama with grade 8 students, that are those attending the last year of lower secondary school. In Italy, all lower secondary schools follow the same curriculum and class groups are mixed-ability ones that is, they are heterogeneous in the composition in terms of proficiency. The reasons for the school choice are mainly related to distance from home or parents' workplace, and schools' reputation. As suggested by Nortvedt and Buchholtz (2018), the learning approach under study is automatic formative assessment (AFA); as suggested by Wright (2016), the learning activities are based on problem-solving and group-working. The setting of the study is experimental; it involves 299 students of grade 8 (treatment group) with their Mathematics teachers, and a control group of 257 students. This study is part of a biggest project, named "Città Educante" (which stands for "Educating City"), promoted by CNR in collaboration with the University of Turin, aimed at investigating the impact of problem solving and formative assessment using innovative digital learning environments under different perspectives (Barana et al., 2019; Barana & Marchisio, 2020). In this paper we will present the results of the project related to equity, linking them to other results on Mathematics learning and student engagement. Our goal is not just focusing on students with low SES and finding ways to engage them in relevant learning. It must not be forgotten that the first aim of equity is that all students should achieve minimum standards, independently on their background. Thus, we try to investigate the role of AFA in supporting learning independently of the students' SES. Using the previously cited definition of equity as composed by fairness and inclusiveness and focusing on the students' socioeconomic background as the main barrier to equity, we will try to understand if AFA activities developed according to the cited model are effective to enhance mathematical learning of all students and if the results depend on SES. Therefore, our research questions are the following ones:

(RQ1) Are digital activities with AFA effective in improving students' learning results?
(RQ2) Does the impact of the interactive activities with AFA on learning depend on the students' socioeconomic status?

The paper is structured as follows: the Method section details the methodology used to answer the research questions, including the structure of the experimentation, the nature of the AFA activities proposed, and the data analysis process. The Results and Discussion section presents the results of the experimentation in relation to the two research questions, and discusses them. The Conclusion section draws the conclusions, answering the research questions and discussing some limits and future directions of this study.

METHOD

Automatic Formative Assessment and Equity

In this study, we follow Black and Wiliam's conceptualization of formative assessment (FA) from (Black & Wiliam, 2009): borrowing their words, "practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited." FA is more than just providing feedback: the information collected through assessment needs to provoke some change in the learning path, with the purpose of reducing the gap between current and desired performance (Wiliam, 2006). Black and Wiliam (2009) also identify 5 key strategies that should guide the development of FA activities:

KS1. clarifying and sharing learning intentions and criteria for success;
KS2. engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding;
KS3. providing feedback that moves learners forward;
KS4. activating students as instructional resources; and
KS5. activating students as the owners of their own learning.

Digital technologies for Mathematics can help automatize part of the process of FA, such as the delivery of tasks, the provision of feedback and the collection of data: this is the case of automatic formative assessment. In a previous paper we gave a conceptualization of AFA (Barana et al., 2021). It draws on a 6-point model for the design of mathematical tasks with AFA using an automatic assessment system running on a mathematical engine, such as Möbius Assessment or STACK. Similar systems are able to generate randomized variables as outputs of mathematical computations, plot graphs and accept mathematical answers for their equivalence to the correct ones (Fahlgren et al., 2021; Sangwin et al., 2010).

The model is based on the following points:

1. Availability of assignments in terms of date, time, and number of attempts;
2. Algorithm-based questions and answers, where students find random values, parameters, or formulas that make questions and the relative answers randomly change at every attempt;
3. Open mathematical answers, whenever multiple-choice ones can be avoided;
4. Immediate feedback, provided while they are still focused on the task;
5. Interactive feedback, that is a step-by-step interactive resolution that shows a possible solving path or a guide for understanding a concept.
6. Real-world contextualization of tasks or relevant applications which make Mathematics more tangible.

Figure 1 shows part of a task carried out with AFA which follows this model. The task involves linear functions and distance-time graphs. It is contextualized in real life so that students have to interpret the results in a real-world situation. The interactive sections are shown one at a time and guide students to describe and interpret the graph, deducing the velocities of the segments and the formulas of the corresponding lines. At each step, students receive feedback about the correct answer and are able to use it in the next step. The whole interactive path can be conceived as feedback, since it provides a good example of a correct solving process and helps students identify their mistakes or misunderstandings (see (Barana et al., 2021) for a complete conceptualization of interactive feedback).

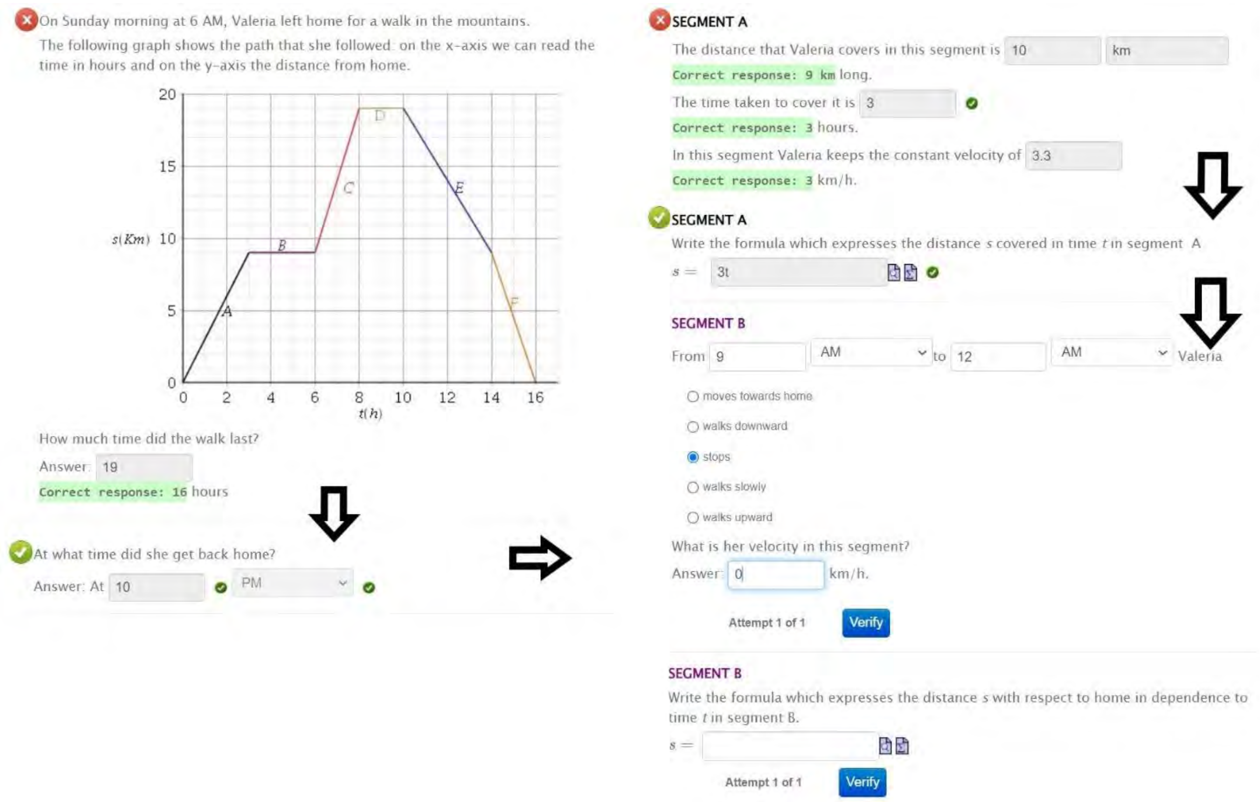

Figure 1. Part of an interactive activity with automatic formative assessment

After the sections visible in Figure 1, the task goes on to describe, with words and formulas, all the segments. Thus, students have to move from the graphic register to the symbolic register through numeric and verbal registers. The response areas vary based on their scope: some of them are multiple-choice, some accept the answer in a numeric form and others check the correctness of mathematical formulas independently of the form in which they are expressed. Thanks to an algorithm running behind the question, numerical values change at each attempt, and the graphic, correct answers and feedback change accordingly. Students have unlimited attempts available, and during subsequent attempts they have to use the acquired understanding and processes, not just recall the correct results.

The tasks developed according to this model can provide useful resources to build equitable classroom activities. Firstly, they engage students in relevant mathematical activity if the task is suitably contextualized and left open (Sacchet, 2022). It can also be used for group activity or as a base for class discussion. The interactive feedback can support the teacher in managing the classroom activity: once solved a task, the most skilled students (or small groups) can autonomously go to the next one, while those with more difficulties can follow the interactive feedback to understand the correct solution. The teacher can provide more tasks to the higher achieving students (or group of students), so that they have something to work on when their classmates are still working on the core activities. A similar adaptive organization can lead to a better management of time; moreover, it avoids the risk that students in difficulty get frustrated if they do not have enough time to complete the activities, meanwhile the fastest ones do not get bored and start disturbing the others while waiting for the rest of the class. Such AFA activities can also support the development of self-regulation, a competence which can boost Mathematics learning (Semana & Santos, 2018). Semana and Santos suggest that some students might need more support than others to develop self-regulation, and that differentiation is key to provide equitable opportunities to learn Mathematics. With AFA, the possibility to repeat the activities, each time finding different values, functions, graphs, or other mathematical objects, offers the opportunity to repeat reasonings until mastered. The immediate feedback helps acknowledge the achieved level and, in case of mistakes, the interactive feedback helps understand where one's reasoning failed and correct it immediately. Therefore, as claimed by Heritage and Wylie (2018), the implementation of AFA could have the benefit of supporting the development of students' mathematical identity as effective and capable learners.

Structure of the Experimentation

In order to answer the research questions, we used an experimental analysis, based on the counterfactual comparison of the outcomes of two groups of participants: the treatment group, composed by students who used the interactive activities with AFA for learning Mathematics, and the control group, who did not use interactive activities and AFA, but traditional teaching methods. The control group, having similar characteristics than the treatment one, serves the purpose of showing what would have happened to the subjects of the intervention if they had not been exposed to it (Angrist & Pischke, 2008).

The participants were in total 547 students belonging to 24 8th-grade classes from 6 different lower secondary schools from various areas of Turin. About 50% of the selected classes belonged to schools located in the suburbs of Turin, attended by at least 90% of students with low SES; the remaining 50% of the sample belonged to schools located in the city center and attended by at least 90% students with middle-high SES and wealthier families. To determine the schools' SES we used data from the INVALSI surveys provided by the schools. INVALSI is the national institute in charge of evaluating the school system, which annually administers standardized tests in Mathematics, Reading and English, as well as surveys to students, teachers and headmasters, to monitor the quality of the school system. Similarly to PISA, INVALSI computes the students' SES using the Economic Social Cultural Status (ESCS) index (Campodifiori et al., 2010) based on the highest parental occupation and education, and on the availability of some material goods, conceived as indicators of family economic condition. Since the literature reports evidence that the group's SES influences the relation between individual SES and achievement (Cascella, 2020; Coleman, 1966), given that the school included in the experiment were attended for the great majority either by students from low social classes, or by families with medium-high socioeconomic background, we split the sample into two parts according to the students' SES, as shown in Table 1. All teachers of Mathematics of the classes were involved. They were in total 23, one for each class except for one teacher in the treatment group who had two classes. Initially, the two groups should have been more balanced in terms of numbers, but some classes did not follow all the activities, so they were removed from the sample for the analyses. It happened especially in the control group, probably because the teachers did not understand the importance of completing all the required activities, not being involved in the experimental part. Since the classes were assigned randomly to the two groups, also the distribution of the teachers in the two groups was random-based and they had similar characteristics in terms of teaching experience (collected through a questionnaire).

Table 1. Numbers of students and classes participating in the experimentation

SES	Schools	Classes			Students		
		Control	Treatment	Total	Control	Treatment	Total
Low	4	4	7	11	84	152	236
Medium-high	2	6	7	13	164	147	328
Total	6	11	13	24	248	299	547

The intervention proposed to the treatment classes consisted in digital activities with AFA about all the topics usually covered at grade 8 in Italy, with a special focus on algebraic formulas and functions, chosen as didactic goals of the experiment. These topics have particular relevance at grade 8 due to their implications for the Mathematics curriculum of upper secondary school. Each treatment class had access to a Moodle page in a dedicated platform, where they could find interactive activities with AFA grouped by topic. All the activities were designed and created by university researchers (the authors of this paper) in collaboration with an INVALSI expert. They were presented to the teachers of the treatment group during some training meetings, carefully showing them how to use the activities. The proposed modalities to use the materials were two:

1. in the classroom, displaying the tasks through the Interactive Whiteboard (IWB). Students, in small groups of 3 or 4, were asked to solve one task in paper-and-pen modality. The teacher collected all the answers and the class agreed on one of them to be checked using the automatic assessment. After verifying whether it was correct, all the groups, in turn, had to show their solving process to the others. The interactive worksheets displayed on the IWB could support the collective discussion and give prompts for deeper reflection;
2. online, as individual homework, using the online assessment and the interactive feedback to check understanding. Students could autonomously navigate within the platform and make one or more attempts to the assignments.

Moreover, the lessons related to algebraic formulas and functions were held face to face by the class teachers in collaboration with the first author of this paper. The activities lasted 10 hours per class and were managed collaboratively. In total, the students worked on the interactive materials for about 40 hours during the entire school year, from December to May.

The paper-and-pen modality for group activities was chosen because in many schools there were not enough computers or tablets available to the students for these activities and it was not feasible neither to bring computers for all students at school every time the teachers wanted to use the interactive materials, nor to ask students to bring their computers or tablets from home. Contrary to expectations, the schools located in the most disadvantaged backgrounds were the best equipped with digital infrastructures (a working Wi-Fi connection, modern computer labs or tablet for students, new IWBs in all rooms), while the schools located in the city center and attended by middle or high-income families lacked modern facilities enabling digital education (the researchers often had to bring their own computers and projectors for the activities). The reason is that having a high number of students of low SES allows schools to be placed in the highest positions of the rankings for receiving fundings from the government or the European Union, and usually this money is used to modernize the schools.

Figure 2 shows an example of a classroom activity on linear functions. The task is a real-word problem concerning homework planning. It asks students to fill tables with data from the text and then to display the graph. After that, 3 questions guide students to the analysis of the graph and to deducing linear functions from the graph. It actively engages students since it asks them to imagine how they would organize their homework and add their homework schedule to the graph. In this way, they can immerse themselves in the task and gain a deeper understanding of it.

Holiday homework

The Mathematics teacher assigned 20 tasks as homework for the 7 days of Easter holiday. Maria solves 3 tasks per day on the first 6 days, on the last day she solves the remaining 2. Daniele solves 2 tasks per day on the first 6 days, on the last day he solves all the remaining ones. Anna takes a rest on the first 3 days. How many tasks shall she do each day in the remaining days in order to finish them all on time?
Fill the following table with the number of tasks to be done each day and with the progressive total of tasks solved each day. How would you organize your holiday homework?

MARIA			DANIELE			ANNA			YOU		
Day	Solved tasks	Total tasks	Day	Solved tasks	Total tasks	Day	Solved tasks	Total tasks	Day	Solved tasks	Total tasks
1	3	3	1	2	2	1	0	0	1	4	4
2	3	6	2	2	4	2	0	0	2	4	8
3	3	9	3	2	6	3	0	0	3	4	12
4	3	12	4	2	8	4	5	5	4	4	16
5	3	15	5	2	10	5	5	10	5	4	20
6	3	18	6	2	12	6	5	15	6	0	20
7	2	20	7	8	20	7	5	20	7	0	20
Draw the points			Draw the points			Draw the points			Draw the points		
Sketch the graph			Sketch the graph			Sketch the graph			Sketch the graph		

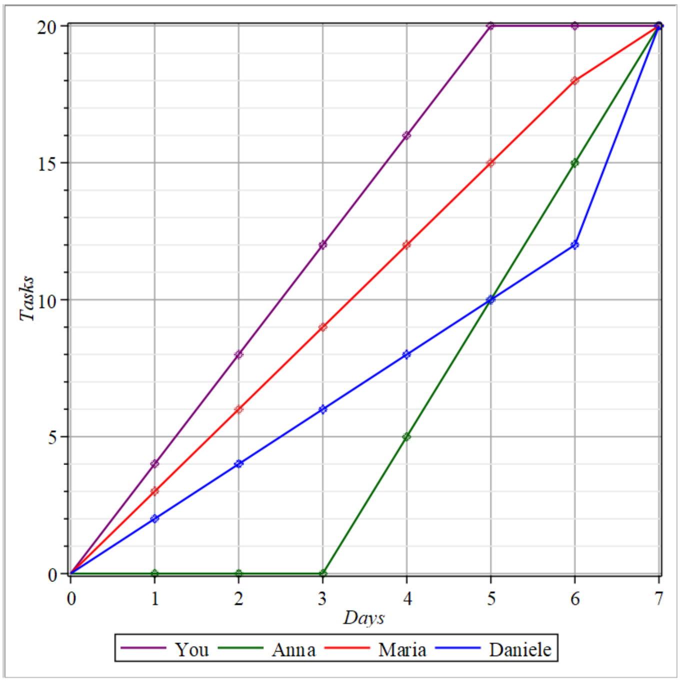

Figure 2. Example of a digital activity proposed to the treatment group during the experimental activities

In the classroom, the teachers displayed the text of the problem and asked students to create a table on their notebook, to fill it with data from the text, and to draw the graphs. When all groups completed this work, the teacher engaged all the students in a collective discussion supported by the IWB. They filled the interactive tables and displayed the graph, so that all students could compare it with their work. After that, the teacher proposed the three questions one at a time. The groups had to agree on an answer and motivate it to the rest of the class. The answers were checked through automatic assessment. The interactive worksheet was then uploaded in the digital learning environment and students were asked to access it from home and repeat the task. Moreover, they could practice with other tasks on linear functions with automatic assessment and interactive feedback, such as the one previously shown in Figure 1. Other examples of the experimental activities are discussed in other papers, such as: (Barana, 2021, 2022).

The teachers of the control classes did not have access to the experimental activities, in order not to influence their teaching methodologies. They were asked to use their usual teaching approach, which was mainly traditional. They covered the same topics as the experimental group, and they also devoted time to formulas and functions, as they are core topics in grade 8.

The experimentation was evaluated through several instruments:

1. a pre-test performed both by the treatment group and by the control group, designed to provide a snapshot of the students' starting level in Mathematics. It was on different mathematical topics and was administered with paper-and-pen modality to all before starting the experimental activities;
2. a mid-term test, administered to both groups with paper-and-pen modality, covering the basics of algebraic formulas and functions. It was taken after the beginning of the activities, but before starting the specific activities on formulas and functions. Its aim was to track the progress in the learning path and represented the students' starting point on these topics;
3. a post-test, performed by both groups with paper-and-pen modality, only focused on formulas and functions, to measure the competences gained through the experimental activities;

The learning tests were proposed through paper and pen and not with the digital assessment in order not to favor the treatment students for their familiarity with the digital platform. They were composed of items coming from the INVALSI national surveys for grade 8, with little adjustments to avoid proposing items that students had already seen. The kinds of questions were the same in the three tests. They consisted in mainly multiple choice or closed answers (numeric or formula), but there were also some open-ended items, such as asking to justify some statement. All of them assess the students' problem-solving and argumentation skills. Most of the items were contextualized in real-life situations. All the items included different tasks with respect to the experimental activities, both in the contexts and in their form. In fact, the experimental activities were mainly explorative or problem-solving activities and students were asked to find and justify multiple solving strategies or interact with the mathematical content, while the test items just asked students to provide the answer to problems, without guiding them in the solving approach. The pre-test was composed of 19 items; the mid-term test included 25 items; the post-test 11 items. For each test, students had 45 minutes to answer the questions. While the initial and intermediate tests were conceived to measure the initial level of student mathematical and problem-solving skills, the final test included high-level reasoning tasks, to compare the development of competence of the control and treatment group. Figures 3, 4 and 5 show examples of questions of the pre-, mid-term, and post-test respectively. We selected three different questions about formulas from a functional point of view, topic on which many of the experimental activities focused.

Data Collection and Analysis

In this paper we will examine the results of the experimentation from a general learning perspective to understand if the digital activities with AFA can enhance equity in learning. Firstly, we will focus on the experiment's net results by comparing the outcomes of the learning tests of the treatment and the control groups to understand if the experimental activities were effective in improving learning results. We will then investigate if the social environment affects the experimentation results.

The formula $L = L_0 + K \cdot P$ expresses the length L of a spring as the applied weight P varies. L_0 represents the rest length of the spring in centimeters; K represents how much the spring is stretched in centimeters when one unit of weight is applied. Which of the formulas listed below best describes the following sentence: "it is a very short and a very hard spring"?

A. $L = 10 + 0{,}5 \cdot P$
B. $L = 10 + 7 \cdot P$
C. $L = 80 + 0{,}5 \cdot P$
D. $L = 80 + 7 \cdot P$

Figure 3. Example of a pre-test question

The English teacher assigned her students a test composed of 25 items, and she explained that the total score s would be calculated by assigning 4 points for each correct answer and subtracting 2 points for each incorrect or missing answer.

a. The maximum possible score is: . . . .
b. If you answered 5 questions incorrectly and all the others correctly, what is your total score?

Figure 4. Example of a mid-term question

The nautical mile, a unit used to measure distances at sea, is equal to 1852 meters.

a. If L denotes the length in kilometers, which of the following formulas allows you to calculate the corresponding length in miles denoted by M?
   A. $M = L \cdot \frac{1000}{1852}$
   B. $M = L \cdot \frac{1852}{1000}$
   C. $M = L \cdot \frac{1852}{1000} : 1000$
   D. $M = L \cdot 1852 \cdot 1000$
b. Explains why the other options are not correct.

Figure 5. Example of a post-test question

In order to discuss the answers to the research questions, we firstly collected the pre, mid-term, and post-tests' individual results, scaled from 1 to 100 to express the percentage of correct answers, of both treatment and control group. These values were saved in the PRE, MT, and POST variables respectively. For each student, we computed the difference between the post and pre-test results, thus defining a new numeric variable: DIFF_POST_PRE. This variable measures the grade of improvement at the end of the year in reference to the initial level measured in the pre-test. It assumes positive values for students who improved their results at the end of the school year and negative values for those who had a lower performance in the post-test than in the pre-test. The sample was then restricted to the only the students who took both the pre and the post tests, being present at school on the days the tests were administered: it is the 83% of the whole sample (80% of the control group and 85% of the treatment group). Similarly, we also computed the differences between the scores of the post- and mid-term tests, and between the mid-term and pre-tests. They were saved in the DIFF_POST_MT and DIFF_MT_PRE variables. These data were cross-checked with the GROUP variable, indicating if students belonged to the treatment or control group, and with the indicator of SES, which we, for simplicity, considered as dichotomous (low and medium-high SES). We used only two values because we observed a large disparity between schools located in the city center and those located in the suburbs, thus they were sufficient to describe the sample. As indicator of SES, we used the INVALSI ESCS index (Campodifiori et al., 2010) provided by schools as explained above. For reasons of clarity, the list of variables used in the analysis are reported in Table 2.

Table 2. List of variables used in the analyses

Variable	Type	Values
GROUP	Dichotomous	Control or Treatment
SES	Dichotomous	Low or Medium-High
PRE	Numeric	Percentage of correct answers to the pre-test
MT	Numeric	Percentage of correct answers to the mid-term test
POST	Numeric	Percentage of correct answers to the post-test
DIFF_POST_PRE	Numeric	Difference between the percentages of correct answers to the post-test and the pre-test (POST – PRE)
DIFF_POST_MT	Numeric	Difference between the percentages of correct answers to the post-test and the mid-term test (POST – MT)
DIFF_MT_PRE	Numeric	Difference between the percentages of correct answers to the mid-term test and the pre-test (MT – PRE)

Cronbach's Alpha was calculated to verify the internal consistency of the results of the three tests. Cronbach's Alpha is a coefficient widely used in educational research to demonstrate the reliability of questionnaire and test items (Taber, 2018). It depends on the number of items considered; usually values higher than 0.7 are considered acceptable, although lower values may be considered acceptable if the scale analyzed is composed of a small number of items (Taber, 2018). One-way ANOVA was used to check if the differences between the scores obtained in the learning tests (DIFF_POST_PRE, DIFF_POST_MT, and DIFF_MT_PRE variables) depended on the group, to investigate if the experimental activities were effective to improve students' learning results, and thus answer (RQ1). To provide more interpretable results, we also used the covariance analysis (ANCOVA) to compute the adjusted mean of the post-test results on the base of the pre-test results. ANCOVA is a test which verifies if the means of a dependent numerical variable (in our case, POST) are equal across different levels of an independent categorical variable (GROUP) and across the levels of another numerical variable which correlates with the dependent variable, called covariate (in our case, PRE and MID-TERM) (Creswell, 2009). To support the results, the effect size was computed to identify the measure of the strength of the relation between the variables (in this case, the differences between the tests scores and the group) (Creswell, 2009). We use Cohen's d index (Cohen, 1969), which is defined as:

$$d = \frac{M_1 - M_2}{S} \quad (1)$$

Where M_1 and M_2 are the means of the two groups and S is an esteem of the standard deviation of the population from which the samples are extracted: in particular,

$$S = \sqrt{\frac{(n_1-1)\cdot S_1{}^2+(n_2-1)\cdot S_2{}^2}{(n_1-1)+(n_2-1)}} \quad (2)$$

where n_1 and n_2 are the numerosity of the two groups and S_1 and S_2 are the standard deviations. We chose this index as the whole sample is numerous (>20), and the standard deviations are not so different (Pellegrini et al., 2018).

To study the effects of the activities in the two different socio-economic contexts and answer (RQ2), we used two-way ANOVA on the difference between post-test and pre-test scores (DIFF_POST_PRE) as dependent variable, using GROUP and SES as independent variables. To investigate the interplay of the two independent variables on the experimentation effects, we ran four ANOVA tests on the DIFF_POST_PRE variable: two considering the GROUP variable as independent variable, on the two groups based on SES separately, and two considering the SES as independent variable, on the two groups based on GROUP (control or treatment group) separately. Lastly, we computed the effect size using the Cohen's d index on the sample restricted to the low SES group.

In the discussion, these results will be also compared with other results of this experiment, already published in other papers with different research goals, and with the qualitative perceptions of the researchers and teachers collected through the focus group. All families signed a consent form for the use of personal data for research purposes. Data were anonymized before treatment and analyzed using SPSS 27; the main results and conclusions will be reported in the following sections.

RESULTS AND DISCUSSION

Effectiveness of the experimentation from a learning perspective

Firstly, we will start the analysis of the experimentation's results by studying the net effects on learning achievements of the experimental activities, answering the research question: (RQ1) Are digital activities with AFA effective in improving students' learning results?

The intent is to understand if AFA could provide an effective strategy to facilitate learning in Mathematics and thus make the first step towards the identification of AFA as an equitable approach. For that purpose, we compared the results of the learning tests carried out by the treatment group to the ones carried out by the control group. The data considered for this analysis are the percentage of correct answers of the pre-test, mid-term test, and post-test, which we registered under the numeric variables PRE, MT, and POST.

The internal consistency of the three tests was measured through Cronbach's Alpha. The pre-test, composed of 19 items, reports 0.77 as Cronbach's Alpha. The mid-term test is composed of 25 items and has 0.80 as Cronbach's Alpha. The post-test, composed of 11 items, has 0.68 as Cronbach's Alpha. We can notice that the three tests have an acceptable level of internal consistency; the third one is slightly lower, but we have to consider that the number of items is lower and the difficulty of the test higher, since the tasks required high-level thinking.

As a first step, we performed the ANOVA on the dependent variable DIFF_POST_PRE, using the dichotomous variable GROUP, expressing the belonging to the treatment or the control group as the independent variable. The results are shown in Table 3: students of the treatment group started from a lower level (39.87 out of 100) than students of the control group (45.83 out of 100), but at the end of the experimentation, their results improved their score significantly more than students of the control group (the mean is 45.77, with an increase of 7.20 out of 100, while the latter obtained a mean of 45.83, with a decrease of 0.58 out of 100). The ANOVA test showed that the experimentation modality influenced the increase in the results between pre and post-tests in a statistically significant way ($F = 19.23$, $p < 0.001$). These data allow us to compute the effect size of the experimental activities, which measure the intervention's efficacy. Using the results in Table 3, Cohen's d is 0.42, a medium value for an effect size (Pellegrini et al., 2018).

If we consider the mid-term test results, we find out that they are more similar to those of the pre-test than the post-test. Like the post-test, it was exclusively focused on formulas and functions, and it was administered before the beginning of classroom activities on that topic. Therefore, it can represent the students' starting point on this topic. We need to underline that the mid-term test was easier than the post-test, so comparisons between these tests' results are meaningless in absolute terms. The difference between the percentages of correct answer to the post-test and the mid-term one (numeric variable DIFF_POST_MT) has a negative mean, and the difference between results of the mid-term test and those of the pre-test (DIFF_MT_PRE) has a much higher mean than the previously studied DIFF_POST_PRE. Data are shown in Table 3.

Table 3. Results of the pre, mid-term and post-test. Values are expressed on a scale from 1 to 100, corresponding to the percentage of correct answers

		Control group	Treatment group	Total
Pre-test (PRE)	N	233	280	513
	Mean	45.83	39.87	42.58
	SD	21.28	19.00	20.22
Mid-term test (MT)	N	233	279	512
	Mean	61.06	57.12	58.91
	SD	18.06	16.98	17.57
Post-test (POST)	N	207	267	474
	Mean	45.77	47.24	46.60
	SD	20.48	19.60	19.98
Difference (DIFF_POST_PRE)	N	197	255	452
	Mean	-0.58	7.20	3.81
	SD	20.30	17.35	19.07
Difference (DIFF_POST_MT)	N	201	255	456
	Mean	-14.92	-9.50	-11.89
	SD	16.30	15.97	16.32
Difference (DIFF_MT_PRE)	N	222	264	486
	Mean	15.03	16.93	16.06
	SD	18.04	14.88	16.41

We computed an ANOVA test on the dependent variables DIFF_POST_MT and DIFF_MT_PRE, considering the dichotomous variable GROUP as the independent variable. We found that the post-test results in the treatment group decreased much less than the ones in the control group compared to the mid-term one: the difference is statistically significant ($F = 12.69$, $p < 0.001$), while there was no evident difference in the improvements from pre-test to mid-term tests between the two groups ($F = 1.624$, $p = 0.20$).

To provide more easily interpretable results, we used the covariance analysis to compute the adjusted mean of the post-test results, considering the pre-test results as the covariate. Then we repeated the test using the results of the mid-term test as the covariate. The results are shown in Table 4. Values are statistically significant (p = 0.002 and p = 0.003 respectively). These means can be interpreted as the grades that students would have obtained in the post-test once we eliminate the variance provided by their initial level, expressed though their grades in the pre (or mid-term) tests. The effects are more evident when we refer to the pre-test than to the mid-term one.

Table 4. Adjusted means of the post-test, using the pre-test and mid-term test as covariates

Covariate	Group	N	Mean	Standard error
Pre-test	Control group	197	43.99	1.19
	Treatment group	255	49.01	1.04
Mid-term test	Control group	201	44.38	1.08
	Treatment group	255	48.78	0.96

Relation between socioeconomic status and learning achievement

We now want to deepen the analysis, investigating how “equitable” the AFA approach has been in relation to Mathematics achievement. We aim to find an answer to the second research question: (RQ2) Does the impact of the interactive activities with AFA on learning depend on the students’ socioeconomic status? In other words, we would like to know if and how the students’ SES alter the results of the experimentation: we check the fairness of the innovative methodologies used for learning in the experimental activities, that is the independence of the students’ outcome from their socioeconomic status. Thus, we ran a two-way ANOVA to test the effect of the students’ socioeconomic status on the results of the experimentation.

Table 5. Results of the learning tests restricted to the students belonging to lower social classes

			Control group	Treatment group	Total
Low SES	Pre-test (PRE)	N	78	140	218
		Mean	42.17	34.74	37.39
		SD	20.06	17.44	18.71
	Post-test (POST)	N	76	136	212
		Mean	35.39	41.90	39.57
		SD	20.19	20.24	20.42
	Difference (POST - PRE)	N	71	130	201
		Mean	-7.90	7.07	1.78
		SD	21.98	17.71	20.57
Medium-high SES	Pre-test (PRE)	N	155	141	296
		Mean	47.67	45.00	46.40
		SD	21.55	19.17	20.47
	Post-test (POST)	N	131	131	262
		Mean	51.79	52.79	52.26
		SD	18.16	17.31	17.71
	Difference (POST - PRE)	N	125	126	251
		Mean	3.55	7.32	5.43
		SD	18.11	17.04	17.62

The aim was to understand if the experimental activities had a different impact in schools located in socially disadvantaged milieux of Turin, attended by students from low socio-economic classes, from others located in the city center and attended by medium or high-income families. As before, we considered the results of the pre and post learning tests, in particular the numeric variables PRE, POST, and DIFF_POST_PRE. We performed the two-way ANOVA test on the increments in the students' grades in the post-test as compared to the pre-test (DIFF_POST_PRE): the results are shown in Table 5. The statistics are significant: p=0.002, meaning that the learning achievements depends both on the SES and the kind of treatment (control/treatment group). Thus, the different social condition did alter the results of the experimentation.

Among students belonging to the low-SES group, while the control group decreased its mean grade of almost 8 points out of 100, the treatment group increased its mean grade of almost the same extent. Considering only students of low SES, the experimentation impacted significantly on achievement: the ANOVA test on the DIFF_POST_PRE variable restricted to the low-SES group returns p<0.001. On the other hand, in the medium-high SES group, both treatment and control group raised their scores, though the treatment group's ones are higher. In this case, the difference is not significant: the ANOVA test run on the DIFF-POST_PRE variable over the medium-high SES group returns p=0.09.

Table 5 can also be read vertically. Among the treatment group, both groups based on SES improved their results by the same extent. This indicates that the digital activities had the same impact on the two groups, or, in other words, that the social factor did not influence the outcomes of the experimental activities. As a proof, the ANOVA test run on the DIFF_POST_PRE variable over the SES variable, restricting the sample to the treatment group only, returns p=0.91. On the other hand, considering the control group only, the difference of the learning results between low and medium-high SES is considerable. The ANOVA test restricted to the control group only returns p<0.001, meaning that the difference is statistically significant. What immediately stands out among these results is the bigger effect of the experimental activities in the low SES group. In order to quantify this finding, we computed the experimentation's effect size restricted to the students from lower social classes. Cohens' d, calculated with these data, is d = 0.77, which is a medium-high value (Pellegrini et al., 2018) and it is nearly the double than the effect size previously computed considering the whole sample.

Discussion

The results show that students who used the interactive activities based on automatic formative improved their learning in Mathematics. While the control group, which studied Mathematics through traditional methodologies for the whole school year, did not improve its results from the beginning to the end of the year, the treatment group made significant improvements. This result corroborates the hypothesis that using similar interactive tasks could provide an inclusive learning environment, helping all students raise their outcomes. The post-test devised for this experiment was composed by high-order reasoning tasks, so its aim was not just to practice routine computations, but to detect the acquisition of mathematical competences in the field of formulas and functions. Most tasks were contextualized in the real world. The results show that, having the opportunity to learn through interactive tasks, discuss them in groups and understand their mistakes through interactive feedback, students of the treatment group developed competences and increased their high-level reasoning. As said before, such kind of reasoning is important to nurture future citizens, able to think clearly and make rational decisions. These findings are in line with what Boaler (2008), Elbers and de Haan (2005), and Wright (2016) suggest to promote equity in Mathematics classroom

Secondly, this analysis confirms that interactive activities that make extensive use of innovative technologies are more successful with students of lower social classes, while we did not observe the same effectiveness in students of upper classes. Observations made during the classroom activities confirm that the level of attention and students' engagement was higher in the schools situated in more disadvantaged areas than those attended by wealthier families. Our feeling was that students, in the former case, understood the great value of the opportunity they had been offered with the participation in this project. Many of their families do not have cultural end economic tools to support other extracurricular activities, unlike their peers of higher social classes, for whom these activities were just one of their many chances of learning. Moreover, the attention given to providing valuable feedback to all, the real-world contextualization of the activities and the high interaction allowed by the digital technologies allowed them to enter the virtuous circle of understanding Mathematics and keeping engaged with it (Ng et al., 2018). As confirmation of these observations, we report the results of another study, (Barana & Marchisio, 2020), which analyzes the same project from the perspective of engagement. Among the findings, we showed that the cognitive engagement level grew significantly more in students from lower social classes than in those from higher ones. Cognitive engagement has been described as the quality of mental investment students make in learning, and it is indicated by their level of self-regulation and their ability to set and pursue learning goals (Appleton et al., 2006; Ng et al., 2018). Moreover, students with low SES tended to resubmit the same assignments available online more times than students with medium-high SES. We considered the rate of submission per assignment as an indicator of engagement - see (Barana & Marchisio, 2020) for more details.

In another study on the same project, we examined the behavior of those students who made little or no access to the digital activities and defined them "reluctant users" of the technologies (Barana et al., 2019). We found that the percentage of reluctant users in students of low SES is similar to that of medium-high SES. From these findings we can affirm that, despite the potentially lower availability of digital technologies in their homes, students from low social classes used the online platform to the same extent, or even more, than students coming from wealthier families. In other words, the digital learning environment did not represent a barrier from the equity point of view and all students had the same chance to access to the activities. This is not in line with the findings of the United Nations report (2022), which report disparities in the access to technologies all over the world. On the contrary, the results are in line with other studies which show that digital technologies, when regularly adopted by teachers, can be beneficial for students of low SES in learning Mathematics (Araya et al., 2015; Huang et al., 2016; Page, 2002; Suppes et al., 2014). Moreover, the findings included in another paper which studies the same experimentation under the perspective of the power of the interactive feedback (Barana et al., 2021) supports the hypothesis that the structure of the AFA activities is the driver for improving learning results in disadvantaged contexts. In fact, (Barana et al., 2021) shows that the activities with interactive feedback were the most effective to close the gap between current and reference performance. Moreover, the positive effects of the interactive feedback were emphasized in the low-SES group. As the present study shows, the same group is also the one which achieved the best results in terms of improvements.

These findings support our conclusions: youngsters from disadvantaged backgrounds were particularly engaged with the innovative methodologies offered to the treatment group and used the interactive activities with more enthusiasm than students from wealthy families, both in the classroom and individually at home. The higher difficulty that they may have had in accessing the technologies necessary to complete the activities (a computer or a tablet) did not limit their use of the interactive path, which, on the contrary, was even wider than students with higher SES.

Going back to the experimentation's results, we can notice that in the treatment group, the improvements made by the students with low SES in the post-test (7.07) is very similar to that of the whole treatment group (7.20). It means that the experimental activities had the same impact on all the students who participated in them, independently of their background: in other words, they were fair, important dimension of equity. On the other hand, in the control group, the decrease in the results of the post-test of students from lower social classes (-7.90) is noticeably worse than the average value of the whole control group (-0.58), and of the students from higher social classes (3.55). This means that the different results of the experimental activities on the two groups based on SES is not due to a greater impact of these innovative methodologies on a particular class of students, but rather to a different impact of traditional methodologies on students with a different social background. This suggests that, especially for students from lower social classes, the use of traditional learning methodologies is not the best strategy to improve their results. Probably, the traditional education received by the control group, especially for students with low SES, was not effective to develop high-order reasoning skills, necessary to correctly answer the post-test questions.

As we noticed before, the activities proposed to the treatment group were not training for the final test, but they supported students in solving contextualized problems and understanding mathematical concepts through interactive feedback. The main focus of the post-test (and of the experimentation) was algebraic formulas and functions. Traditionally, these topics have often been dealt with through repetitive tasks, focusing on the manipulation of algebraic formulas rather than on the meaning that formulas convey (Cusi et al., 2011). The experimental activities were very far from similar traditional teaching models, which fail when the goal is developing understanding and competences, as shown by several studies (Gates, 2014; Nortvedt & Buchholtz, 2018; Yang Hansen & Strietholt, 2018). Focusing on students of medium-high SES, we can notice that the treatment group achieved higher results than the control group, but the control group did not suffer any particular disadvantages from the use of traditional methodologies. This result is in line with the studies supporting the idea that when students have a culturally rich home environment and a wide availability of resources, the choice of the learning methodologies used at school has a lower impact on their achievement levels (Giusti et al., 2015; OECD, 2020).

Lastly, one could argue that these results are due to the Hawthorne effect, that is, the set of modifications of a phenomenon or of a behavior which occurs due to the presence of observers, but which do not last in time (Cook, 1962). If the Hawthorne effect could have influenced a small portion of the experimentation results, we need to consider that:

1. the project lasted the whole school year, while the Hawthorne effect would have had a more limited effect;
2. only a few hours were held in the presence of the researchers, while the teachers adopted the methodologies and used the interactive materials for the whole school year;
3. the students had little perception that they were the treatment group of an experimentation, they just knew they were part of a University project for learning Mathematics;
4. the pre and post-tests were considered by teachers and students as regular class tests;
5. also the students from higher social classes would have suffered from this effect.

Thus, we think that the main part of these results is not compromised by the Hawthorn effect.

CONCLUSIONS

Summary of the obtained results

In this paper, AFA has been presented as an innovative approach to foster equity in school Mathematics.

We tackle equity from the perspective of SES, which is one of the principal factors affecting learning achievements. The power of this methodology has been evaluated through an experiment design involving 546 8th grade students. The results of the analyses helped us answer the two research questions, namely:

(RQ1) Are digital activities with AFA effective in improving students' learning results?
(RQ2) Does the impact of the interactive activities with AFA on learning depend on the students' socioeconomic status?

In the first place, we could observe a clear improvement of the learning results of the treatment group, while the control group did not increase their results, with a moderate effect size of the experimentation (0.41). Thus, the answer to the first research question is positive: the digital activities based on AFA had a moderate impact on the students' learning achievement. This question is aimed at investigating how "inclusive" this approach can potentially be. It would be unrealistic to expect that by using AFA all students would have improved their results; however, we found that they improved more than those who learned through traditional teaching approaches.

Moreover, we observed that the results of the experimentation depend on SES. In particular, the gap between treatment and control group sensibly increased when considering the students from lower social classes, while this difference is not significant in the medium-high SES. The effect size of the experimentation, considering only students from low SES, is high (0.77). Moreover, we observed that the treatment group's results do not depend on the SES, while the control group's results are heavily dependent on it. These findings show that AFA is a "fair" approach since the learning results of students who used AFA did not depend on their SES. On the contrary, traditional methodologies are not fair, since the results of the students who did not use the digital materials strongly depend on their SES. In conclusion, we can positively answer the second research question: there is a significant relation between the impact of the experimentation on learning and SES; in particular, in the low-SES group, the effects of interactive activities nearly doubled, while the effects were less visible in the medium-high SES group. However, in the treatment group, the SES did not affect the results, while in the control group it did.

Challenges and new directions

The results of this research open up several challenges. First, we can consider that the AFA activities included in the interactive path were mainly developed by experts in digital education and mathematics education. This resulted in top-quality activities, both from a didactic and from a technical point of view; however, they do not always reflect didactic activities elaborated by teachers themselves, whose results could have had different impacts on students' learning. The creation of digital tasks and activities to be used with formative purposes requires technical skills and knowledge of the tools, as well as a pedagogical preparation in the strategies and models of formative assessment. Otherwise, there is the risk of merely replicating traditional instruction with digital tools without taking advantage of the benefits gained from a correct, informed, and aware use of these technologies. This can be tackled through a specific training dedicated to the teachers or the instructors to author the learning activities and to use the existing ones in an appropriate way.

Secondly, the sample considered in this project is limited in number and geographical area. The experience of Città Educante can be replicated in other contexts (including different instructional grades, geographical areas, and mathematical contents) to verify that AFA activities are effective to promote equity and, more generally, to improve the quality of Mathematics teaching and learning. The didactic methodologies used in Città Educante have been proposed to enhance the teaching and learning of Mathematics of all Italian teachers through the Problem Posing and Solving (PP&S) Project, supported by the Ministry of Education to renew Mathematics teachers' practices (Brancaccio et al., 2015). The materials used during this experimentation were made freely available to all the Italian teachers enrolled in the PP&S Project; they can use, edit, and adapt them to their needs. Specific training on these methodologies is offered to the participants. In March 2020, when all Italian schools were closed due to the COVID-19 pandemic, the PP&S Project platform was open to all Italian teachers of every subject, even non-scientific ones. Thus, AFA interactive materials helped many Italian students keep on learning in an emergency situation.

Lastly, from an equity perspective, more in-depth qualitative research would help understand if and how the adoption of AFA as a regular teaching strategy can affect students' identity and how it can empower them to change their social conditions.

Further reflections and conclusions

One significant result that this work highlighted is the effect of interactive activities with AFA on students coming from low social classes. As we observed above, during the classroom activities we perceived a higher level of attention in schools located in disadvantaged areas of Turin than in schools located in the city center attended by wealthy families. From another study on the same data (Barana & Marchisio, 2020), we observed that the behavioral and emotional engagement levels increased at the same extent for the two groups of students, while the cognitive engagement level increased significantly more in students from lower socio-economic conditions. The high value of the effect size found in the low-SES group compared with the whole group reflects this trend in the cognitive engagement. Developing an interest in and an understanding of Mathematics in poorly educated families may improve young people's ability to actively engage in society and in the workplace, thus giving them a chance to improve their social status. Moreover, the fact that in both groups based on SES there were a similar number of reluctant users of the digital activities (Barana et al., 2019) contributes to characterize the digital learning environment as inclusive.

From the discussions with the teachers, it turned out that another category of students who highly benefited from AFA were those with learning disorders. We did not examine this aspect statistically, however we considered relevant to include this information here, since it corroborates the hypothesis that said didactic methodologies can be effective under the perspectives of inclusion and improvement of chances. In the treatment classes, all students had access to intellectually challenging learning, all students were provided with the opportunity to work collaboratively with their peers in a community of practice, and formative assessment gave them consistent occasion of being aware of their progresses through their learning path: these results, according to Heritage & Wylie (2018), are the primary vehicles to pursue equity goals. On the other hand, the learning results of students who used traditional methodologies (control group) are in line with a consistent body of research which points out that that schooling seemingly perpetuates SES inequalities in mathematics performance (Gates, 2014; Nortvedt & Buchholtz, 2018; Yang Hansen & Strietholt, 2018). In fact, among the students of the control group, those belonging to the low-SES group were the most disadvantaged.

Through these results, the use of AFA in a digital learning environment is proposed as a way to foster the 4th sustainable development goal of Agenda 2030 (quality education), since it promotes quality learning opportunities for all. It also fosters the 10th goal (reduced inequalities), since the use of the learning activities proposed here could reinforce the participation of disadvantaged students to meaningful learning mathematical activities, and therefore contribute to uneven social inequalities through education. Thus, this paper shows that innovation in education can be a powerful weapon to promote equity and overturn the social situation of disadvantaged students. Moreover, it also shows that traditional education can increase inequalities in school results, and from the literature we know that they will probably reflect into higher social and economic inequalities. Thus, this paper suggests a possible approach to Mathematics Education which could help reduce the gap between advantaged and disadvantaged youngsters. This paper wants to send a message to schoolteachers, who are encouraged to use similar activities in their classrooms; it is a message for other researchers, who are invited to expand this research to different and larger samples, and to develop further the proposed approaches; finally, it is a message aimed at policymakers, because the research can shed light on the good practices that deserve to be supported to build sustainable communities.

Acknowledgments

The authors desire to thank the other members of the research group of the Città Educante Project: Stefano Boffo, Francesco Gagliardi, Rossella Garuti, and Rodolfo Zich.

Declarations

Author Contribution : The authors declare that their contributions are equal.

Funding Statement : This research was funded by Compagnia di San Paolo through the OPERA project (Open Program for Education Research and Activities).

Conflict of Interest : The authors declare no conflict of interest.

REFERENCES

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction Based on Adaptive Learning Technologies. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (2nd ed., pp. 522–560). Routledge.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Araya, R., Gormaz, R., Bahamondez, M., Aguirre, C., Calfucura, P., Jaure, P., & Laborda, C. (2015). ICT Supported Learning Rises Math Achievement in Low Socio Economic Status Schools. In G. Conole, T. Klobučar, C. Rensing, J. Konert, & E. Lavoué (Eds.), *Design for Teaching and Learning in a Networked World* (Vol. 9307, pp. 383–388). Springer International Publishing. https://doi.org/10.1007/978-3-319-24258-3_28

Barana, A. (2021). From Formulas to Functions through Geometry: A Path to Understanding Algebraic Computations. *European Journal of Investigation in Health, Psychology and Education*, *11*(4), 1485–1502. https://doi.org/10.3390/ejihpe11040106

Barana, A. (2022). Understanding linear functions in an interactive digital learning environment. In U. T. Jankvist, R. Elicer, A. Clark-Wilson, H. G. Weigand, & M. Thomsen (Eds.), *Proceedings of the 15th international conference on technology in mathematics teaching (ICTMT 15)* (pp. 255–262). Danish School of Education, Aarhus University.

Barana, A., & Marchisio, M. (2020). An interactive learning environment to empower engagement in Mathematics. *Interaction Design and Architecture(s) Journal - IxD&A, 45*, 302–321.

Barana, A., Marchisio, M., & Sacchet, M. (2019). Advantages of Using Automatic Formative Assessment for Learning Mathematics. In S. Draaijer, D. Joosten-ten Brinke, & E. Ras (Eds.), *Technology Enhanced Assessment* (Vol. 1014, pp. 180–198). Springer. https://doi.org/10.1007/978-3-030-25264-9_12

Barana, A., Marchisio, M., & Sacchet, M. (2021). Interactive Feedback for Learning Mathematics in a Digital Learning Environment. *Education Sciences*, *11*(6), 279. https://doi.org/10.3390/educsci11060279

Baya'a, N. F. (1990). Mathematics anxiety, mathematics achievement, gender, and socio-economic status among Arab secondary students in Israel. *International Journal of Mathematical Education in Science and Technology*, *21*(2), 319–324. https://doi.org/10.1080/0020739900210221

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Boaler, J. (2008). Promoting 'relational equity' and high mathematics achievement through an innovative mixed-ability approach. *British Educational Research Journal*, *34*(2), 167–194. https://doi.org/10.1080/01411920701532145

Brancaccio, A., Marchisio, M., Palumbo, C., Pardini, C., Patrucco, A., & Zich, R. (2015). Problem Posing and Solving: Strategic Italian Key Action to Enhance Teaching and Learning Mathematics and Informatics in the High School. *Proceedings of 2015 IEEE 39th Annual Computer Software and Applications Conference*, 845–850. https://doi.org/10.1109/COMPSAC.2015.126

Campodifiori, E., Figura, E., Monica, P., & Ricci, R. (2010). Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia. *INVALSI Working Paper Series*, *2010*(2), 1–25.

Cascella, C. (2020). Intersectional effects of Socioeconomic status, phase and gender on Mathematics achievement. *Educational Studies*, *46*(4), 476–496. https://doi.org/10.1080/03055698.2019.1614432

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press.

Coleman, J. S. (1966). *Equality of educational opportunity* (No. 0E-36001). U. S. Government Printing Office.

Cook, D. L. (1962). The Hawthorne Effect in Educational Research. *The Phi Delta Kappan*, 44(2).

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed). Sage Publications.

Cusi, A., Malara, N. A., & Navarra, G. (2011). Early Algebra: Theoretical Issues and Educational Strategies for Bringing the Teachers to Promote a Linguistic and Metacognitive approach to it. In

J. Kai & E. Knuth, *Early Algebraization: Cognitive, Curricular, and Instructional Perspectives* (pp. 483–510). Springer.

Elbers, E., & de Haan, M. (2005). The construction of word meaning in a multicultural classroom. Mediational tools in peer collaboration during mathematics lessons. European Journal of Psychology of Education, 20(1), 45–59. https://doi.org/10.1007/BF03173210

Fahlgren, M., Brunström, M., Dilling, F., Kristinsdóttir, B., Pinkernell, G., & Weigand, H.-G. (2021). Technology-rich assessment in mathematics. In A. Clark-Wilson, A. Donevska-Todorova, E. Faggiano, J. Trgalová, & H.-G. Weigand (Eds.), Mathematics Education in the Digital Age: *Learning, Practice and Theory* (pp. 69–83). Routledge.

Field, S., Kuczera, M., & Pont, B. (2007). *No more failures: Ten steps to equity in education*. OECD.

Gaona, J., Reguant, M., Valdivia, I., Vásquez, M., & Sancho-Vinuesa, T. (2018). Feedback by automatic assessment systems used in mathematics homework in the engineering field. *Computer Applications in Engineering Education*, *26*(4), 994–1007. https://doi.org/10.1002/cae.21950

Gates, P. (2014). Equity and Access in Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 217–221). Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_58

Giusti, S., Gui, M., Micheli, M., & Parma, A. (2015). *Gli effetti degli investimenti in tecnologie digitali nelle scuole del Mezzogiorno* (Vol. 33). Collana Materiali Uval.

Gutstein, E. (2006). *Reading and writing the world with mathematics: Toward a pedagogy for social justice*. Routledge.

Heritage, M., & Wylie, C. (2018). Reaping the benefits of assessment for learning: Achievement, identity, and equity. *ZDM*, *50*(4), 729–741. https://doi.org/10.1007/s11858-018-0943-3

Hoogland, K., & Tout, D. (2018). Computer-based assessment of mathematics into the twenty-first century: Pressures and tensions. *ZDM*, *50*(4), 675–686. https://doi.org/10.1007/s11858-018-0944-2

Huang, X., Craig, S. D., Xie, J., Graesser, A., & Hu, X. (2016). Intelligent tutoring systems work as a math gap reducer in 6th grade after-school program. *Learning and Individual Differences*, *47*, 258–265. https://doi.org/10.1016/j.lindif.2016.01.012

Lerman, S. (2014). Socioeconomic Class in Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 553–558). Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_141

McConney, A., & Perry, L. B. (2010). Socioeconomic status, self-efficacy, and mathematics achievement in Australia: A secondary analysis. *Educational Research for Policy and Practice*, *9*(2), 77–91. https://doi.org/10.1007/s10671-010-9083-4

MIUR. (2012). Indicazioni Nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione.

Ng, C., Bartlett, B., & Elliott, S. N. (2018). *Empowering engagement: Creating learning opportunities for students from challenging backgrounds*. Springer.

Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: Responding to issues regarding methodology, policy, and equity. *ZDM*, *50*(4), 555–570. https://doi.org/10.1007/s11858-018-0963-z

OECD. (2020). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. OECD. https://doi.org/10.1787/d5f68679-en

Osadebe, P. U., & Oghomena, D.-E. (2018). Assessment of Gender, Location and Socio-Economic Status on Students' Performance in Senior Secondary Certificate Examination in Mathematics. *International Education Studies*, *11*(8), 98. https://doi.org/10.5539/ies.v11n8p98

Page, M. S. (2002). Technology-Enriched Classrooms: Effects on Students of Low Socioeconomic Status. *Journal of Research on Technology in Education*, *34*(4), 389–409. https://doi.org/10.1080/15391523.2002.10782358

Pellegrini, M., Vivanet, G., & Trinchero, R. (2018). Gli indici di effect size nella ricerca educativa. Analisi comparativa e significatività pratica. *Educational, Cultural and Psychological Studies*, *18*, 275–309. http://dx.doi.org/10.7358/ecps-2018-018-pel1

Rohn, D. (2013). Equity in Education: The Relationship Between Race, Class, and Gender in Mathematics for Diverse Learners. *Urban Education Research & Policy Annuals*, *1*(1), 13–22.

Sacchet, M. (2022). Ten Tips for Successful Creation of Contextualized Problems for Secondary School Students with Maple. *Maple Transactions*, *2*(1). https://doi.org/10.5206/mt.v2i1.14446

Sangwin, C. (2015). Computer Aided Assessment of Mathematics Using STACK. In S. J. Cho (Ed.), *Selected Regular Lectures from the 12th International Congress on Mathematical Education* (pp. 695–713). Springer. https://doi.org/10.1007/978-3-319-17187-6_39

Sangwin, C., Makar, K., Cazes, C., Lee, A., & Wong, K. L. (2010). Micro-level Automatic Assessment Supported by Digital Technologies. In C. Hoyles & J.-B. Lagrange (Eds.), *Mathematics education and technology: Rethinking the terrain: The 17th ICMI study* (Vol. 13, pp. 227–250). Springer.

Scherer, R., & Siddiq, F. (2019). The relation between students' socioeconomic status and ICT literacy: Findings from a meta-analysis. *Computers & Education*, *138*, 13–32. https://doi.org/10.1016/j.compedu.2019.04.011

Semana, S., & Santos, L. (2018). Self-regulation capacity of middle school students in mathematics. *ZDM*, *50*(4), 743–755. https://doi.org/10.1007/s11858-018-0954-0

Stacey, K., & Wiliam, D. (2013). Technology and Assessment in Mathematics. In M. A. Clements (Ed.), *Third International Handbook of Mathematics Education* (Vol. 27, pp. 721–751). Springer.

Suppes, P., Liang, T., Macken, E. E., & Flickinger, D. P. (2014). Positive technological and negative pre-test-score effects in a four-year assessment of low socioeconomic status K-8 student learning in computer-based Math and Language Arts courses. *Computers & Education*, *71*, 23–32. https://doi.org/10.1016/j.compedu.2013.09.008

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Research in Science Education, 48(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

United Nations. (2016). *Transforming our world: The 2030 agenda for sustainable development*. https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf

United Nations. (2022). *The Sustainable Development Goals Report 2022*. https://unstats.un.org/sdgs/report/2022/

Valli Jayanthi, S., Balakrishnan, S., Lim Siok Ching, A., Aaqilah Abdul Latiff, N., & Nasirudeen, A. M. A. (2014). Factors Contributing to Academic Performance of Students in a Tertiary Institution in Singapore. *American Journal of Educational Research*, *2*(9), 752–758. https://doi.org/10.12691/education-2-9-8

Wang, L., Li, X., & Li, N. (2014). Socio-economic status and mathematics achievement in China: A review. *ZDM*, *46*(7), 1051–1060. https://doi.org/10.1007/s11858-014-0617-8

Wiliam, D. (2006). Formative Assessment: Getting the Focus Right. *Educational Assessment*, *11*(3–4), 283–289. https://doi.org/10.1080/10627197.2006.9652993

Wright, P. (2016). Social justice in the mathematics classroom. *London Review of Education*, *14*(2), 104–118. https://doi.org/10.18546/LRE.14.2.07

Yang Hansen, K., & Strietholt, R. (2018). Does schooling actually perpetuate educational inequality in mathematics performance? A validity question on the measures of opportunity to learn in PISA. *ZDM*, *50*(4), 643–658. https://doi.org/10.1007/s11858-018-0935-3

Yerushalmy, M., Nagari-Haddif, G., & Olsher, S. (2017). Design of tasks for online assessment that supports understanding of students' conceptions. *ZDM*, *49*(5), 701–716. https://doi.org/10.1007/s11858-017-0871-7

Zhu, Y. (2018). Equity in Mathematics Education: What Did TIMSS and PISA Tell Us in the Last Two Decades? In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu (Eds.), *Invited Lectures from the 13th International Congress on Mathematical Education* (pp. 769–786). Springer International Publishing. https://doi.org/10.1007/978-3-319-72170-5_43