Game-based education promotes sustainable water use

Ennio Bilancini (ennio.bilancini@imtlucca.it)
IMT School for Advanced Studies Lucca
Leonardo Bconcinelli
University of Florence
Roberto Di Paolo
IMT School for Advanced Studies Lucca, University of Alicante https://orcid.org/0000-0002-6081-6656

Analysis

Keywords: Water consumption, Prosocial behaviour, Field quasi-experiment, Game based learning, Games for social change, Primary school.

Posted Date: May 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-490913/v1

Game-based education promotes sustainable water use

Ennio Bilancini* Leonardo Boncinelli† Roberto Di Paolo‡

Abstract

In this study we estimate the impact of a game-based educational program aimed at promoting sustainable water usage among 2nd-4th grade students and their families living in the municipality of Lucca, Italy. To this purpose we exploited unique data from a quasi-experiment involving about two thousand students, one thousand participating (the treatment group) and one thousand not participating (the control group) in the program. Data were collected by means of a survey that we specifically designed and implemented for collecting students' self-reported behaviors. Our estimates indicate that the program has been successful: the students in the program reported an increase in efficient water usage and an increase in the frequency of discussions with their parents about water usage; moreover, positive effects were still observed after six months. Our findings suggest that game-based educational programs can be an effective instrument to promote sustainable water consumption behaviors in children and their parents.

Keywords: Water consumption, Prosocial behaviour, Field quasi-experiment, Game-based learning, Games for social change, Primary school.

Acknowledgements: The authors gratefully acknowledge the financial support of the IMT School for Advanced Studies Lucca through the PAI project Pro.Co.P.E. "Prosociality, Cognition, and Peer Effects". Also, we sincerely want to thank Lucca Crea s.r.l. and the Provveditorato degli Studi di Lucca for their kind support and essential advise.

*IMT School of Advanced Studies, Piazza S.Francesco 19, 55100 Lucca, Italia. Tel.: +39 0583 43 26 737, email: `ennio.bilancini@imtlucca.it`.

†Department of Economics and Management, University of Florence, Via delle Pandette 9, 50127 Firenze, Italia. Tel.: +39 055 2759578, fax: +39 055 2759910, email: `leonardo.boncinelli@unifi.it`.

‡Department of Economics, University of Alicante; IMT School for Advanced Studies Lucca. Email: roberto.dipaolo@imtlucca.it.

1 Introduction

Sustainable water consumption is relevant for the general sustainability of current and future societies (Wada and Bierkens, 2014; Kummu *et al.*, 2016; Liu *et al.*, 2017; Greve *et al.*, 2018). Sustainable water consumption is, in many cases, an instance of prosocial behaviour in a social dilemma (Hardin, 1968): a situation in which a conflict exists between maximizing one's individual benefits and maximizing the benefits of the present and future generations. Individuals who are purely self-interested are less likely to adopt the prosocial behaviors that lead to sustainable water consumption, unless social norms exert sufficient social pressure to push self-interested individuals to do otherwise. Since the acquisition of preferences for prosocial behaviors as well as the internalization of social norms take place, in a substantial part, during childhood (House and Tomasello, 2018; House *et al.*, 2020), it becomes a critical goal to create opportunities for young children to develop such preferences and internalize norms of sustainable water consumption (Copple *et al.*, 2013; Cobo-Reyes *et al.*, 2020).

Early childhood education is the natural starting point for a life-long learning. During the past years, a variety of educational methods to promote prosociality in children have been successfully implemented. These methods include play space, multi-use toys, dedicated books, group play, and organized gaming (Orlick, 1983). In particular, the kind of social interactions that come from group play and organized gaming, as well as the time that gaming can occupy in children's daily lives, make game-based educational programs a natural candidate tool for promoting desirable behaviors. Some studies, in recent years, evaluated the relevance of programs which encourage good practices in environmental benchmarks, such as the use of water (Niles *et al.*, 2013; Cuadrado *et al.*, 2017). In a field experiment (Schultz *et al.*, 2016) the role of social norms in promoting water conservation was studied, finding that people who received normative information about similar household in their neighborhoods consumed less water than the control group; moreover, people with already

strong personal norms were less affected by the normative information than those with low personal norms. Importantly, children are able to recognize if prosocial norms apply to specific situations (Blake *et al.*, 2015), so that it becomes important that children understand what is sustainable water consumption and can relate their behavior to concrete and specific situations such as water collection or body washing.

In this paper we provide evidence regarding the effectiveness of a game-based educational program implemented during the first eleven months of the year 2019 in the municipality of Lucca, Italy. The program was named *BLUTUBE: Who brings the water home* and was aimed at promoting sustainable water consumption as well as awareness about the municipal water system and its usage. The targets of the program were about 1000 students from 2nd-4th grades and their families. The program relied primarily on ludic engagement for the specific objectives of improving students' awareness about the water cycle in nature, the water system of the municipality of Lucca, and the daily usage of water.

Our approach to the empirical assessment of the program's impact is based on the quasi-experiment methodology (Campbell and Stanley, 2015): we had no possibility to intervene directly on the organization of the program, but we were able to implement a simple two-group design (treatment and control) and collect three distinct measurements of target outcome variables over a period of eleven months. In particular, we elicited the students' awareness and their behaviors about water consumption with three waves of surveys administered, respectively, immediately before the program started, some days after the main activities were over, and after six further months. Responses to this kind of questionnaires have been shown to be a reliable source of information on children's perspectives and perceptions (Danielson and Phelps, 2003).

Our main finding is that the program had a positive impact on the awareness of water usage. This effect is primarily driven by an increase in the frequency of self-reported virtuous behaviors regarding water consumption and discussions with parents about water. Moreover,

such positive effect appears to be persistent: six months after the end of the main activities of the program the effect is still positive and of appreciable size.

2 Results

Our final sample consists of 52 classes in the treatment group (one class envelope was lost during the collection process) and 53 in the control group, for a total of 105 classes and 5273 questionnaires (up to three per student). Figure 1 reports the timing of the program and the three survey waves used to measure reported behavior. The pre-program wave involved 869 students in both treatment and control group. The post-program wave involved 895 students in the treatment group and 860 students in the control group. After 6 months, the last wave involved 908 students in the treatment group and 872 in the control group. The final sample consists of 5273 questionnaires (see Table 2 in Appendix A).

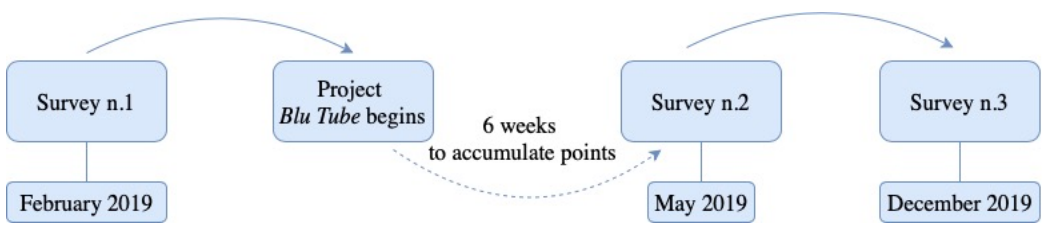

Figure 1: Timeline of quasi-experimental study of the intervention program.

Summary statistics by treatment and control groups for the pre-program survey show that the two groups are not well balanced (see Table 3 in Appendix A): while the difference in the number of students per class is only marginally not statistically significant ($Z = -1.95, p = .051$), the difference in the measured students' cognitive skills is statistically significant ($Z = -2.30, p = .031$) as well as the distribution of grades ($Z = 4.99, p < .0001$). These differences are mainly due to the fact that the distribution of students across grades is quite different between the treatment and the control group (for the 2nd grade there are 818 students in the treatment group and 855 in the control group; for the 3rd grade there are 621 students in the treatment group and 1123 in the control group; for the 4th grade

there are 1162 students in the treatment group and 694 in the control group). In the light of this, we checked whether there is any difference in the reported behavior in the pre-program survey. Importantly, there is no statistically significant difference in the aggregate reported behavior between the control and the treatment group ($Z = -1.30, p = .193$). Aggregate reported behavior is constructed summing up the answers to all 7 questions of relevance here, so that (with a Likert scale going from 1 to 5) the aggregate variable ranges from a minimum of 7 (least sustainable reported behavior) to a maximum of 35 (most sustainable reported behavior).

We also looked at the distribution of answers in the pre-program survey for each of the 7 questions, testing for statistically significant differences. In four cases we found that the distribution of answers are not statistically different between the treatment and the control group, namely: *Shower* ($Z = -0.18, p = .849$); *Fountain* ($Z = 0.84, p = .397$); *Vegetables* ($Z = -0.69, p = .488$); *Waste* ($Z = 1.69, p = .091$), while in 3 cases we found statistically significant differences: *Teeth* ($Z = -3.05, p = .002$), *Hands* ($Z = -2.36, p = .018$) and *Parents* ($Z = -2.27, p = .023$).

In the light of these results we adopt a two-step strategy. First, we carry out a non-parametric analysis of the treatment effect on the aggregate reported behavior. This is possible because, although the treatment and control groups are not perfectly balanced, the aggregate variable comes with similar levels in the two groups for the pre-program survey. We then check the robustness of non-parametric results by running regressions for each wave, including controls for the sample characteristics in order to correct for the lack of sample balancedness.

Second, we study the treatment effect on the reported behavior for each of the 7 questions using ordered logit regressions where we pool all data and we control for sample characteristics, the 3-survey structure, and their interaction with the treatment. This allows us to obtain indications about the source of the treatment effects estimated at the aggregate level,

taking into account the fact that some reported behaviors do not come with similar levels in the pre-program survey. Also, we previously carry out a non-parametric analysis of the treatment effect for each of the 7 questions in order to give a complete picture about the differences in reported behavior across both the three surveys and the treatment and control groups.

Finally, one might wonder if the answers to the 7 questions can be accounted for by a few common factors. Correlation analysis and principal component analysis suggest that this is not quite the case (see Appendix B.1).

2.1 Aggregated reported behavior

Figure 2 reports the cumulative distribution function of the aggregated reported behavior in the three waves (pre-program, post-program, and post6-program, i.e., 6 months after post-program) for both control and treatment groups. While the distributions of treatment and control groups in the pre-program do not appear to be different, in the post-program and post6-program the distributions of the treatment group are shifted to the right; in particular, the distribution of the treatment group appears to first order stochastically dominate the distribution of the control group. Epps-Singleton test of the equality of the distributions confirms this: we reject the hypothesis that the distributions of treatment and control groups are the same in both the post-program survey and the post6-program survey ($W2 = 62.243, p < .001$ and $W2 = 30.943, p < .001$, respectively), while we cannot reject the hypothesis that the distributions of treatment and control groups are the same in pre-program survey ($W2 = 2.331, p = .675$).

In Figure 3 the means of the aggregated reported behavior are compared between treatment and control groups, by survey wave. No statistically significant difference is found for the pre-program survey ($Z = -1.300, p = 0.193$). In contrast, for the post-program we find that the treatment group has a statistically higher average of about 2.11 with

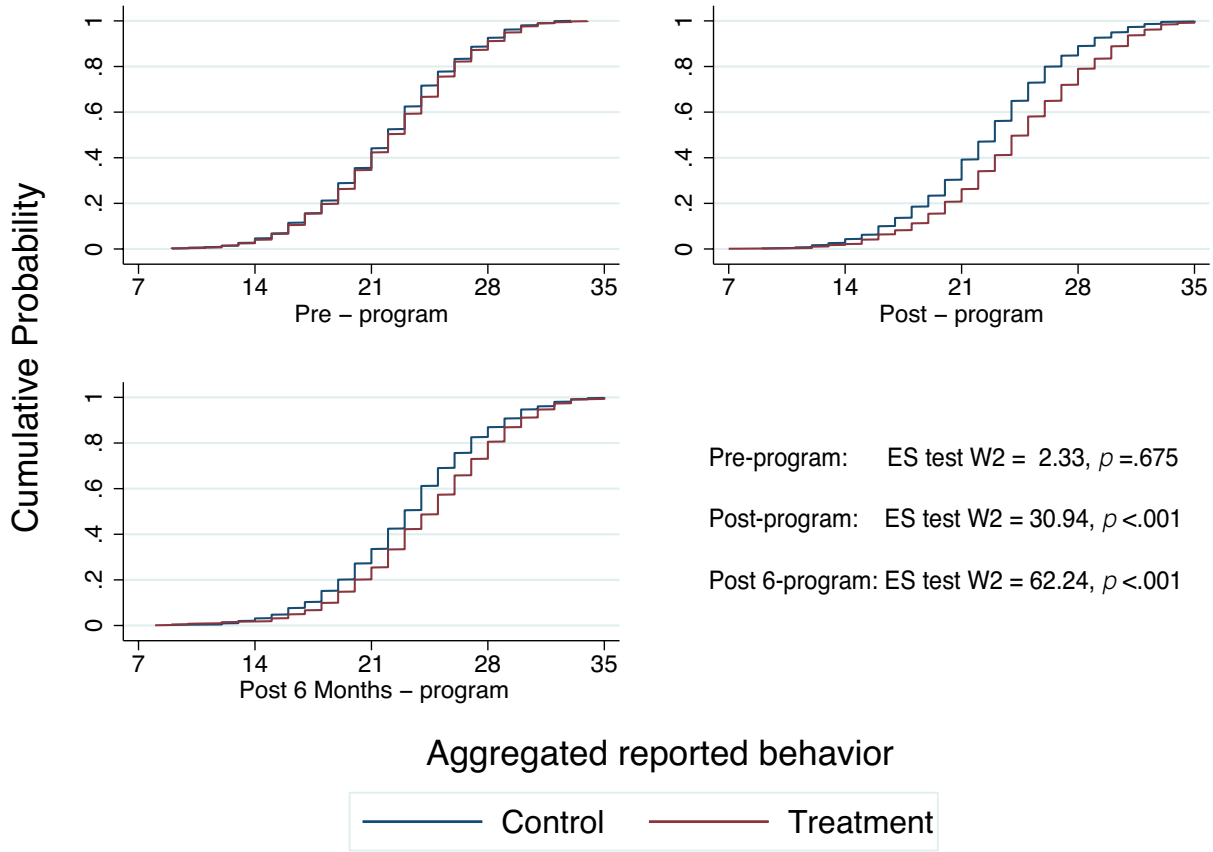

Figure 2: Cumulative distribution function of the aggregated reported behavior by conditions and waves. Distributions in the post-program and 6 months after are shifted to the right in the treatment group, with a statistically significant differences between conditions. ES stands for Epps-Singleton test.

respect to the pre-program treatment group ($Z = -9.055, p < 0.001$) and a statistically significant higher average of about 1.72 with respect to the post-treatment control group ($Z = -7.479, p < 0.001$). These numbers range from 1.32% to 7.04% of the pre-program average, suggesting that the treatment has had a substantial impact between the pre-program and the post-program surveys.

Furthermore, Figure 3 shows that there is no appreciable difference between the aggregated behavior in the treatment group between the post-program survey and the post6-program survey ($Z = 0.165$, $p = 0.869$). Also, although the the average aggregated behavior of the control group increases of about 0.56 points between the post-program and the post6-

program surveys, we still find a statistically significant difference between the treatment and the control groups in the post6-treatment survey ($Z = 5.271$, $p < 0.001$). Together, these findings suggest that the effect of the treatment is persistent, at least until the official end of the program (about 9 months after its start).

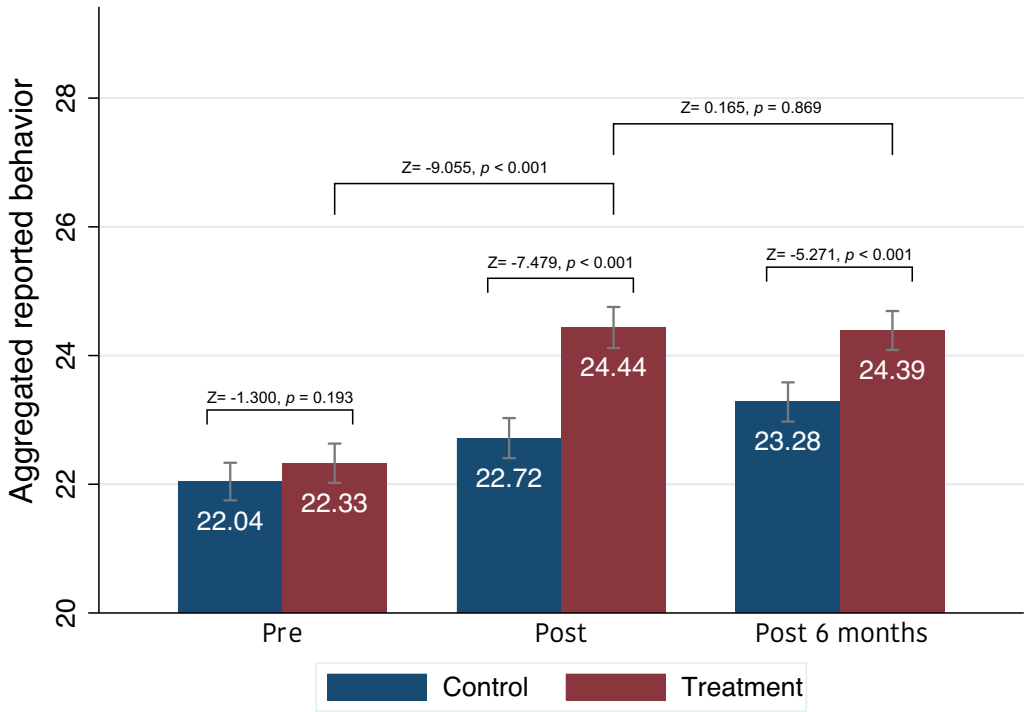

Figure 3: Average of the aggregated reported behavior by conditions and waves. In the pre-program period, the aggregated reported behavior in the treatment group is not significantly different from the control group (Mann-Whitney test, $Z = -1.300, p = 0.193$). In the post-program period and after 6 months, the aggregated reported behavior in the treatment group is significantly higher respect to the control group (Mann-Whitney test, $Z = -7.479, p < 0.001$ and $Z = -5.271, p < 0.001$, respectively). The treatment effect is stable after 6 month (Mann-Whitney test, $Z = 0.165, p = 0.869$). Error bars represents the 95% confidence interval.

The findings described above rely on the assumption that the lack of balance between treatment and control groups did not bias our estimates. In order to control for such potential problem we run linear regression models where aggregated reported behavior is predicted by the treatment and a number of controls. Importantly, since students came from different schools and classes, and that in one school there is the possibility to have more than one

class treated, we are able to control for schools including school fixed effects. In addition, besides a dummy variable for the treatment (which is equal to 1 if the student belongs to the treatment group), we include a dummy for the grade (omitted category is 2nd grade), an index of cognitive skills (fraction of correct answers in logical/mathematical questions), and the number of students in the class. We run similar regressions for the pre-program, the post-program, and the post6-program surveys. Results are reported in Table 1.

Table 1: Linear Fixed Effect Regression

	Pre		**Post**		**Post 6 Months**	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	.534	1.13	1.89***	2.17***	1.12**	1.27*
	(.379)	(.593)	(.394)	(.550)	(.384)	(.594)
3rd Grade		1.07		.927		1.21
		(.610)		(.574)		(.703)
4th Grade		2.05***		1.65***		1.10*
		(.418)		(.443)		(.492)
TR × 3rd		-.811		-.481		-.645
		(.802)		(.782)		(.809)
TR × 4th		-.088		.182		.018
		(.770)		(.736)		(.776)
Cognitive Skills		.187		-.011		.647*
		(.227)		(.233)		(.261)
Students		-.002		.055		.092
		(.045)		(.038)		(.048)
Constant	21.9***	20.6***	22.6***	20.7***	23.3***	20.5***
	(.262)	(.830)	(.251)	(.727)	(.268)	(.793)
N	1685	1685	1732	1732	1765	1765

The dependent variable is the aggregated reported behavior on good/bad practices of water usage. *Treatment* is equal to 1 if the students are in the treatment group, 0 otherwise. *Grade* is the students' year group. *2nd Grade* is the reference category. *Cognitive Skills* is equal to 1 if the result obtained in the logical and mathematical questions are higher than the median, 0 otherwise. *Students* is the number of students in each class. In all cases, we control for school fixed effects. Standard errors (in parenthesis) are clustered at class level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Column (1) reports the results of the school-fixed effect regression in the pre-program survey. No statistical significant effect of the treatment is found in this case. The result is confirmed by the estimates reported in Column (2) where the regressors include controls for

the grade of the students, their interaction with the treatment, an index of cognitive skills and the number of students in each class. Among these, the only regressor with statistically significant coefficient is *4th Grade*, suggesting that it may be the source of potential pre-program differences in reported behavior.

Column (3) reports the results of the school-fixed effect regression in the post-program survey. The coefficient of the treatment variable is positive (1.89) and statistically significant, confirming the results of the non-parametric test. Similar results are found in column (4) where the regressors include the controls used for the regression in column (2). In particular, the coefficient of the treatment variable is positive (2.17) and statistically significant. Again, the only regressor with statistically significant coefficient is *4th Grade*, in line with the idea that it may be the source of potential pre-program differences in reported behavior.

Column (5) reports the results of the school-fixed effect regression in the post6-program survey. The coefficient of the treatment variable is positive (1.12) and statistically significant, somewhat lower than in column (3). This confirms the result showed in the non-parametric test that the effect of the program is persistent after 6 months, although it may be of reduced magnitude. Similar results are found in column (6) where the regressors include the controls used for the regression in column (2) and (4). Specifically, the coefficient of the treatment variable is positive (1.27) and statistically significant. Again, the coefficient of *4th Grade* is positive and statistically significant, but in this case it is not the only one: also the coefficient of *Cognitive Skills* is positive and statistically significant. The sum of the these two coefficients is about of the same magnitude that the one of *4th Grade* in column (4), suggesting that in the longer run cognitive skills might be a substitute for grade seniority.

2.2 Disaggregated reported behaviors

Figure 4 reports the means of reported behaviors for each of the 7 questions comparing treatment and control groups, by survey wave. As already noted in Table 3, 3 out of 7

reported behaviors (*Teeth*, *Hands*, and *Parents*) appear to be statistically different in the pre-program survey, with the treatment group coming with a higher mean.

Looking at the differences between treatment and control groups in the post-program survey, we find that 4 out of 7 variables show a statistically significant difference, with a higher mean for the treatment group: *Teeth* ($Z = -4.248, p < 0.001$); *Fountain* ($Z = -3.149, p = 0.0016$) ; *Hands* ($Z = -5.429, p < 0.001$); *Parents* ($Z = -6.115, p < 0.001$) and *Waste* ($Z = -5.284, p < 0.001$). Moreover, 3 of these 4 variables appear to be statistically different also in the post6-program survey: *Teeth* ($Z = -2.587, p = 0.009$); *Hands* ($Z = -5.020, p < 0.001$) and *Parents* ($Z = -3.881, p = 0.001$); in addition, we also find a statistically significant difference for the variable *Shower*, again with a higher mean in the treatment group ($Z = -5.125, p < 0.001$).

In order to control for potential confounding factors that potentially persisted across the three waves – and which could explain the differences described above – we pool data of the three survey waves and we run ordered logit regressions for each of the 7 variables, also adding the control variables used in the analysis of aggregated reported behavior. In this case we prefer not to use a liner regression models because of the 5-tier ordinal structure of answers.

Figure 5 reports the estimates of the relevant coefficients of the ordered logit regressions (detailed estimates can be found in Table 6 in Appendix B). Specifically, the coefficients of interests are those of the interactions between *Treatment* and *Post* (the treatment effect just after the end of the program) and between *Treatment* and *Post 6* (the treatment effect 6 months after the end of the program), whereas the base of reference is the control group in the pre-program survey. According to this analysis the program has had a positive effect on *Fountain*, *Hands*, *Parents* and *Waste*. These effects are still detectable after six months for *Fountain* and *Waste*, when also a positive treatment effect on *Shower* is found.

These results suggest that the program has had a positive effect especially on two di-

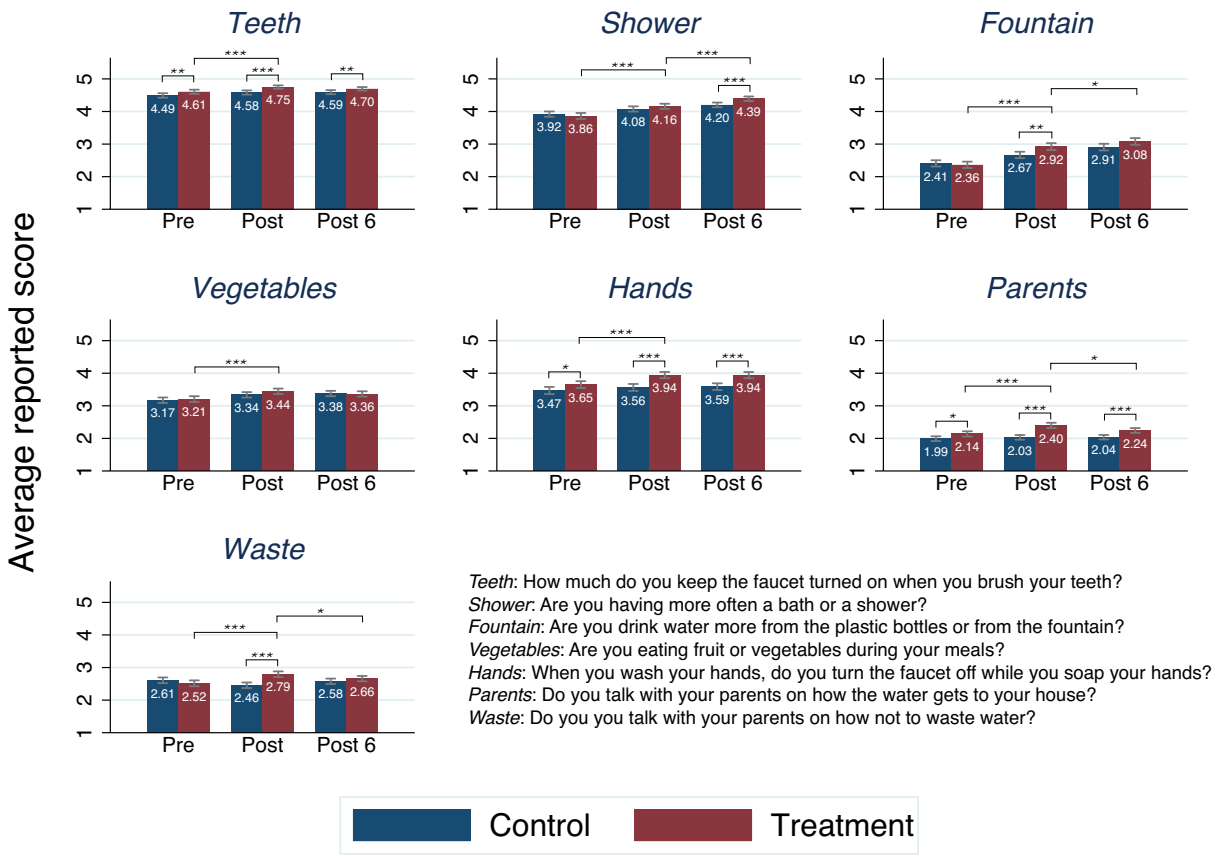

Figure 4: Average reported behavior by questions, conditions and survey wave. Each answer assume values from 1 to 5. Questions are reported in the figure. Statistically significant difference between conditions are reported above columns (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Error bars represent the 95% confidence interval.

mensions, namely the habits and behaviors that involve massive or frequent use of water (full body washing, hands washing, drinking) and the discussions with parents about water (from where it comes, how not to waste it), while other dimensions involving more indirect or limited use of water (eating products requiring water to be produced, teeth brushing) seem to have been less affected. Moreover, while the effect on the discussions with parents seems to have faded away towards the end of the program, the effect on the habits and behaviors that involve massive or frequent use of water seems to have persisted beyond the end of the program.

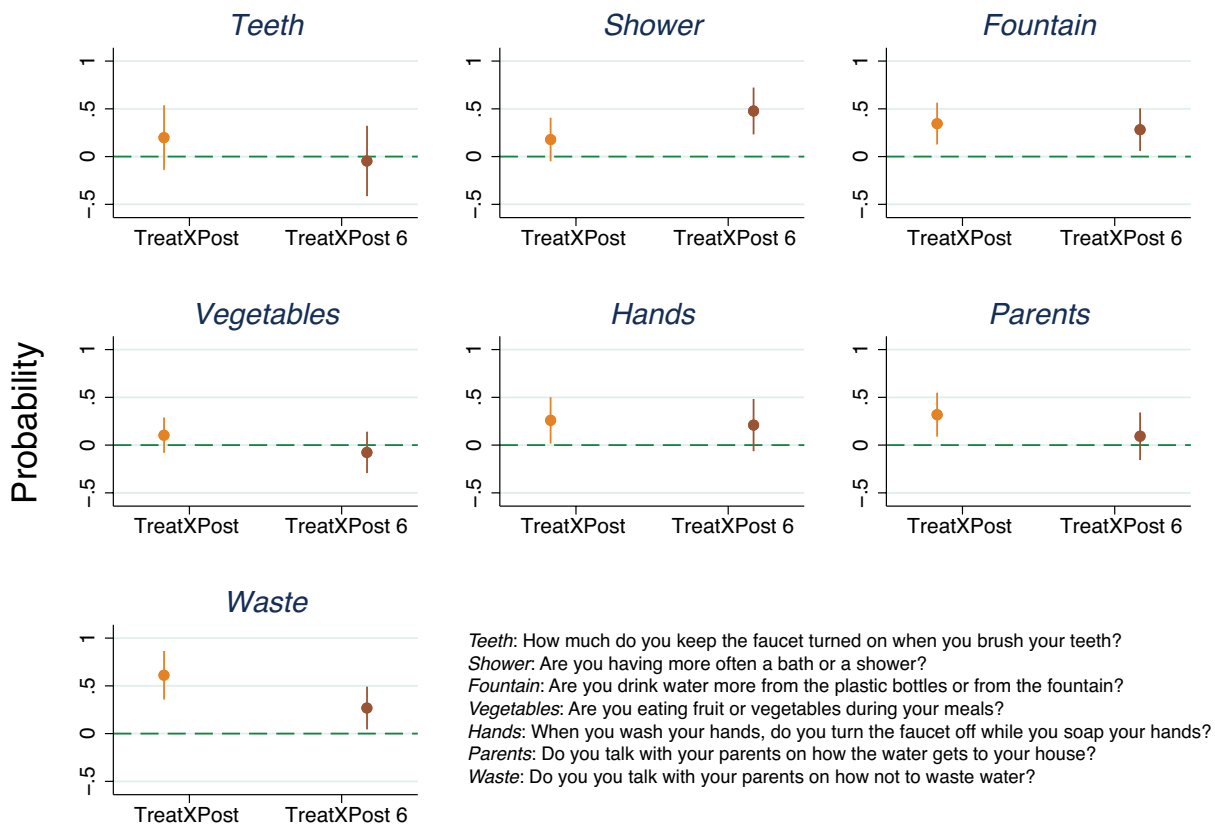

Figure 5: Estimated coefficient of the Ordered Logit regression in Table 6 in Appendix B.2. The dependent variables are the 7 questions, which assume values from 1 to 5. Questions are reported in the figure. Error bars represent the 95% confidence interval.

3 Discussion

Our results provide field evidence about the effectiveness of promoting sustainable behaviors regarding water consumption by means of game-based educational programs. Our analysis exploited a unique dataset built from a quasi-experiment involving about two thousand Italian students of 2nd-4th grades, all from the same municipality (Lucca, Italy). Specifically, our findings suggest that the program has had positive, sizeable and persistent effects, especially with regard to habits and behaviors that involve massive or frequent use of water (full body washing, water drinking). We believe that such evidence strongly pushes towards

a greater consideration of game-based education programs as policy instruments to promote sustainable habits and behaviors, especially when children and their families can be targeted.

It is worth emphasizing that the program had not just provided a chance to play with sustainability-themed games. Instead, the structured ludic activities were designed to engage students in specific settings (at home, at school, during time spent with the family) and this was properly incentivized in terms of the game rewards that materialized over a rather long period of time (several months). The resulting take-home message is that game-based programs aiming at promoting sustainable behaviors should be designed to engage participants in their daily life, for a substantial length of time, and with social activities involving people with whom they have stable relationships.

One important aspect of our results which deserves to be highlighted is that the decline in the treatment effect over the last part of the program is entirely due to an improvement of the reported behavior in the control group, and not to a progressive deterioration of the reported behavior in the treatment group – which actually does not decline. This dynamic could have at least two different sources. One is independent learning by students over the nine months of the program, which might have led students to improve their behaviors over time just through standard channels which have nothing to do with the program and that are common to all classes and schools. Some evidence of this is found in the positive correlation between the 4th grade and virtuous behaviors. If this is the correct explanation, then the program has had in part the effect of accelerating such learning in the first months, implying a deceleration in the last months. Another source of explanation is the presence of peer effects beyond students' own classes, that is, students in the control group might have been exposed indirectly to the program through their social connections outside their own classes. This latter explanation would imply that the treatment effect is far larger that our estimates indicate. With our data we cannot establish which explanation works better. Additional specific data have to be collected for this purpose.

A standard limitation of quasi-experiment is that, since the randomization protocol cannot be managed directly, one cannot conclude about the causal effect of the treatment. We think that such limitation, although not absent, is less severe in our study because the assignment procedure was largely exogenous to students' and teachers' desires, with constraints for eligibility and required participation that left little room for self-selection. Moreover, we could control for systematic differences in the characteristics of control and treatment groups, such as grade, cognitive abilities, class size, and school.

Another limitation of this study is that we could only use self-reported behavior and not directly observe relevant behaviors. Unfortunately, it turned out that the observation of direct water consumption by the families involved in the program was impossible, mostly due to the absence of a reliable way to collect these data, either from the local water utility or from the families themselves.

Perhaps the most important limitation of this study is the fact that we were not allowed to connect individual response in the three surveys for the control group (while we could do so for the treatment group since the names of the students were public). This has forced us to rely on class averages to get a longitudinal structure of the data, greatly reducing the statistical power and necessarily limiting the scope of our analysis (e.g., we could not properly exploit individual characteristics). We cannot do much in this regard if not stressing that such information should be made a priority in future studies .

Starting from the results of this study there are at least three avenues of future research that seem promising. Firstly, one may dig into the collected data regarding ludic habits and preferences to see whether these modulate the effects of the program, and whether they are affected by the participation in the program. Ludic habits and preferences are important for students' wellbeing and life-long learning. Secondly, one may want to run follow-up field experiments with the aim of observing actual behavior regarding water use. This can only be done with a substantial smaller number of students, but full randomization is likely to

be more easily implementable in such a case. Lastly, one may want to run similar studies employing game-based educational programs aimed at promoting different sustainable behaviors and habits, such as waste production, recycling, and energy consumption, in order to check to what extent our results can be generalized.

4 Methods

4.1 The game-based educational program

The program was designed and implemented by the Provincial Education Office of Lucca (Provveditorato agli Studi), Lucca Crea s.r.l. (a company 100% owned by the municipality of Lucca which is in charge of organizing and managing cultural events),[1] and GEAL s.p.a. (the water utility company of the municipality of Lucca).

The program was titled "*BLUTUBE - Chi porta l'acqua a casa*" (*BLUTUBE - Who brings the water home*) and had its main engine made of gaming activities, for which an urban and a board game were developed ad hoc by Lucca Crea and its collaborators, also in partnership with GEAL and the municipality of Lucca. The gaming activities were tuned to fit 2nd, 3rd, and 4th grades students from the primary schools in Lucca. The main aim of the program was to bring about greater awareness of the daily use of water resources and their sustainable consumption together with knowledge of the integrated water system of the municipality of Lucca and the water cycle in general. Games and gaming activities were specifically designed for this purpose, although the board game (also named BLUTUBE) was designed to be playable, and enjoyable, as a stand alone game too (more details on the games can be found in Appendix C).

The program was divided in three distinct phases. The first phase was titled *How not to dodge in a glass of water*. In a given day a group of educators, specifically selected for

[1] Lucca Crea s.r.l. primary job is to organize Lucca Comics & Games, one of the largest transmedia shows in the world focusing on comics, games and pop culture.

the program, went to each class participating in the program to give a short talk on the importance of water resources and their consumption as well as to explain the working of the gaming activities (program phases, allocation of game points, publication of rankings) and, in particular, to teach students how to play the board game BLUTUBE. Moreover, each student got its own box of the board game (for playing at home) and each class was also endowed with a copy of the board game (for playing in class).

The second phase was titled *Bring the water to your mill* and lasted 6 weeks during which the students participating in the program had the chance to play as much as they wanted, and accumulate points accordingly, for two distinct rankings: the individual ranking and the class ranking. There were four different ways to obtain points:

- *playing the board game BLUTUBE at school:* each student can play during school time. The teacher records each time a student plays on a scoreboard and each week a picture is sent to the program organizers. For each recorded play a student earns 10 points, up to a total of 2500 for the whole phase also considering the points earned for playing at home (see below);

- *playing the board game BLUTUBE at home:* each student can play at home with their family or friends and gain points every time they send a picture of the playing to the program organizers, also indicating the name, the surname, the school and the class. For each appropriate picture sent a student earns 10 points, up to 2500 points in total also considering the points earned for playing at school (see above);

- *visiting the "hidden water places" in Lucca:* each student can visit, together with parents or other family members, a number of specific places labelled as "water places" in municipality of Lucca. Such places are reported in the map describing the program and distributed at the beginning with the board game. A student can send to the program organizers a picture proving a visit in one distinct water place indicated in

the map, also indicating the student's name, the surname, the school and the class. For each appropriate picture sent the student earns 150 points, up to 2500 points in total.

- *providing evidence of sustainable behavior:* each student can send to the program organizers a picture where the student is making a sustainable use of water, e.g., eating vegetables, fulfilling the can at the fountain, turning the faucet off when they are brushing their teeth. The picture has also to indicate the student's name, the surname, the school and the class. A student gains between 10 to 200 points for each appropriate picture, depending on the actual behavior, up to 5000 points in total.

Starting from the second week of the second phase both individual and class scores were published in a dedicated website and in local newspapers. In this way, the participating students, their parents, and others in their schools could see their weekly progress and compare their scores with those of other participants.

The last phase of the program was titled *BLUTUBE Tournaments* and consisted in a tournament with restricted participation where the only way to accumulate points was playing with the board game BLUTUBE. Specifically, the 16 classes with the highest total score in the second phase (among the 53 classes participating) were selected to participate in four distinct group stage tournaments (each comprising 4 of the 16 classes). The winner of each group stage tournament qualified to participate in the final stage tournament which took place during the Lucca Comics and Games festival held in 2019. The final stage tournament allowed to win a full paid holiday trip themed "Environment", where students could learn methods to create electricity through the use of heat while respecting the environment.

All activities related to the game-based educational program had been carried out between January and November, 2019. The participation protocol was as follows. Most primary schools in the municipality of Lucca were involved. Actual participation in the program was

determined at the class level, under consent by the school head teacher. Lucca Crea, which was in charge of promoting the program across the schools, talked to the head teacher of each school asking for classes who were available to participate in the program. In most cases, the decision about whether to participate or not was taken by the head teacher of each class, and in no case there was a possibility for the students of the class to affect such decision, which was made on the basis of the overall workload of the class in terms of extra-curricular activities. A few remarks are worth doing. First, the participation protocol led to a situation where in the same schools there were classes which participated and classes which did not participate. Second, participation was exogenous to the students' desire to participate. Third, actual participation was often exogenous to the teachers' desire to participate too. This is because the teachers' decision was often constrained by the fact that their class was already involved in a number of extra-curricular activities, and hence could not actually participate, or by the fact that it had to add extra-curricular activities with the program being the only possibility, and hence it was actually forced to participate.

This participation protocol allows the applicability and effectiveness of our method of analysis, in that the assignment to the program, although not fully randomized, is to a good extent exogenous to schools, students' and teachers' preferences.

4.2 Data and empirical strategy

The program described in Subsection 4.1 qualifies as a natural quasi-experiment (Meyer, 1995) for which we designed a pre/post control-treatment study that we implemented using a questionnaire (designed ad hoc) administered three times: just before the program, immediately after the end of phase two, and then again at the end of the program (six months later).

The study includes 28 primary schools. From those schools, 53 classes were directly involved in the program, forming the treatment group. For the control group we selected

other 53 classes that were not directly involved in the program, trying to build the best possible counterfactual. This was not an easy task because the total of 106 classes covers about the 90% of the entire population of 2nd-4th grades students in the municipality of Lucca (the overall number of classes being 116). So, together the treatment and control groups represent almost the entire student's population.

Students' awareness about the efficient use of water was elicited by means of a paper-based survey regarding students' behaviours and habits related to water use and consumption (the original and the English-translated questionnaires can be found in Appendices C.3 and D, respectively). Specifically, the survey contained seven distinct questions about water consumption in familiar circumstances, the extent to which students talk about water with their parents, and the extent to which students eat food containing water (fruit and vegetables). These questions are: "How much do you keep the faucet turned on when you brush your teeth?"; "Are you having more often a bath or a shower?"; "Do you drink water more from plastic bottles or from fountains/faucets?"; "Are you eating fruit or vegetables during your meals?"; "When you wash your hands, do you turn the faucet off while you soap your hands?"; "Do you talk with your parents on how the water gets to your house?"; "Do you talk with your parents on how not to waste water?". Answers were recorded using a 1-to-5 Likert scale.

The survey also contained questions related to relational activities, ludic habits and ludic preferences, that we do not exploit in the following analysis as they were meant for different research purposes. In addition, we tried to measure cognitive skills using logical and mathematical questions taken from the tests produced by the INVALSI (Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione) and the ones developed by TIMSS (Trends in Mathematics and Science Study).

The first survey was collected during February 2019, before the beginning of the program. The parents of students involved signed an informed consent form, with the specific consent

for the possibility to link students' answers to their scores in the program. Teachers received only general information about the research project, and specifically no details about what we were trying to elicit. The second survey was administered at the end of the second phase, during the month of May 2019. The survey was identical to the previous one but for the questions aiming at eliciting cognitive skills which we opted to substitute with new ones of comparable difficulty. To ensure consistency, the second survey was administered to the classes involved following the same procedures as in the first wave. Lastly, a third survey was administered six months after, when the program was officially over. This last survey was identical to the previous two but for the questions aiming at eliciting cognitive skills. Also in this case the survey was administered to the classes involved following the same procedures as in the first two waves.

References

Blake, P. R., Piovesan, M., Montinari, N., Warneken, F. and Gino, F. (2015). Prosocial norms in the classroom: The role of self-regulation in following norms of giving. *Journal of Economic Behavior & Organization*, **115**, 18–29.

Bucciol, A. and Piovesan, M. (2011). Luck or cheating? a field experiment on honesty with children. *Journal of Economic Psychology*, **32** (1), 73 – 78.

Campbell, D. T. and Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research*. Ravenio Books.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A. and Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological science*, **11** (4), 302–306.

Cobo-Reyes, R., Dominguez, J. J., García-Quero, F., Grosskopf, B., Lacomba,

J. A., Lagos, F., Liu, T. X. and Pearce, G. (2020). The development of social preferences. *Journal of Economic Behavior & Organization*, **179**, 653–666.

Cohn, A. and Maréchal, M. A. (2016). Priming in economics. *Current Opinion in Psychology*, **12**, 17–21.

Copple, C., Sigel, I. E. and Saunders, R. (2013). *Educating the young thinker: Classroom strategies for cognitive growth*. Routledge.

Cuadrado, E., Tabernero, C., García, R., Luque, B. and Seibert, J. (2017). The role of prosocialness and trust in the consumption of water as a limited resource. *Frontiers in Psychology*, **8**, 694.

Danielson, C. K. and Phelps, C. R. (2003). The assessment of children's social skills through self-report: A potential screening instrument for classroom use. *Measurement and Evaluation in Counseling and Development*, **35** (4), 218–229.

Funtowicz, S. O. and Ravetz, J. R. (1995). Science for the post normal age. In *Perspectives on ecological integrity*, Springer, pp. 146–161.

Gill, D. and Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, **124** (6), 1619–1676.

Greve, P., Kahil, T., Mochizuki, J., Schinko, T., Satoh, Y., Burek, P., Fischer, G., Tramberend, S., Burtscher, R., Langan, S. *et al.* (2018). Global assessment of water challenges under uncertainty in water scarcity projections. *Nature Sustainability*, **1** (9), 486–494.

Hardin, G. (1968). The tragedy of the commons. *science*, **162** (3859), 1243–1248.

House, B. R., Kanngiesser, P., Barrett, H. C., Broesch, T., Cebioglu, S., Crittenden, A. N., Erut, A., Lew-Levy, S., Sebastian-Enesco, C., Smith,

A. M. *et al.* (2020). Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour*, **4** (1), 36–44.

— and Tomasello, M. (2018). Modeling social norms increasingly influences costly sharing in middle childhood. *Journal of experimental child psychology*, **171**, 84–98.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, **47** (1), 5–86.

Kummu, M., Guillaume, J. H., de Moel, H., Eisner, S., Flörke, M., Porkka, M., Siebert, S., Veldkamp, T. and Ward, P. (2016). The world's road to water scarcity: shortage and stress in the 20th century and pathways towards sustainability. *Scientific reports*, **6** (1), 1–16.

Liu, J., Yang, H., Gosling, S. N., Kummu, M., Flörke, M., Pfister, S., Hanasaki, N., Wada, Y., Zhang, X., Zheng, C. *et al.* (2017). Water scarcity assessments in the past, present, and future. *Earth's future*, **5** (6), 545–559.

Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of business & economic statistics*, **13** (2), 151–161.

Mussen, P. and Eisenberg-Berg, N. (1977). *Roots of caring, sharing, and helping: The development of pro-social behavior in children.* WH Freeman.

Niles, M. T., Lubell, M. and Haden, V. R. (2013). Perceptions and responses to climate policy risks among california farmers. *Global Environmental Change*, **23** (6), 1752–1760.

Orlick, T. (1983). Enhancing love and life mostly through play and games. *Journal of Humanistic Counseling, Education & Development.*

Schultz, P. W., Messina, A., Tronu, G., Limas, E. F., Gupta, R. and Estrada, M. (2016). Personalized normative feedback and the moderating role of personal norms:

A field experiment to reduce residential water consumption. *Environment and Behavior*, **48** (5), 686–710.

Wada, Y. and Bierkens, M. F. (2014). Sustainability of global water use: past reconstruction and future projections. *Environmental Research Letters*, **9** (10), 104003.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf