Ecological Economics 208 (2023) 107801

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Analysis

Game-based education promotes practices supporting sustainable water use

Ennio Bilancini [a,1], Leonardo Boncinelli [b,1], Roberto Di Paolo [a,*,1]

[a] *IMT School for Advanced Studies Lucca, Piazza S.Francesco 19, 55100 Lucca, Italy*

[b] *Department of Economics and Management, University of Florence, Via delle Pandette 9, 50127 Firenze, Italy*

ARTICLE INFO

Keywords:
Water
Prosocial behavior
Field quasi-experiment
Game-based learning
Games for social change

ABSTRACT

We estimate the impact of a game-based educational program aimed at promoting practices for 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 to record 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 practices supporting sustainable water use behaviors in students and their parents.

1. Introduction

Sustainable water use 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 use is, in many cases, an instance of prosocial behavior 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 use, unless social norms exert sufficient social pressure to push self-interested individuals to do so. 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 use (Houser et al., 2012; 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 that encourage good practices in environmental benchmarks (see Ardoin and Bowers, 2020; Otto and Pensini, 2017, for a recent review), such as the use of water (Niles et al., 2013; Cuadrado et al., 2017). Wang and Chermak (2021) show that workshop-based education programs are effective methods for influencing household behaviors and demonstrate the efficiency of those programs on water conservation. Otto et al. (2019) examine how children's environmental attitude and behavior develop from around the age of 7, finding an increase until the age of 10. In a field experiment, Schultz et al. (2016) analyze the role of social norms in promoting water conservation, finding that people who received normative information about similar households 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 weak 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 use and can relate their behavior to concrete and specific situations such as water collection or body washing.

A recent literature has emerged exploring the effectiveness of game-based learning for environmental sustainability, in particular for sustainable water governance (Medema et al., 2019). Aubert et al. (2019)

* Corresponding author.
E-mail addresses: ennio.bilancini@imtlucca.it (E. Bilancini), leonardo.boncinelli@unifi.it (L. Boncinelli), roberto.dipaolo@imtlucca.it (R. Di Paolo).

[1] All authors contributed equally to the work presented in this paper.

https://doi.org/10.1016/j.ecolecon.2023.107801
Received 21 September 2021; Received in revised form 31 December 2022; Accepted 6 March 2023
Available online 21 March 2023

develop a framework to guide and reflect on the design and assessment of game-based approaches, starting from the observation that factors such as expected aims, targeted audience, and spatial and temporal scales, strongly influence the design of game-based approaches and the research developed to assess them. More recently, de Kraker et al. (2021) focus on the effectiveness of a multi-player serious game on river management to support social learning, which is essential to the social sustainability of water management strategies by achieving a convergence in perspectives among societal stakeholders. Robertson (2022) find evidence that gaming activities can help students familiarize with terminology related to transport, fluxes, and storage, which in turn is fundamental to understand the global water cycle. With a broader perspective, Janakiraman et al. (2018) provide a survey of the literature on game-based learning as a tool to facilitate attitude change for environmental sustainability. Fjællingsdal and Klöckner (2020) use qualitative data from focus groups to conclude that board games can be highly effective tools in some aspects of environmental communication. Ho et al. (2022) provide experimental evidence that students understand the relevance of the SDGs by playing a board game designed to simulate the real world, including national and international policies. Our game-based intervention stands out from the other contributions in the literature in that (i) it combines game-based learning (employing a purposely designed board game) with gamification (implemented with a public ranking based on points gained through 4 different activities) and (ii) it involves a large number of students (around one thousand in the treatment group and one thousand in the control group) in a quasi-experimental study.

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 use as well as awareness about the municipal water system and its usage. The targets of the program were around 1000 students from 2nd–4th grades and their families. The program relied primarily on ludic[2] 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 use with three waves of surveys administered, respectively, immediately before the program started, two days after the end of the program activities at the schools, 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; Di Riso et al., 2010; Bevans et al., 2020; Alan and Kabasakal, 2020).

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 use 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] In this case we follow the definition by Caillois (2001), where *ludus* refers to a structured activity with explicit rules (games), while *paidia* refers to unstructured and spontaneous activity (playfulness).

2. Methods

2.1. 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),[3] 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 game[4] 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 A).

The program was divided in three distinct phases. The first phase was titled *How not to drown in a glass of water*. At the beginning of the project, a group of educators, specifically selected for 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 labeled 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

[3] 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.

[4] Urban games are a branch of "pervasive games" with the characteristic of having the city or portions of it as a playground. "Pervasive games" are playful activities that use everyday spaces and objects to create an active narrative, leading the user to see reality in a different form.

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 focusing on playing the board game BLUTUBE. Specifically, 16 classes (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 months later 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.

2.2. Data and empirical strategy

The program described in Section 2.1 qualifies as a natural quasi-experiment (Cook et al., 1979; 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, at the very beginning of phase three (after one day of the stage tournament), 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' behaviors and habits related to water use (the original and the English-translated questionnaires can be found in Appendices A.4 and B, respectively). Specifically, the survey contained seven distinct water-related questions regarding 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?". Each question was chosen to measure the primary purposes of the program. The main targets that Lucca Crea and GEAL had in mind while they were developing the educational program were the students' habits. Thus, we constructed the seven questions to measure all the particular practices (and not the frequency, which can be assumed randomly distributed between families) involved in the urban and board game. Answers were recorded using a 1-to −5 Likert scale which was proposed in three cases with categories going from the least virtuous to the most virtuous and in the remaining four cases in the reverse order. For the analysis presented in Section 3 we recoded all answers such that category 1 is always the least virtuous and category 5 is always the most virtuous (the original scale for each question can be found in Appendix B).

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 measured 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. All parents of students involved provided an informed consent form (Fig. 6 in Appendix A.1), with the specific consent for the possibility to link students' answers to their scores in the program. All the signed consent forms were collected and stored by the promoter of the program, Lucca Crea. 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 immediately after the stage tournaments were over, 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.

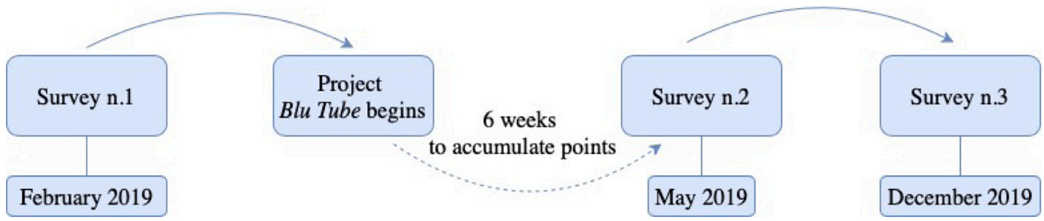

Fig. 1. Timeline of the quasi-experimental study of the BLUTUBE program.

3. 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). Fig. 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.[5] The final sample consists of 5273 questionnaires (see Table 2 in Appendix C).

Summary statistics by treatment and control groups for the pre-program survey show that the two groups are not well balanced[6] (see Table 3 in Appendix C): 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 D.2).

3.1. Aggregated reported behavior

Fig. 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 Fig. 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 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 an impact between the pre-program and the post-program surveys.

Furthermore, Fig. 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 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

[5] The number of students in the three waves is different due to those students being absent during the survey administrations.

[6] We run a robustness analysis of our results with a randomly drawn sub-sample balanced between control and treatment groups. Results are reported in Appendix D.1.

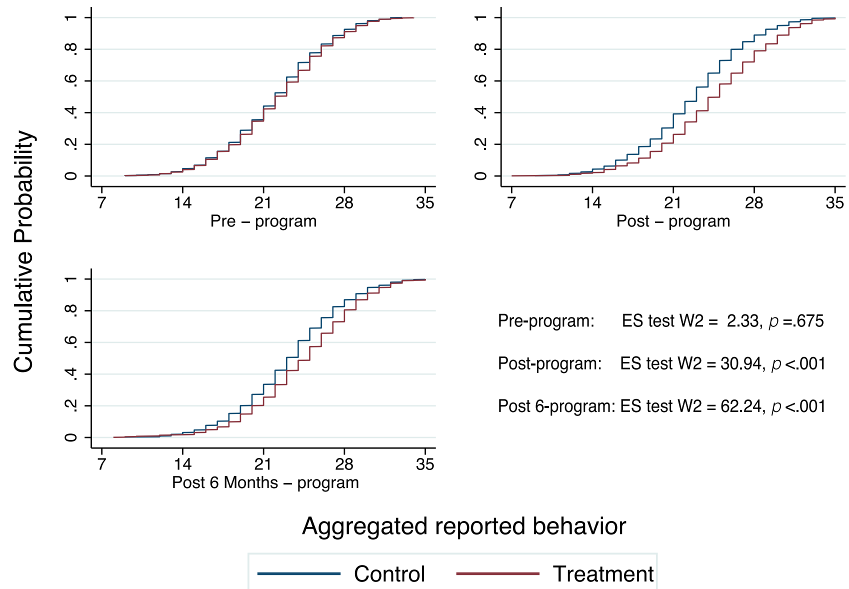

Fig. 2. Cumulative distribution function of the aggregated reported behavior by conditions and waves. Distributions in the post-program and 6 months after are shifted to the right in the treatment group, with a statistically significant differences between conditions. ES stands for Epps–Singleton test.

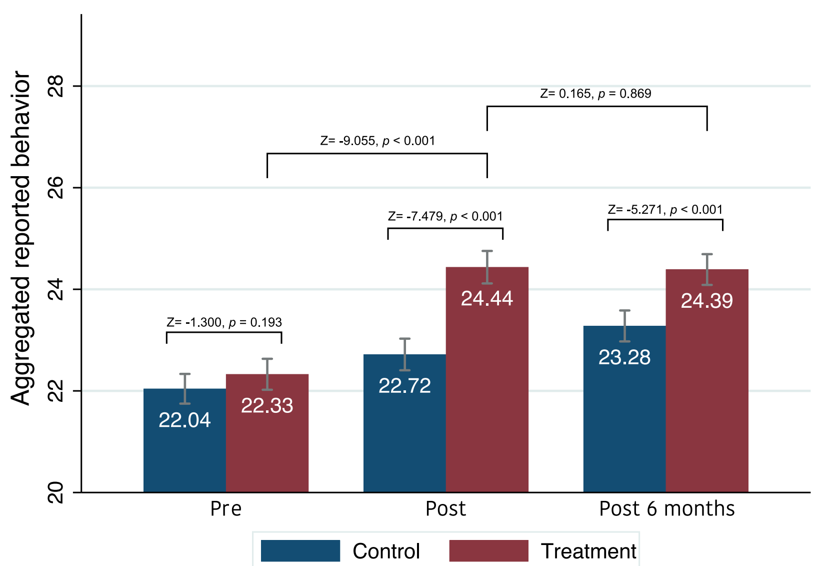

Fig. 3. Average of the aggregated reported behavior by conditions and waves. In the pre-program period, the aggregated reported behavior in the treatment group is not significantly different from the control group (Mann–Whitney test, $Z = -1.300, p = 0.193$). In the post-program period and after 6 months, 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.

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).

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 variable for each survey wave (omitted category is the pre-program survey), a dummy for each 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. Results of the regression analysis are reported in Table 1.

Model 1 reports the results for the main regressors. No statistically significant difference is found between treatment and control groups before the intervention. Post and Post 6 show a positive and statistically significant effect that may be related to the mere passing of time. We notice that the size effect of Post 6 (1.245) is roughly double than the size effect of Post (0.682), consistently with the idea that, as students grow older, their reported behavior improves as an accumulation process of learning good practices of water usage, irrespective of treatment. Importantly, the positive and statistically significant coefficient of the interaction term Treatment X Post reveals that the treatment intervention has an effect that goes beyond learning over time. Moreover, such an effect lasts after 6 months from the intervention, as shown by

Table 1
Linear fixed effect regression.

	Model 1	Model 2
Treatment	0.435	0.767
	(0.374)	(0.525)
Post	0.682***	0.701***
	(0.199)	(0.198)
Post 6	1.245***	1.226***
	(0.223)	(0.221)
Treatment X Post	1.418***	1.415***
	(0.293)	(0.293)
Treatment X Post 6	0.812*	0.801*
	(0.309)	(0.308)
3rd Grade		1.085
		(0.551)
4th Grade		1.604***
		(0.366)
Treatment X 3rd Grade		−0.627
		(0.684)
Treatment X 4th Grade		0.039
		(0.655)
Cognitive Skills		0.275
		(0.154)
Students		0.051
		(0.035)
Constant	21.966***	19.965***
	(0.264)	(0.604)
N	5182	5182

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. *Post* and *Post 6* are equal to 1 if the survey is taken in the post-program or post6-program wave, respectively. *Pre* is the reference category and refers to the survey taken in the pre-program wave. *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. Standard errors (in parenthesis) are clustered at class level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the interaction Treatment X Post 6, even if with a smaller effect size and a reduced significance level. Model 2, where we include controls for the grade of the students, their interaction with the treatment dummy, an index of cognitive skills and the number of students in each class, confirms the previous findings. Among these, as an additional remark, we observe that the coefficient of the 4th Grade is positive and statistically significant, suggesting that older students report a more responsible water usage. Quite interestingly, it appears to be no difference in the effect of the treatment across 2nd, 3rd and 4th grades.

3.2. Disaggregated reported behaviors

Fig. 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 linear regression models because of the 5-tier ordinal structure of answers.

Fig. 5 reports the estimates of the relevant coefficients of the ordered logit regressions (detailed estimates can be found in Table 8 in Appendix D). 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 dimensions, 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, namely teeth brushing and eating fruit and vegetables, seem to have been less affected. Looking at the *Teeth* question, it must be noted that the baseline level in the pre-program phase was already really high ($M = 4.49$ in the control group and $M = 4.60$ in the treatment group). This may have implied a ceiling effect, which may have precluded an increase in that measure. In the case of the *Vegetables* question, it can be argued that such question is, actually, not immediately or clearly related to responsible water usage. Following this line of reasoning, we can expect that the treatment intervention had indeed not much bite on it, which is what we find in the empirical results. A robustness check can be found in Table 5 of the Appendix D.1, where the same regression analysis of Table 1 is performed after excluding the answer to the *Vegetable* question from the aggregate self-reported behavior. 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.

4. Discussion

Our results provide field evidence about the effectiveness of promoting sustainable water-related behaviors 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). Although we cannot say if game-based educational policies are relatively more or less effective than traditional ones — since we do not observe such a counterfactual — we believe that the evidence that we collected strongly pushes towards a greater consideration of game-based educational programs as policy instruments to promote sustainable habits and behaviors, especially when students 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.

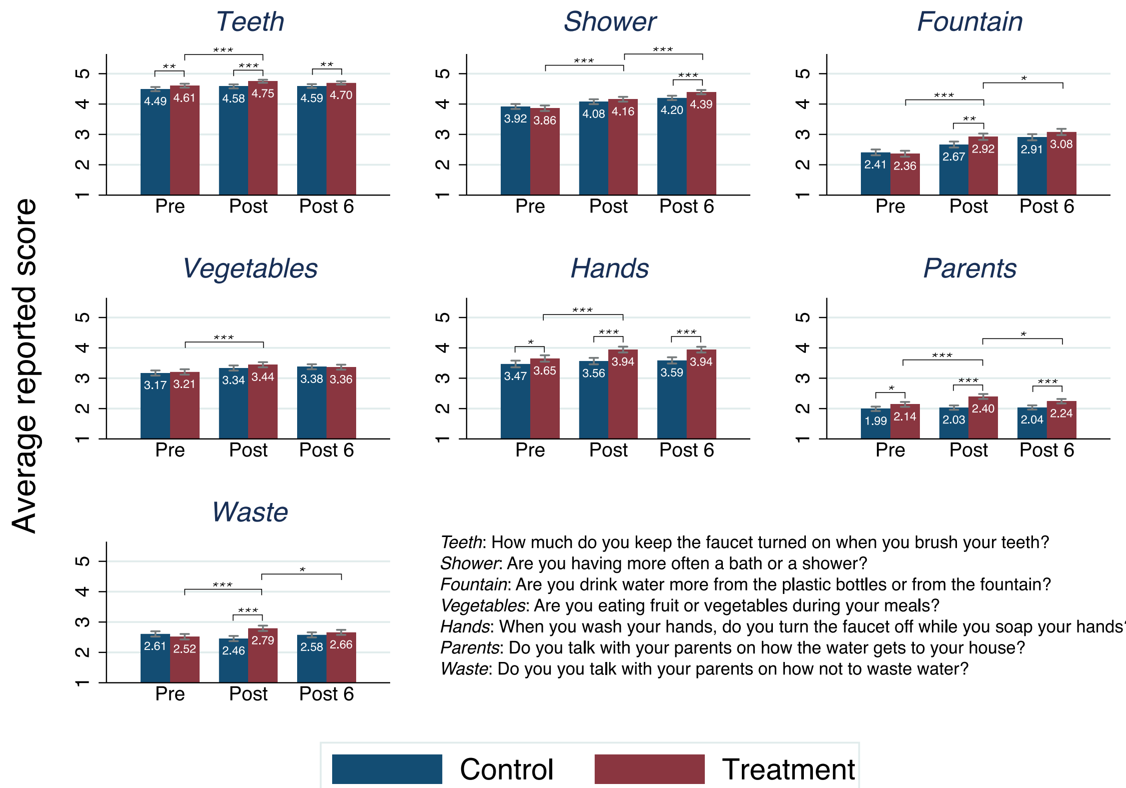

Fig. 4. Average reported behavior by questions, conditions and survey wave. Each answer assume values from 1 to 5. Questions are reported in the figure. Statistically significant difference between conditions are reported above columns (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Error bars represent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

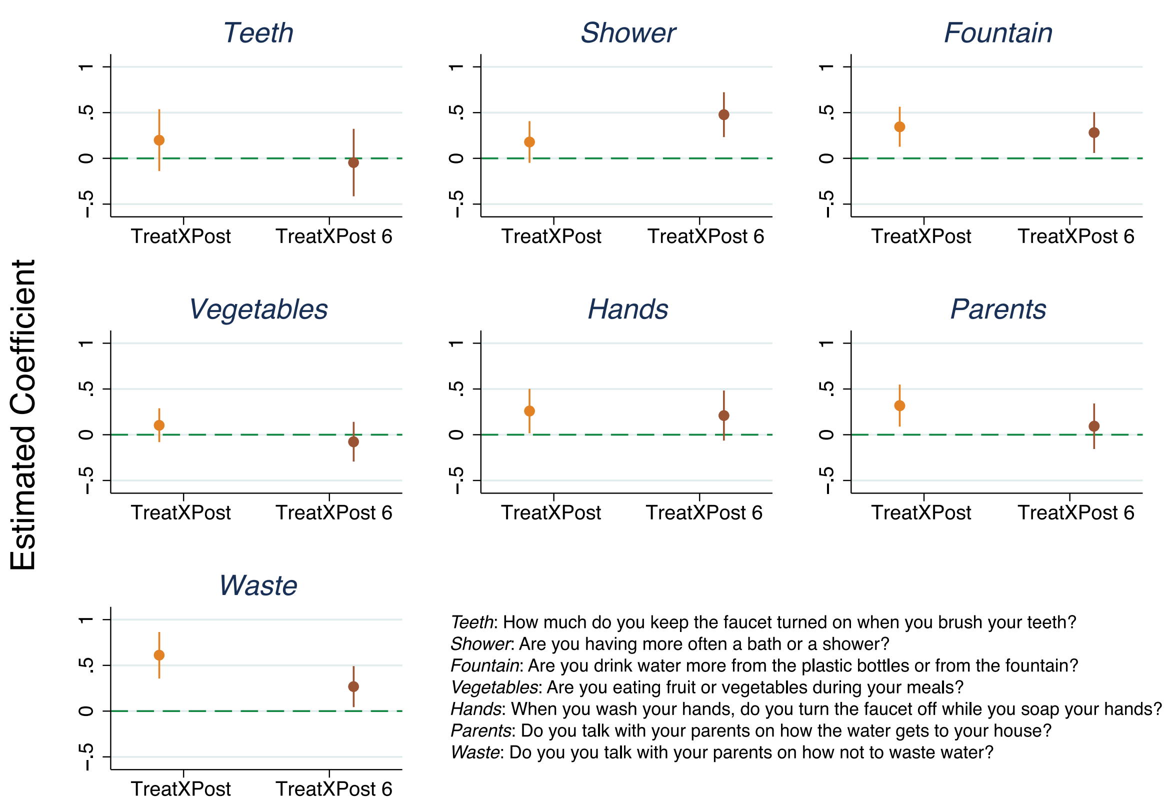

Fig. 5. Estimated coefficient of the Ordered Logit regression in Table 8 in Appendix D.3. The dependent variables are the 7 questions, which assume values from 1 to 5. Questions are reported in the figure. Error bars represent the 95% confidence interval.

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. Thanks to the fact that some schools had classes participating in both treatment and control groups, while other schools had classes participating in either the control group only or the treatment group only, we could check for the presence of faster improvement for the classes in the control group which were in schools where also classes in the treatment groups were present. We did not find evidence of such faster growth and, hence, we are inclined to support the first explanation, namely that the program has had in part the effect of accelerating learning of water related behaviors and issues.

A standard limitation of quasi-experiments 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. While it is not uncommon to rely on reported behavior in studies on pro-environmental preferences and attitudes (Meyer, 2015; Giannetti et al., 2021; Silvi and Padilla, 2021), the limitation here lies in the fact that the observed treatment effect may be generated by students learning what is better for society and then answering accordingly, without really changing their behavior. Unfortunately, in our case 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. To have a more concrete idea of the impact that the BLUTUBE program may have had on water consumption, one can consider the estimated reduction in water use implied by the self-reported measures that we collected. For instance, if we consider shower vs. bath, hand washing, and teeth brushing — and we take a conservative view on water consumption and use by students — we can estimate that the water saved may be between 3000 and 3500 liters per year per student (about 2500 for shower, 200 for hand washing and 500 for teeth brushing), which sums up to more than 3 million liters per year for the Municipality of Lucca.

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.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge Lucca Crea s.r.l., the Ufficio Scolastico Territoriale of the Province of Lucca, and GEAL s.p.a. for their kind support and essential advise. A special thank goes to the non-profit association LudoLega Lucchese who made it possible to implement the BLUTUBE program. We thank Silvia Ceccarelli, Veronica Pizziol, Simone Violi and all the educators for their precious assistance.

Fundings

This work was supported by the IMT School for Advanced Studies Lucca through the PAI project Pro.Co.P.E. "Prosociality, Cognition, and Peer Effects" and by the Italian Ministry of Education, University and Research (MIUR) through the PRIN project Co.S.Mo.Pro.Be. "Cognition, Social Motives and Prosocial Behavior" (grant n. 20178293XT).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecolecon.2023.107801.

References

Alan, Ü., Kabasakal, K.A., 2020. Effect of number of response options on the psychometric properties of likert-type scales used with children. Stud. Educ. Eval. 66, 100895.

Ardoin, N.M., Bowers, A.W., 2020. Early childhood environmental education: A systematic review of the research literature. Educ. Res. Rev. 100353.

Aubert, A.H., Medema, W., Wals, A.E.J., 2019. Towards a framework for designing and assessing game-based approaches for sustainable water governance. Water 11 (4), 869.

Bevans, K.B., Ahuvia, I.L., Hallock, T.M., Mendonca, R., Roth, S., Forrest, C.B., Blackwell, C., Kramer, J., Wakschlag, L., 2020. Investigating child self-report capacity: A systematic review and utility analysis. Qual. Life Res. 29 (5), 1147–1158.

Blake, P.R., Piovesan, M., Montinari, N., Warneken, F., Gino, F., 2015. Prosocial norms in the classroom: The role of self-regulation in following norms of giving. J. Econ. Behav. Organ. 115, 18–29.

Caillois, R., 2001. Man, Play, and Games. University of Illinois Press.

Campbell, D.T., Stanley, J.C., 2015. Experimental and Quasi-Experimental Designs for Research. Ravenio Books.

Cobo-Reyes, R., Dominguez, J.J., García-Quero, F., Grosskopf, B., Lacomba, J.A., Lagos, F., Liu, T.X., Pearce, G., 2020. The development of social preferences. J. Econ. Behav. Organ. 179, 653–666.

Cook, T.D., Campbell, D.T., Day, A., 1979. Quasi-Experimentation: Design & Analysis Issues for Field Settings, vol. 351. Houghton Mifflin Boston.

Copple, C., Sigel, I.E., Saunders, R., 2013. Educating the Young Thinker: Classroom Strategies for Cognitive Growth. Routledge.

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

Danielson, C.K., Phelps, C.R., 2003. The assessment of children's social skills through self-report: A potential screening instrument for classroom use. Meas. Eval. Couns. Dev. 35 (4), 218–229.

Di Riso, D., Salcuni, S., Chessa, D., Raudino, A., Lis, A., Altoè, G., 2010. The strengths and difficulties questionnaire (SDQ). Early evidence of its reliability and validity in a community sample of Italian children. Personal. Individ. Differ. 49 (6), 570–575.

Fjællingsdal, K.S., Klöckner, C.A., 2020. Green across the board: Board games as tools for dialogue and simplified environmental communication. Simul. Gaming 51 (5), 632–652.

Giannetti, C., Guarnieri, P., Luzzati, T., et al., 2021. Pro-Environmental Attitude and Behaviours: An Investigation on the Role of Pro-Sociality. Tech. Rep., University of Pisa.

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 Sustain. 1 (9), 486–494.

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

Ho, S.-J., Hsu, Y.-S., Lai, C.-H., Chen, F.-H., Yang, M.-H., 2022. Applying game-based experiential learning to comprehensive sustainable development-based education. Sustainability 14 (3), 1172.

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. Nat. Hum. Behav. 4 (1), 36–44.

House, B.R., Tomasello, M., 2018. Modeling social norms increasingly influences costly sharing in middle childhood. J. Exp. Child Psychol. 171, 84–98.

Houser, D., Montinari, N., Piovesan, M., 2012. Private and public decisions in social dilemmas: Evidence from children's behavior. PLoS One 7 (8), e41568.

Janakiraman, S., Watson, S.L., Watson, W.R., 2018. Using game-based learning to facilitate attitude change for environmental sustainability. J. Educ. Sustain. Dev. 12 (2), 176–185.

de Kraker, J., Offermans, A., van der Wal, M.M., 2021. Game-based social learning for socially sustainable water management. Sustainability 13 (9), 4646.

Kummu, M., Guillaume, J.H., de Moel, H., Eisner, S., Flörke, M., Porkka, M., Siebert, S., Veldkamp, T., Ward, P., 2016. The world's road to water scarcity: Shortage and stress in the 20th century and pathways towards sustainability. Sci. Rep. 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.

Medema, W., Mayer, I., Adamowski, J., Wals, A.E., Chew, C., 2019. The potential of serious games to solve water problems: Editorial to the special issue on game-based approaches to sustainable water governance. Water 11 (12), 2562.

Meyer, B.D., 1995. Natural and quasi-experiments in economics. J. Bus. Econom. Statist. 13 (2), 151–161.

Meyer, A., 2015. Does education increase pro-environmental behavior? Evidence from Europe. Ecol. Econom. 116, 108–121.

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

Orlick, T., 1983. Enhancing love and life mostly through play and games.. J. Humanist. Couns. Educ. Dev..

Otto, S., Evans, G.W., Moon, M.J., Kaiser, F.G., 2019. The development of children's environmental attitude and behavior. Global Environ. Change 58, 101947.

Otto, S., Pensini, P., 2017. Nature-based environmental education of children: Environmental knowledge and connectedness to nature, together, are related to ecological behaviour. Global Environ. Change 47, 88–94.

Robertson, W.M., 2022. Increasing student engagement and comprehension of the global water cycle through game-based learning in undergraduate courses. J. Geosci. Educ. 70 (2), 161–175.

Schultz, P.W., Messina, A., Tronu, G., Limas, E.F., Gupta, R., Estrada, M., 2016. Personalized normative feedback and the moderating role of personal norms: A field experiment to reduce residential water consumption. Environ. Behav. 48 (5), 686–710.

Silvi, M., Padilla, E., 2021. Pro-environmental behavior: Social norms, intrinsic motivation and external conditions. Environ. Policy Gov..

Wada, Y., Bierkens, M.F.P., 2014. Sustainability of global water use: Past reconstruction and future projections. Environ. Res. Lett. 9 (10), 104003.

Wang, J., Chermak, J.M., 2021. Is less always more? Conservation, efficiency and water education programs. Ecol. Econom. 184, 106994.