J Cult Econ
DOI 10.1007/s10824-016-9285-6

ORIGINAL ARTICLE

Nudging museums attendance: a field experiment with high school teens

Patrizia Lattarulo[1] · **Marco Mariani**[1] · **Laura Razzolini**[2]

Received: 17 February 2016 / Accepted: 28 September 2016

Abstract This paper reports results from a field experiment conducted to study the effect of incentives offered to high school teens to motivate them to visit art museums. A vast literature exists on the design of incentives to modify the behavior of firms and consumers, but not much is known about incentives offered to adolescents and young adults to affect their cultural consumption behavior. Students in the first treatment receive a flier with basic information and opening hours of a main museum in Florence, Italy—Palazzo Vecchio. Students in the second treatment receive the flier and a short presentation conducted by an art expert about the exhibit; students in the third treatment, in addition to the flier and the presentation, receive also a non-financial reward in the form of extra-credit points toward their school grade. The analysis yields two main findings. First, non-financial reward is more effective at inducing the students to undertake the encouraged visit than either the simple presentation or the basic information with the flier. Second, over a longer time horizon the non-financial reward does not induce a significant change in behavior with respect to the simple presentation.

✉ Laura Razzolini
lrazzolini@vcu.edu

Patrizia Lattarulo
patrizia.lattarulo@irpet.it

Marco Mariani
marco.mariani@irpet.it

[1] IRPET – Tuscany's Regional Institute for Economic Planning, Villa La Quiete alle Montalve, Via Pietro Dazzi, 1, 50141 Firenze, Italy

[2] Department of Economics, School of Business, Virginia Commonwealth University, 301 West Main Street, Box 4000, Richmond, VA 23284, USA

Published online: 07 October 2016

Keywords Incentives · Non-monetary rewards · Cultural consumption · Cluster-randomized experiment · Randomization inference

JEL Classification M52 · D82 · C93 · I28 · Z10 · Z18

1 Introduction

Adolescents and teens often visit museums with their families or school teachers and classmates, but when it is up to them to choose how to allocate their free time, museums rarely appear at the top of their preferences (Gray 1998).This happens in spite of the "open doors" policy followed by most museums and their attempts to portray an image of educational and entertaining institutions. The literature on cultural consumption reports a correlation between participation to cultural events and household income and status (Falk and Katz-Gerro 2016). Much less attention, however, has been paid to incentives offered by cultural organizations, such as museums and art galleries, to pursue the goal of increasing attendance and visits, particularly by adolescents and teens. Our paper begins to fill this gap by providing evidence from a field experiment designed to study incentives offered to high school teens to motivate them to visit an art museum in Florence, Italy. The aim of the experiment is to encourage individual museum attendance during the high school years, and to identify best practices to transform this behavior into a long run cultural consumption. Studying museums' attendance and cultural consumption, in general, is interesting from a social point of view for several reasons: cultural consumption may lead to an increase in cultural capital (Throsby 1999; Kisida et al. 2014), may increase the quality of citizenship (Duffy 1992), and may entail positive education spillovers (Bowen et al. 2014).[1]

The field experiment was conducted in Florence, Italy during the Spring and Fall of 2014. The experiment was designed to identify the best incentives to offer to high school teens to motivate them to visit an art museum. Students in the first treatment received a flier containing basic information and opening hours of a main museum in Florence—Palazzo Vecchio. Students in the second treatment received the flier and a short presentation by an art expert from the museum; students in the third treatment group, in addition to the flier and the presentation, received a non-financial reward in the form of extra-credit points toward their school grade.

Field experiments have been increasingly popular in economic studies (see Levitt and List 2009, for a general survey). However, their application to this area of research is novel. A growing literature explores the use and effects of financial and non-financial incentive programs to change individuals' health behavior (Charness and Gneezy 2009; Babcock and Hartman 2010; Royer et al. 2015), to improve students' performance in school (Levitt et al. 2012), or to encourage public voluntary service (Ashraf et al. 2014). Recently published contributions in the field of cultural economics by Suárez-Vázquez (2011), Bakhshi and Throsby (2014), and Berlin et al. (2015) report results from

[1] Bowen et al.'s (2014) field study shows how museum attendance and art exposure may lead to "significantly stronger critical thinking skills."

laboratory or field experiments, but none focuses on museum attendance and means to increase its practice. Close to our work is a recent field experiment by Kisida et al. (2014), which focused on the benefits of museum experience rather than on motivators toward museum attendance. Kisida et al. shows that student's early exposure to museums raises their further engagement with museums and art in general. In this paper, we focus on incentives and nudges that can be provided to young adults to lead them to a greater engagement.

In our field experiment, about 300 high school students from 15 different classrooms[2] were offered one of three different levels of encouragement inviting them to visit the museum for free during a given period of time. Students were also asked to complete a questionnaire about their and their families' background characteristics and attitudes toward museum attendance and cultural consumption in general. The three levels of encouragement defined three treatments for the experiment. Treatment assignment was randomized at the class level, which configures the experiment as a cluster-randomized trial (Duflo et al. 2008; Gerber and Green 2012) and the analysis was performed with randomization inference techniques (see Gerber and Green 2012; Imbens and Rubin 2015), which allow statistical tests over small samples. The students were asked to prove their visit to the museum by returning the admission ticket. Finally, six months after the experiment, the students were contacted again and information was collected regarding any additional number of museum visits since the original intervention.

The aim of this research is to estimate which form of encouragement is associated with a higher probability of visiting the museum, and to establish whether and how each encouragement regime affects the students' subsequent behavior.

Results suggest that presentations in the classroom by museum personnel are very likely to increase future voluntary museum attendance, independent on whether the students did or did not undertake the encouraged visit at the time of the intervention. The additional provision of non-financial reward (extra-credit points) is shown to boost immediate execution of the suggested museum visit, but has limited effects on future behavior.

The paper is organized as follows. Section 2 describes the experimental design and presents the variables of interest. Section 3 introduces the randomization inference approach for the estimation, while Sect. 4 reports the results of the analysis. Section 5 concludes the paper.

2 Experimental design

The field experiment was run in Florence, Italy, at three different points in time during 2014. The experiment involved 297 students from 15 different classes and from three different high schools. Three students chose not to participate, thus

[2] In the Italian school system, high school students are assigned to a class as freshmen and remain with the same group of peers all five years of high school. Students choose the type of secondary school they wish to attend, as different high school offer a different curriculum specializing in classics, scientific, fine arts or technical studies.

leading to 294 participants. All students attended the fourth year of high school and were aged 17–18. All high schools offered a similar program of studies involving a mix of humanities, mathematics and scientific subjects.[3] Students were offered the opportunity to visit Palazzo Vecchio. Palazzo Vecchio, located in the city center, houses the city's main offices and is one of the most visited museums in Florence. The entrance is free to individuals 17 and younger.[4]

The experimental design consisted of three treatments, depending on the different type of encouragement, *W,* received:

- $W =$ *flier—f*: students receive a flier containing basic information, as opening hours and a brief description of the museum and a short text written by the experimenters stating the importance of museum attendance;
- $W =$ *presentation—p:* in addition to the previous—flier and text, students receive a short presentation about the exhibit conducted by an art expert from the museum;
- $W =$ *reward—r*: in addition to the previous—flier, text and presentation, students receive a non-financial reward in the form of extra-credit points toward their final school grade.

The flier is Palazzo Vecchio's official brochure and it was distributed to all students. All students also received a text written by the experimenters, to be voluntarily read at home and shared with parents, containing rather general statements stressing the importance of museum attendance.

The presentation was done by an art expert, specialized in communicating to a young audience. It followed a strict protocol and used the support of videos, audio materials and illustrations. The focus of the presentation was on enhancing the students' curiosity about museum visits in general and Palazzo Vecchio in particular, by portraying the visits as an intriguing and entertaining experience.[5]

The reward treatment consisted of extra-credit points toward the final class grade. In the Italian high school system, final class grade is based on (1) final subjects' grades on a 1–10 points scale and (2) extra-curricular points on a 0–1 point scale. Extra-curricular points can be obtained as reward for voluntary extra-curricular activities, such as language or music courses attended, competitive sport participation, volunteering or cultural activities performed. The extra-curricular activities must be approved by the school principal. Both curricular and extra-curricular

[3] The high school type involved in the experiment is named Liceo Scientifico and is, by far, the most popular choice among Italian students that wish to follow a general education program. This high school provides no vocational education, as its main purpose is to prepare students for college. For recent official data on high school enrolments in Italy see http://www.istruzione.it/allegati/2014/focus_iscrizioni_as_2014_2015.pdf.

[4] For students over 17 years old, a free ticket was provided, to guarantee equal access to all participants.

[5] For example, Palazzo Vecchio was presented as a place of art and history, but also as the location of contemporary fiction, movies and videogames. In addition, the presentation created bridges between 16th century and today's culture in imagery and language and, with the support of images, pointed to the presence of irony in the 16th century artworks. The presentation was conducted by an educational expert from Mus.e (http://musefirenze.it/en), an association dedicated to the enhancement of the cultural heritage of the Florentine Civic Museums and, more in general, of the city of Florence.

credits are relevant for the final graduation degree (GPA) at the end of the five year high school program.

Under the flier treatment, students received the weakest possible incentive to perform the visit. The presentation treatment is aimed at stimulating the students' intrinsic motivation to visit a museum, while the reward treatment provided an additional extrinsic stimulus. Following Ryan and Deci (2000), the basic distinction between intrinsic and extrinsic motivation refers to doing something because it is inherently interesting and/or enjoyable (intrinsic), versus doing something to earn or avoid a specific separable outcome (extrinsic).

All students and teachers were told that they were participating in a study on teen's cultural consumption and were not informed that they were part of an experiment with different forms of encouragement. Students were notified that participation was voluntary and they could withdraw at any time during the study. Instructions are reported in the "Appendix."

Table 1 shows the number of classes and students assigned to each treatment and describes the main characteristics of the students in the sample, by treatment received. The assignment was done by clusters at the class level using a simple randomization procedure, thus guaranteeing the independence between treatment's assignment and potential outcomes, which is required to identify causal effects (Imbens and Rubin, 2015). The 15 classes were randomly assigned to the three treatments in groups of five.[6] Treatment's assignment was designed in a way to avoid or limit any interference between classes assigned to alternative treatments.[7]

Overall, 35 % of the students are male, only a minority is born abroad (6.1 %), 42 % live in the surroundings of the city, outside of the city center.[8] The majority of students are 17 years old, with only a few older (5.8 %).

Average reported students' GPA is 6.8 out of 10 points and around 26 % of the sample considers him/herself interested only in humanities. About half of the students report having the majority of friends in the same class (45 %), and each student lists having about 3.5 close friends in the class. More than half of the friends listed by each student do the encouraged visit (56 %). Each student (on average) had visited more than three museums during the previous year. Most of them had already visited the proposed museum (Palazzo Vecchio) with their families in the past.

[6] The size of the sample (15 observations/classes) was primarily driven by budget and organizational constraints, which made it impossible to boost size. We chose not to perform power analysis to determine appropriate sample size because normally distributed outcomes could not be assumed given the few observations per treatment and expected effect size and variability could not be set in a reasonable way due to lack of previous studies on the topic to provide with reference values.

[7] More complex randomization approaches to improve ex-ante covariate balance and precision with small samples—matched pair designs, blocking/stratifications or re-randomizations—were not possible, because the experimenters had no a priori knowledge of students' characteristics in each class (see Bruhn and McKenzie 2009; Gerber and Green 2012; Imbens and Rubin 2015).

[8] Florence is a relatively small city (about 350,000 inhabitants) with an extensive public transportation network that serves the city and its suburbs. The museum is easy to reach from all schools and different part of the city. The distance between the museum and the students' place of residence is unlikely to negatively affect the execution of the proposed museum visit.

Table 1 Experimental design and descriptive statistics about students

	Treatment			
	Overall	Flier	Present	Reward
Information about the sample				
No. of classes	15	5	5	5
No. of Students (%)	294 (100 %)	93 (32 %)	97 (33 %)	104 (35 %)
Information about the students				
Male (1/0)	0.354	0.215	0.299	0.529
Born abroad (1/0)	0.061	0.022	0.093	0.067
Suburban resident(1/0)	0.422	0.387	0.320	0.548
Older than cohort (1/0)	0.058	0.075	0.062	0.039
Information about the high school				
Most friends are classmates (1/0)	0.486	0.452	0.474	0.529
No. classmates who are friends	3.582	4.441	2.635	3.773
Current GPA (out of 10)	6.817	6.720	6.789	6.931
Interested only in humanities (1/0)	0.262	0.226	0.392	0.173
Leisure and Cultural habits				
Interested in politics (1/0)	0.228	0.215	0.206	0.260
Volunteer (1/0)	0.167	0.204	0.144	0.154
No. museum visits during last year	3.867	3.269	4.742	3.587
Visited Palazzo Vecchio previously (1/0)	0.721	0.677	0.763	0.721
Parents' education				
At least one parent unemployed	0.157	0.151	0.175	0.144
Both parents high school degree	0.327	0.376	0.247	0.356
At least one has college degree	0.455	0.398	0.516	0.452
Parents regularly go to museums	0.197	0.183	0.247	0.163

Table 2 reports the timeline for the field experiment. Students completed a survey at two different points in time: (1) when the experiment was administered, to collect general information about individual characteristics—habits of cultural and leisure consumption for the students and their parents, school performance, friendship ties within and outside of the classroom; and (2) after eight months from the original intervention, to assess the number of additional visits to museums done during the six months following the experiment. Two months after the initial intervention, the admission tickets to the museum were collected.

Table 3 reports by treatment the average number of previous museums visits, the participation rate by class to visit the recommended museum and the average number of additional museum visits completed after 6 months: 40 % of students in the reward treatment did visit Palazzo Vecchio, while only 10 % of those in the presentation treatment and 3 % of those who received the flier did complete the visit. The striking difference in the number of visits to the recommended museum across treatments is more evident if compared to the average number of visits completed during the previous year and during the following 6 months.

Table 2 Timeline

When	W = flier	W = presentation	W = reward
First visit late March/early April 2014	Students are informed that they will be involved in a study about cultural consumption. A flier of Palazzo Vecchio with opening hours is distributed. A brief text written by the experimenters stating the importance of museums' attendance is also distributed		
		A museum operator talks about Palazzo Vecchio for 15 min	
			Students are told that the visit to Palazzo Vecchio will be rewarded with extra-credit points
	Students complete a questionnaire about their background characteristics, cultural consumption habits and within-classroom friendship ties		
	Students who visit Palazzo Vecchio within two months bring the entry ticket back to their teacher		
Second visit After two months	Collect entry tickets from the assigned teachers		
Third visit After eight months (six months after	Students complete a questionnaire about the number of individual visits to museums done in town and out of town in the past six months		

Table 3 Further descriptive statistics across all students

Treatment	Average No. of Museum visit in previous year (SD)	% of students visiting the recommended museum per class	Average No. of Museum visit in the subsequent 6 months (SD)
Flier	3.27 (3.14)	3.23	1.49 (2.56)
Presentation	4.74 (4.03)	10.31	4.39 (4.22)
Reward	3.59 (4.15)	40.38	3.00 (2.50)
Total	3.87 (3.86)	18.71	2.95 (3.38)

3 A randomization inference approach for the analysis of a cluster-randomized experiment

In cluster-randomized experiments, randomization occurs at the cluster level, as the unit of assignment to a treatment is a group (Murray 1998; Arcenaux 2005; Raudenbush et al. 2007). In our study, the unit of assignment to each treatment is the high school class. We assigned 15 different high school classes, each with an average of 19.8 students, to one of three treatments, each corresponding to a different encouragement. All students in the same class received the same type of encouragement. Thus, the class is the natural unit of inference. In this way, we avoid additional complications connected with individual analysis and stick to the goal of evaluating which form of encouragement works best, gross of the peer effects that

could be triggered by the encouragement itself within each class.[9] The choice of keeping the unit of analysis at the cluster level limits the size of our sample. To perform statistical inference with the small number of classes available, we chose to use randomization inference. Unlike other testing procedures, randomization inference enables exact inference for our finite population of classes. It provides us with p values that are exact with respect to this population and valid, irrespective of the sample size, as they are recovered without resorting to distributional assumptions about the test statistics or to other large sample approximations (Gerber and Green 2012; Imbens and Rubin 2015).

In the remainder of this section, we first place our analysis in the formal statistical framework of causal inference based on potential outcomes; then, within this framework, we briefly describe the randomization inference approach (Gerber and Green 2012; Imbens and Rubin 2015). Let $K = 15$ denote the clusters or classes, each containing n_k students. For each class k, let W_k denote the treatment received: $(k = f, p, r)$. The classes were randomly and evenly assigned to the three treatments, so that $M_k = 5$ $(k = f, p, r)$ denotes the five classes assigned, respectively, to the flier, the presentation and the reward treatment. The outcome variables of interest in the analysis are: (1) the share of students undertaking the encouraged visit to the proposed museum and (2) the classroom average number of subsequent visits to other museums undertaken in the 6 months after the experiment (with the encouraged visit excluded from this count). Both variables are defined at the cluster level. For each class, while three potential outcomes can be defined for each variable of interest, only one outcome is actually observed associated with the assigned treatment.

Indeed, the problem is that we can only observe one of the potential outcomes for each unit. For example, if the class receives $W_k = r$, we can only observe the potential outcome $Y_k(r)$, and not $Y_k(f)$ and $Y_k(p)$, i.e., there is a problem of missing information about what the outcome of interest would have been had the class received an alternative treatment. Our goal is to compare the observable $Y_k(r)$ to the unobserved potential outcomes to establish whether, for class k and with respect to a given response variable, the reward encouragement works better than an alternative form of encouragement. Thus, the main challenge consists in finding the most credible approximation of class k counterfactuals $Y_k(f)$ or $Y_k(p)$. To address this missing data problem, we usually shift the focus on estimable quantities, such as expectations under different treatments, and compare these expectations. If we assume that (1) potential outcomes respond only to the treatment and not to other features of the experiment (excludability assumption), and that (2) the potential outcomes for any class do not vary with the treatment assigned to any other class, as it would happen in the presence of between-cluster interference and externalities (Rubin 1980), then the effect of receiving treatment r rather than f can be approximated by the difference in the expected value of the outcome variable in the two treatment groups. The credibility of the approximation depends on the process

[9] It could be interesting to carry out the analysis at the individual level rather than the cluster level, by evaluating if and how the different forms of encouragement affect individual decisions. Such analysis would add the issue of intra-cluster correlation, which in our case corresponds to some sort of peer pressure that can affect both the individual decision to undertake the encouraged visit and the later visits.

that determines how units are assigned to treatments. Since classes are assigned at random by the experimenters, exogeneity is ensured, as assignments are independent of the unit's characteristics. The process of random assignment addresses the missing data problem by creating groups of observations that are a priori identical. Based on this, we define the average treatment effect (ATE) of a particular form of encouragement over another as the average of all unit-level effects. For instance, the ATE for encouragement r with respect to encouragement $e = f, p,$ using a simple difference-in-means estimator (DIM), is:

$$\begin{aligned} \text{ATE}_{r,i,\text{DIM}} &= E[[Y_k(r)|W = r] - E[Y_k(e)|W = e]] \\ &= E[Y_k(r)|W = r] - E[Y_k(e)|W = e]. \end{aligned} \tag{1}$$

Randomization guarantees, with large samples, that pretreatment variables are well balanced across the subsamples defined by treatment assignments and there is no need to adjust for background characteristics. However, with small samples, the process may lead to pretreatment variables which are not perfectly balanced. Since treatment assignment is controlled by the experimenter and, therefore, exogenous, the background variables are not required for unbiased treatment effect estimation. Still, somewhat imprecise estimates may happen if, although uncorrelated to the form of encouragement received, background variables affect the outcomes of interest. To address the unbalances in background characteristics that persist across treatment groups after randomization, we chose to rescale the outcome by subtracting from the observed dependent variable its lagged value prior to (and thus independent of) the experiment, as the lagged value is very likely to be a good predictor of the outcome itself. We, therefore, also used a difference-in-differences estimator (DID) of the ATE. When comparing, for instance, r to $e = f, p$:

$$\text{ATE}_{r,i,\text{DID}} = E\big[\big[Y_k(r) - Y_{k,\text{prior}}\big]|W = r\big] - \big[Y_k(e) - Y_{k,\text{prior}}|W = e\big]. \tag{2}$$

The DID estimator discards class-level fixed effect and focuses on change scores alone, which ensures precision gains.

To calculate p values, we then use the randomization inference approach (Fisher 1935), which requires a sharp null hypothesis, a test statistic, and a measure of extremeness.

The usual null hypothesis is that of no effect whatsoever of the treatment.[10] It is a sharp null hypothesis that allows to infer values for all unobserved potential outcomes: If the treatments have no effect for any unit, the unobserved potential outcomes are identical to the observed one whatever the treatment. With respect to our experiment, the sharp null hypothesis states that:

H0 The treatment effect is zero for all classes, for all treatments, i.e., $Y_k(f) = Y_k(p) = Y_k(r)$ for all k.

The alternative, one-tailed hypotheses are, respectively:

[10] This hypothesis differs from the weaker null hypothesis that the average treatment effects are zero, where the class-level treatment effects could all differ from zero but could be zero on average.

H1$_{p,f}$ There exists at least one class k for which the "presentation" works better than the "flier" treatment, i.e., $Y_k(p) > Y_k(f)$ or, equivalently, $Y_k(p) - Y_k(f) > 0$.

H1$_{r,f}$ There exists at least one class k for which the "reward" works better than the "flier" treatment, i.e., $Y_k(r) > Y_k(f)$ or $Y_k(r) - Y_k(f) > 0$.

H1$_{r,p}$ There exists at least one class k for which the "reward" works better than the "presentation" treatment, i.e., $Y_k(r) > Y_k(p)$ or $Y_k(r) - Y_k(p) > 0$.

As for the choice of a test statistic, the average treatment effects presented earlier in this section represent the most natural candidates. However, as stressed by the literature, also other test statistics could be chosen that are unaffected by possible outliers, such as the difference in medians and the difference in average ranks. We will use these alternative test statistics to perform sensitivity analysis.

Since we know, under the sharp null hypothesis, the complete schedule of potential outcomes, we can simulate all possible randomizations that could have taken place, each time calculating the test statistic, so as to obtain the exact sampling distribution of the latter under the sharp null hypothesis.

In our experiment, with five out of 15 classes assigned to each treatment, there were $[K!/(M_f!\ \ M_p!\ \ M_r!)] = [15!/(5!\ \ 5!\ \ 5!)] = 756{,}756$ alternative treatment assignments possible, but only one of these actually occurred. By looking at the distribution of these hypothetical ATEs, the randomization distribution, and contrasting it to the ATE estimated for the treatment assignment that actually occurred, we can calculate the probability of obtaining—under the sharp null hypothesis—a fictional ATE that is at least as large as the one obtained from the actual experiment. This probability is equivalent to an exact p value and represents a measure of extremeness. A small value (close to zero) of the p value suggests that the observed value of the test statistic is very unlikely in the randomization distribution of the test statistic simulated under the null hypothesis of no effect. This constitutes evidence against sharp null and in favor of the alternative hypothesis.

4 Results

Descriptive statistics from the field experiment at the class level, by treatment, are reported in Table 4: namely, the variables of interest (a and b) and an important background variable, the class average number of museum visits undertaken during the 12 months preceding the experiment (c).

Table 4 shows that the reward encouragement induces the largest share of students (40.2 %) to visit the proposed museum, while participation rates are lower under the presentation (11.8 %) and the flier (3.3 %) treatments. However, if we consider the number of visits to museums done in the 6 months following the experiment, then the average number of visits is higher for classes in the presentation treatment (4.303) than in the flier (1.756) or the reward (2.98) treatments.

Table 4 also shows that, prior to the experiment, classes randomly assigned to the presentation treatment reported a higher number of museum visits in the previous

Table 4 Descriptive statistics at the class level by treatment

Variable	Obs	Mean	SD	Min	Median	Max
Flier						
(a) share of students undertaking the encouraged visit	5	0.033	0.075	0.000	0	0.167
(b) No. of voluntary museum visits six months later	5	1.756	2.013	0.000	1.158	4.875
(c) No. of voluntary museum visits in previous 12 months	5	3.286	0.677	2.760	3.158	4.444
Difference (b, c)	5	−1.530	1.541	−3.222	−2.000	0.431
Presentation						
(a) share of students undertaking the encouraged visit	5	0.118	0.263	0.000	0	0.588
(b) No. of voluntary museum visits six months later	5	4.303	2.346	2.500	3.263	8.368
(c) No. of voluntary museum visits in previous 12 months	5	4.749	1.237	3.760	4.533	6.857
Difference (b, c)	5	−0.446	1.303	−2.147	−0.522	1.511
Reward						
(a) share of students undertaking the encouraged visit	5	0.402	0.180	0.222	0.455	0.650
(b) No. of voluntary museum visits six months later	5	2.980	0.622	2.190	2.909	3.909
(c) No. of voluntary museum visits in previous 12 months	5	3.543	1.058	2.667	3.050	5.000
Difference (b, c)	5	−0.564	0.679	−1.409	−0.491	0.111

12 months than the classes assigned to the other two treatments. This difference, due to chance, constitutes an example of unbalances in background variables, despite randomization. Since the number of museum visits prior to the experiment is a reasonably good predictor of the number of museum visits during the experiment, it makes sense to consider the differences (b, c) and employ a DID in addition to the DIM estimator.[11]

Table 5 shows the results of the randomization inference analysis. Column (1) reports the ATEs observed values for each treatment and columns (2) show the 95 % confidence intervals calculated through the test-inversion procedure suggested by Rosenbaum (2002). The reward encouragement appears again as the most effective way to promote the one-time visit to the proposed museum, but not the best way to boost future museums' visits. The poor performance of the reward, relative to the presentation, realizes not only with the DIM, but also with the more precise DID estimator.

Column (3) shows the one-sided p values, while the randomization distributions of the test statistics of interest, in the form of Kernel probability density functions, are shown in Fig. 1. For example, the observed differential effect (ATE) of receiving the reward rather than the presentation on the share of students

[11] These differences have negative signs because the reference period prior to the experiment is 12 months, while the reference period during the experiment is six months. However, the related ATEs can be positive, since they are defined as the difference in *Ys* between each pair of treatments.

Table 5 Observed average treatment effects, confidence intervals and p values for the sharp null hypothesis *H0:* $Y_k(f) = Y_k(p) = Y_k(r)$ for all classes

Treatment	(1) ATE	(2) ATE 95 % C.I.		(3) p value(right tail)
Effect on the share of students undertaking the encouraged visit (DIM)				
Presentation versus flier	0.084	−0.023	0.208	0.293
Reward versus flier	0.368	0.261	0.497	0.005
Reward versus presentation	0.284	0.176	0.412	0.030
Effect on subsequent voluntary museum visits (DIM)				
Presentation versus flier	2.547	1.494	3.810	0.016
Reward versus flier	1.224	0.173	2.494	0.187
Reward versus presentation	−1.323	−2.379	−0.053	0.836
Effect on subsequent voluntary museum visits (DID)				
Presentation versus flier	1.084	0.031	2.347	0.015
Reward versus flier	0.967	−0.084	2.236	0.189
Reward versus presentation	−0.117	−1.173	1.152	0.840

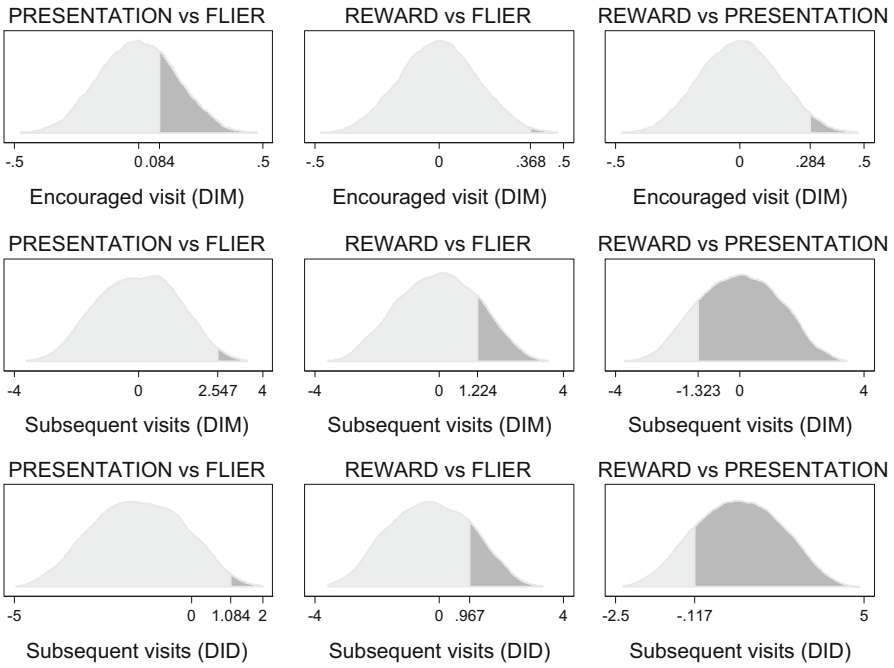

Fig. 1 Randomization distributions of average treatment effects for the sharp null hypothesis *H0:* $Y_k(f) = Y_k(p) = Y_k(r)$ for all classes

undertaking the encouraged visit is 0.284, corresponding to 28.4 % points higher under reward than under presentation. The probability to find, in the randomization distribution simulated under the sharp null hypothesis of no effect, a fictional value

of this ATE that is at least as high as 0.284 is 0.030 (3 %) (see column (3) in Table 5). In Fig. 1, this probability corresponds to the dark gray area in the relative randomization distribution: the smaller the p value (or, equivalently, the dark gray area), the stronger the evidence against the null hypothesis and in favor of the alternative one.[12] Over hypothetical replications of this experiment, there is a 95 % probability that the interval 0.176–0.412 includes the true ATE (see column (2) in Table 5).

Applying this argument to all treatments' pairs we conclude that, to the end of promoting the one-time visit to the museum, the reward not only works better than the simple flier encouragement, but also considerably better than the presentation. On the other hand, there is no substantial difference between the effects of the presentation and the basic flier, as the probability of finding an ATE of 0.084 or higher in the randomization distribution is 29.3 %.

With regard to future visits, the observed ATE of receiving a presentation instead of a flier, using a DIM estimator, is 2.5, which means that the mean of future visits under the presentation is 2.5 points higher than under the flier encouragement. Since the related p value is 1.6 %, we can infer that the presentation encouragement works. This differential effect narrows with the more precise DID estimator, as we control for pre-experiment museum attendance: the difference in the number of visits after the presentation versus the flier is just 1.084, still highly significant (p value $= 1.5$ %). When contrasting the reward to the flier, the effect is still positive (the actual ATE $= 0.9$ with the DID estimator), but loses statistical significance as the probability of finding equal or greater values in the randomization distribution increases to 18.9 %. Finally, we find no evidence of a differential effect of the reward versus the presentation to the end of boosting subsequent museums visits.

In the analysis conducted so far, the outcome at the class level is a mean and the test statistic is a difference between the average outcome of classes assigned to different treatments. As both the class outcome and the test statistic might be affected by the presence of outliers, we examine the sensitivity of randomization inference to the choice of other class outcomes and other test statistics that are not sensitive to possible outliers. As for the class outcome, we use medians. With regard to the test statistic, we consider the difference in average ranks and the difference in medians, which constitute appropriate alternatives to the difference in means (Imbens and Rubin 2015).[13] Table 6 reports the results of the sensitivity analysis. It shows that the reward treatment now significantly outperforms the flier on subsequent visits, but it continues to be as good motivator as the simple presentation.

In summary, the effect of the reward encouragement is large and significantly positive in affecting a one-time likelihood of a visit to the proposed museum. The same is not true with respect to future visits, where the simple presentation may suffice. For this, we conclude that a policy based on an extra-credit reward is not the

[12] Note that a p value of 0.030 satisfies conventional requirements of statistical significance.

[13] While magnitudes of differences in medians are comparable to the differences in means as they are expressed in the same measurement unit, the same cannot be said of ranks.

Table 6 Sensitivity of randomization inference to the choice of alternative class outcomes and test statistics

Treatment	Class outcome: Mean		Class outcome: Mean		Class outcome: Median	
	Test statistic: Diff. in average ranks	*p* value (right tail)	Test statistic: Diff. in medians	*p* value (right tail)	Test statistic: Diff. in medians	*p* value (right tail)
Effect on the share of students undertaking the encouraged visit (simple difference)						
Presentation versus flier	1	0.365	0	0.623	0	0.619
Reward versus flier	6.8	0.002	0.455	0.066	0.455	0.063
Reward versus presentation	5.8	0.010	0.455	0.065	0.455	0.063
Effect on subsequent voluntary museum visits (simple difference)						
Presentation versus flier	5.6	0.022	2.105	0.008	3	0.015
Reward versus flier	2.8	0.178	1.751	0.024	2	0.025
Reward versus presentation	−2.8	0.845	−0.354	0.677	−1	0.925
Effect on subsequent voluntary museum visits (DID)						
Presentation versus flier	0.4	0.027	0.730	0.007	1	0.015
Reward versus flier	2.6	0.173	1.859	0.021	2	0.023
Reward versus presentation	2.2	0.843	1.129	0.683	1	0.928

most effective instrument to increase teens' attendance to museums over time and affect their long run behavior. Instead, classroom visits and direct communication to students by museum staff appear to be a more successful way to increase and affect cultural consumption in teens and adolescents.

5 Concluding remarks

This paper reports results from a field experiment conducted to study incentives offered to high school teens to motivate them to visit an art museum. This is a very novel area of application of field experiments and contributes evidence to the literature that studies the effects of the use of incentives to nudge behavior and, in particular, to increase cultural consumption.

By means of a cluster-randomized trial, with randomization at the class level, we compared the immediate and the subsequent effects on visits to a museum induced by three different types of encouragement: students in the first treatment with low encouragement received a flier containing basic information and opening hours of a main museum in Florence; students in the second treatment with intermediate encouragement received the flier and a short presentation conducted by an art expert to stimulate the students' intrinsic motivations; students in the third treatment with high encouragement, in addition to the flier and the presentation, received also an extrinsic, non-financial reward in the form of extra-credit points toward their school grade. The analysis yielded two main findings. First, the reward is more effective at inducing the students to visit the museum than either the presentation or the basic information with the flier. Second, over a longer time horizon, the extra-credit reward does not induce a significant change in behavior with respect to the simple presentation, which appears to be successful in increasing museum visits. This kind of result stresses the importance of intrinsic determinants of behavior and suggests not overemphasizing the benefits associated to extrinsic rewards, in line with other results found in the literature (see Deci et al. 1999; Gneezy and Rustichini 2000; Fryer 2011).[14]

Access and consumption of cultural goods are important for young adults and teenagers. They may exert effects that extend over a lifetime and benefits that are relevant both from a personal and a social point of view. Unfortunately, young adults and teens often show little interest and awareness of the rich cultural and artistic endowment available to them, even in cities like Florence. This study confirms that policies based on nudging individual behavior with appropriate incentives may be promising in terms of changing individual attitudes toward cultural consumption. Since young adults appear to be positively affected by information provided by experts, partnerships between museums and their educational staff and schools or school districts seems to be a promising avenue to affect teen's behavior. In this vein, museums should invest in educational and communication strategies targeted at outreaching young adults in school and

[14] These findings have received some attention in economic theory (e.g., Kreps 1997; Frey and Jegen 2001; Benabou and Tirole 2003; James 2005).

stimulating youth interest in arts and culture with curiosity-enhancing and intriguing presentations.

Acknowledgments The authors thank schools, teachers and students who participated in the study, as well as the educational experts of the Mus.e Association for the preparation and execution of classroom presentations. A previous version of this work has been presented at the 7th European Workshop on Applied Cultural Economics, held at the Austrian Institute of Economic Research, Vienna (Austria), 3–5 September 2015. The authors wish to thank participants in this event and, in particular, Karol Jan Borowiecki (University of Southern Denmark) for providing useful comments. The authors are also grateful to Alessandra Mattei (University of Florence) and Chiara Bocci (IRPET) for valuable statistical suggestions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix: Instructions[15] (reward treatment)

Welcome!

Your opinions matter to us. In spite of the large historic and artistic endowment and the vast cultural opportunities, in Tuscany the consumption of cultural services by teen and young adults is still poor. This research aims to understand and encourage a particular form of cultural consumption by young adults: visits to museums.

These instructions are simple. Take advantage of the opportunity to enrich your cultural background by visiting in the next few weeks the museum that will be proposed to you. Listen to the presentation that will be delivered in class. (This sentence was removed in the flier treatment). *As reward for visiting the museum, you will receive extra points which will be added to you final class average and will count towards your graduating academic curriculum*. (This sentence was removed in the flier and presentation treatments).

You will also complete a questionnaire asking basic information about you and your family characteristics and preferences. You participation is voluntary and the questionnaire will be completely anonymous. Your name and personal identity will not be associated to any of the answers in the questionnaire. No information will be shared with other parties not associated with this project. In the course of this experiment, you will be identified by a number and no one, including the researchers, will be able to identify your decisions once the experiment is completed.

From today till the end of May you are invited to visit the Museum of Palazzo Vecchio. Entrance to the museum is free if you are younger than 18.

If you do visit the museum, please keep the admission ticket and turn it into your teacher at school to prove your visit, marking the ticket with your ID number.

[15] Instructions have been translated from Italian. For an original version, please contact the corresponding author.

MUSEO DI PALAZZO VECCHIO—TIME for VISITS
Piazza della Signoria, Firenze

October 1–March 31
Every day excluded Thursdays: 9–19
Thursday: 9–14
April 1–September 30
Every day excluded Thursdays: 9–24
Thursday: 9–14

Visit to the Tower (Visit to the tower is suspended in case of rain)

October 1–March 31
Every day excluded Thursdays: 10–17
Thursday: 10–14
April 1–September 30
Every day excluded Thursdays: 9–21
Thursday: 9–14

In preparation for your future visit to the museum, please consider the following.

Did you know?

- By paying taxes, we all share the cost of maintaining and enriching our art and cultural endowment, which is source of wealth for our region and for which we are responsible toward future generations.
- Even though in school and academic programs, humanities subjects are still very important, recent studies show that young adults' and teens' consumption of museums and other cultural exhibits is limited. This is true even in places, like Florence and Tuscany, with a rich historic and cultural heritage, well-known and well appreciated all over the world.
- The city of Florence houses world-renowned museums, such as the Uffizi or Accademia. In Florence, you can find many more museums and exhibits, all extremely interesting, because they tell the history of our city.
- It is scientifically proven that individuals with a rich cultural and artistic background and heritage are more successful in life.

Art enriches your critical thinking

Curiosity moves individuals to look beyond what is observable and known. A museum may inspire in you a new interest in an object, in history or even in a new idea. A museum may help you discover your own preferences, develop your thinking, your attitudes and values. We are offering not just a learning activity, but the possibility of a life enriching experience.

Art offers incentives to individuals for creativity and uniqueness in life

Individual creativity and original thinking set us apart from one another. Creativity and originality are main characteristics of our personality and determine our identity. Many identify themselves as creative, original and innovative. However, they are not aware of how art consumption can enhance the understanding and appreciation of such qualities.

Art lets you know the past to change the future

> I believe that the more you know about the past, the better you are prepared for future.
> Theodore Roosevelt, United States President, 1901–1909.

Art displayed in museums is a great source of information about the past. In Cicero's (106–43 BC) words "History is life's teacher." Knowledge of history, with its lights and shadows, allows us to navigate through the present to build a better future. Museums help us to go beyond barriers in time and space. Museums build a bridge between us and those before us who have created the world in which we live now. We cannot change the world for the better if we do not know it well.

Why contribute to the financing and maintenance of our artistic endowment without enjoying its benefits?

Perhaps, while on vacation abroad you have visited beautiful museums in other countries. You did well! Visits to museums allow you to understand the country you are visiting. Museums are financed by the taxpayers, citizens of those countries. Similarly, in our city, the taxes paid by you and your parents go in part to fund museums. Why don you take advantage of such service?

References

Arcenaux, K. (2005). Using cluster randomized field experiments to study voting behavior. *The Annals of the American Academy of Political and Social Science, 601*, 169–179.

Ashraf, N., Bandiera, O., & Jack, B. K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics, 120*, 1–17.

Babcock, P., & Hartman, J. (2010). Networks and workouts: Treatment size and status specific peer effects in a randomized field experiment. NBER Working Papers 16581.

Bakhshi, H., & Throsby, D. (2014). Digital complements or substitutes? A quasi-field experiment from the Royal National Theatre. *Journal of Cultural Economics, 38*(1), 1–8.

Benabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies, 70*(3), 489–520.

Berlin, N., Bernard, A., & Fürst, G. (2015). Time spent on new songs: Word-of-mouth and price effects on teenager consumption. *Journal of Cultural Economics, 39*(2), 205–218.

Bowen, D. H., Greene, J. P., & Kisida, B. (2014). Learning to think critically. A visual art experiment. *Educational Researcher, 43*(1), 37–44.

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics, 1*(4), 200–232.

Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica, 77*(3), 909–931.
Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.
Duffy, C. T. (1992). The rationale for public funding of a national museum. In R. Towse & A. Khakee (Eds.), *Cultural economics* (pp. 37–48). Berlin Heidelberg: Springer.
Duflo, E., Glennerster, R., & Kremer, M. (2008). Using randomization in development economics research: A toolkit. In P. Schultz & J. A. Strauss (Eds.), *Handbook of development economics* (Vol. 4, pp. 3895–3962). Amsterdam: Elsevier.
Falk, M., & Katz-Gerro, T. (2016). Cultural participation in Europe: Can we identify common determinants? *Journal of Cultural Economics, 40*(2), 127–162.
Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.
Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys, 15*(5), 589–611.
Fryer, R. (2011). Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics, 126*(4), 1755–1798.
Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis and interpretation*. New York: W.W. Norton.
Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics, 115*(3), 791–810.
Gray, C. M. (1998). Hope for the future? Early exposure to the arts and adult visits to art museums. *Journal of Cultural Economics, 22*(2–3), 87–98.
Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. New York: Cambridge University Press.
James, H. S. (2005). Why did you do that? An economic examination of the effect of extrinsic compensation on intrinsic motivation and performance. *Journal of Economic Psychology, 26*(4), 549–566.
Kisida, B., Greene, J. P., & Bowen, D. H. (2014). Creating cultural consumers. The dynamics of cultural capital acquisition. *Sociology of Education, 87*(4), 281–295.
Kreps, D. M. (1997). Intrinsic motivation and extrinsic incentives. *The American Economic Review, 87*(2), 359–364.
Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review, 53*(1), 1–18.
Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2012). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. NBER Working Papers 18165.
Murray, D. M. (1998). *Design and analysis of community trials*. Oxford: Oxford Univ. Press.
Raudenbush, S. W., Martinez, A., & Spybrook, J. (2007). Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis, 29*, 5–29.
Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer.
Royer, H., Stehr, M. F., & Sydnor, J. R. (2015). Incentives, commitments and habit formation in exercise: Evidence from a field experiment with workers at a Fortune-500 company. *American Economic Journal: Applied Economics, 7*(3), 51–84.
Rubin, D. B. (1980). Discussion of Basu's "Randomization analysis of experimental data: The Fisher Randomization Test". *Journal of the American Statistical Association, 75*, 591–593.
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.
Suárez-Vázquez, A. (2011). Critic power or star power? The influence of hallmarks of quality of motion pictures: an experimental approach. *Journal of Cultural Economics, 35*(2), 119–135.
Throsby, D. (1999). Cultural capital. *Journal of Cultural Economics, 23*(1), 3–12.