European Economic Review 159 (2023) 104562

Contents lists available at ScienceDirect

European Economic Review

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

Improving children's food choices: Experimental evidence from the field

Gary Charness[a,*], Ramón Cobo-Reyes[b], Erik Eyster[a], Gabriel Katz[c], Ángela Sánchez[d], Matthias Sutter[e]

[a] *University of California, Santa Barbara, United States*
[b] *American University of Sharjah, United Arab Emirates, and Georgetown University, Quatar*
[c] *University of Exeter, UK*
[d] *Universidad Pontificia Comillas, Spain*
[e] *Max Planck Institute For Research on Collective Goods Bonn, Germany, University of Cologne, Germany, University of Innsbruck, Austria, and IZA Bonn, Germany*

A R T I C L E I N F O

JEL-codes:
C93
I12

Keywords:
Improving food choices
Health
Children
Parents
Information interventions
Field experiment

A B S T R A C T

We present a field experiment to study the effects of different information conditions on food choices of 282 children in elementary schools. Previous interventions have typically paid participants for healthy eating, but this often may not be feasible. We introduce a system where food items are graded based on their nutritional value, involving parents or classmates as change agents by providing them with information regarding food choices. We find parents' involvement in the decision process to be particularly beneficial in boosting healthy food choices, with very strong results that persist months after the intervention.

1. Introduction

Poor diet has been identified by the World Health Organization (2009) as a major determinant of global risks to health and one important reason for the rising costs of healthcare. According to the most recent Global Burden of Disease study (2016), poor diet is linked to one in five deaths worldwide, with low intake of healthy foods being the leading risk factor for mortality. Forty-five percent of all cardio-metabolic deaths in the U.S. are associated with suboptimal intake of dietary elements (Micha et al., 2017).

A poor diet has been shown not only to have consequences for the adult population but also to have long-term implications for children, affecting both their physical well-being and cognitive development. Poor diet weakens the immune system (Sorahindo and Feinstein, 2006) and contributes to the development of dental caries and diabetes (Noble and Kanoski, 2016). It also hinders growth and even undermines intellectual performance (Weinreb et al., 2002; Whitaker et al., 2006).

Recent data on children's eating patterns are far from encouraging. Despite progress in fruit intake, children still fail to meet recommendations for the amount of both fruit and vegetables they should eat daily (Kim et al., 2014). In Spain, the country of our study, only 10% of children consume vegetables on a daily basis, and 29% eat fruit (Williams et al., 2020). The childhood obesity rate in Spain is 17.3%, with 23.3% of children overweight (Aladino Study, Spanish Government, 2019). The data patterns look similar in

* Corresponding author.
E-mail address: charness@econ.ucsb.edu (G. Charness).

https://doi.org/10.1016/j.euroecorev.2023.104562
Received 13 September 2022; Received in revised form 26 July 2023; Accepted 7 August 2023
Available online 14 August 2023

the United States, where the National Health and Nutrition Examination Survey reported for 2011–2012 that over 30% of U.S. children between 2 and 19 years old were overweight, a percentage that has tripled since the period 1971–1974. Because dietary habits are typically set during childhood, it is crucial that improvements are made during this period (Haire-Joshu and Tabak, 2016).

Designing, implementing, and evaluating interventions that produce durable changes in children's eating behavior is critical not only to better understand the drivers and the evolution of eating habits, but also to support policy efforts aimed at tackling the long-term individual and social consequences of poor nutrition. While it has been shown that paying people leads to better nutrition (Belot et al., 2016; Loewenstein et al., 2016), doing so is not always practical or feasible. In particular, monetary incentives cannot easily and permanently be rolled out on a large scale. In this paper, we take a different approach to promoting healthy eating that eschews financial incentives in favor of providing non-financial incentives or nutritional information. Our approach can be implemented with relatively minimal funding and so could be widely adopted.

We conduct a field experiment in Spanish elementary schools to study whether and how it might be possible to influence or incentivize school children to make healthier food choices when eating in school. A total of 282 children, aged nine to ten, participated in our study.

Our design includes four treatments. In all treatments, children were presented with five different food trays containing an assortment of snacks; they made four choices from all snacks across the trays. The intervention lasted for three weeks, during which children chose snacks on two pre-scheduled days per week. In the *Baseline* treatment, nothing else was done. In the *Nutritionist* treatment, a nutritionist explained the benefits of healthy eating at the beginning of the first day. In the *Grades* treatment, a clearly-labeled "grade" was associated with each of the trays, with healthier foods receiving higher grades. Finally, in the *Parents* treatment, the children saw the same grades as in the *Grades* treatment, and parents also received weekly reports on their children's average food grade that week; the children knew that their parents would receive their grades. Four months after the end of the intervention, we returned for a surprise visit to measure the longer-term effects after the incentives had been removed.

Because school children routinely compete for grades, we hypothesized that the *Grades* and *Parents* treatments would generate competition between participants. Based on the positive effect of competition on children's choices of fruits and vegetables found in Belot et al. (2016), we conjectured that these two treatments would lead to healthier choices than the *Baseline*. We also expected the information released by the nutritionist to significantly change the children's behavior, and so we predicted differences between the *Nutritionist* treatment and the *Baseline*.

We find an initial improvement in healthy eating in the *Nutritionist* treatment, followed by a downward trend over time. The *Grades* treatment shows no effect on healthy choices at first, but it generates a positive trend that improves behavior over time. Overall, the proportion of healthy choices is 36% in the *Baseline*, 45% in the *Nutritionist* treatment, and 46% in the *Grades* treatment. The *Parents* treatment has far larger effects, with 74% of the choices being healthy foods—more than double the rate of the *Baseline*. Perhaps most critically, these effects persist over time, even after the removal of additional information. In "surprise" sessions conducted four months later (and without any stimulus), the respective proportions of healthy choices are 41%, 47%, 54%, and 69% in *Baseline, Nutritionist, Grades* and *Parents*, respectively. This suggests lasting benefits for the interventions, especially when parents were initially involved.

The contribution of this paper is threefold. First, it proposes a sustainable and inexpensive way of improving healthy choices in children. We show that engaging parents as change agents and providing them (with their children's knowledge) with information about children's food choices substantially increases healthy eating. In the previous literature discussed in the next section, the role of parents has been primarily restricted to two strategies to influence children's food intake: restriction and pressure. These measures tend to backfire: when parents pressure children to eat a food, the children come to like it less; when parents restrict a food, children desire it more and eat more of it after removal of the restriction (DeCosta et al., 2017). In our study, we incorporate parents into the decision process in a non-invasive way. Parents receive information about their children's average grade but not their exact choices. So, while parents certainly influence their children's decisions, they cannot dictate or even precisely monitor their children's snacks during school.

Second, the paper demonstrates longer-lasting effects of the intervention, as the improvements in diet are on average *fully present* four months after the end of the intervention and removal of any extra information during surprise visits, despite these visits falling into the next academic year. Incentivizing children through non-financial methods seems to engender good eating habits or learning, which sheds some light on how to achieve longer-term impacts. To the best of our knowledge, previously only material incentives have been shown to generate a post-intervention effect in the domain of healthy choices.[1]

Finally, this paper also contributes from a methodological perspective. Previous studies providing nutritional information have done so through labels that either simply manipulate attractiveness (Morizet et al., 2012; Pelchat and Pliner, 1995) or convey nutritional information that requires a high level of literacy and numeracy to interpret (Rothman et al., 2006). These interventions have been shown to have either little or no effect at all on choices. An alternative approach has been to use color codes that classify food items into red, yellow, and green depending on their healthiness. This approach shows only slightly better results than the baseline in a hospital cafeteria (Thorndike et al., 2012), using weekday data. There is no conclusive evidence that any color-code interventions have had longer-term effects. Our intervention departs from those in the literature through the use of numerical grades identical to those assigned for school courses. Unlike caloric information, children had copious experience with these numerical grades. Unlike color labels, these numerical grades gave children a yardstick by which they could impress their parents as well as compete with one

[1] Yet even this evidence is inconclusive, as shown in the next section.

another. Just like children (and adults) compete in exercise, they turn out to compete in heathy eating when provided a means of keeping score.

Our intervention, particularly the treatment involving parents, points to a low-cost and highly-effective means of battling the problems of poor childhood diet and obesity that have been expanding in the developed world. If sustainable healthy eating habits are established in childhood, we should expect less obesity in the adult population as well. Our intervention appears to be a promising avenue to achieve this.

The structure of the remainder of this paper is as follows. Section 2 offers a literature review of previous related work, while Section 3 describes our experimental design in detail. We present our hypotheses in Section 4 and results in Section 5. Finally, we discuss these results and conclude in Section 6.

2. Previous literature

Previous literature on parental control of food choices has investigated two strategies used by parents to influence their children's behavior: restriction and pressure. Restriction has been found to have negative consequences on eating behavior; prohibition leads to increased desire and consumption when the forbidden food becomes available (Fisher and Birch, 1999; Jansen et al., 2007, 2008; Ogden et al., 2013). We know of no studies identifying restriction as an effective strategy for effecting dietary change. Similarly, pressuring children to eat particular foods has not been found successful (Galloway et al., 2006; Rigal et al., 2016).

Researchers also have explored the effectiveness of providing information in improving healthy choices. Wisdom et al. (2010) studied the effect of calorie information and "asymmetric paternalism" on the type of food picked in a chain restaurant. They find that providing the calorie information had no significant effect on the probability of picking a low-calorie sandwich. However, participants were more likely to pick it in a paternalism treatment in which it was more convenient to pick the low-calorie sandwich. In other cases, putting calorie information on food has shown more positive effects. Bollinger et al. (2011) find that mandatory calorie posting in Starbucks led to reductions in calories per transaction, which persisted for ten months after implementation. Fichera and von Hinke (2020) find that introducing nutrition labels on a retailer's store-brand products reduced the monthly consumption of calories from labelled store-brand foods. On the other hand, Downs et al. (2013) and Roberto et al. (2010) report on settings in which the provision of calorie information did not improve healthy eating.[2]

Many nutritional labels require a high level of literacy and numeracy to interpret (Rothman et al., 2006), which may inhibit their effectiveness. Some research suggests that simpler labels have greater effects. Vyth et al. (2011) investigated (in Dutch cafeterias) whether labeling foods with the Choices nutrition logo (a logo introduced in several large catering organizations in the Netherlands in 2006) affected food choices. While they find a significant effect of the logo on fruit sales, they find no significant effects on other items (soup, bread, salad and snacks). Thorndike et al. (2012) introduced a color-code system in a Massachusetts General Hospital cafeteria over 3 months, using the labels red, yellow and green to label unhealthy, less unhealthy and healthy items, respectively. The sales of red items decreased by 9.2%, and green ones increased by 4.5%. Two years of keeping the labels and positioning of foods to make healthy (unhealthy) items more (less) accessible led to a modest shift to green items (Thorndike et al., 2014). Because the incentive was kept in place for the *entire duration* of the study (unlike in our design), it cannot speak to the potential persistence of effects after the end of the intervention.

The previous studies on the effect of information on healthy choices primarily target an adult population. Mora and Lopez-Valcarcel (2018) study the impact of a nutrition workshop on adolescents' food and drink choices. The workshop was somewhat effective, with 31% of the students in the workshop chose the healthy option compared to only 20% in the control group. However, that paper does not analyze the effect of the workshop in the longer-term, nor does it explore the use of grades or parents.[3]

Finally, although not directly related to our paper, the "first generation" of interventions in behavioral economics provided material incentives aiming to promote healthy behavior.[4] These incentives have been shown to have an effect in the short-term, with little exploration of long-term effects. For example, Just and Price (2013) conducted a field experiment at fifteen elementary schools in Utah. They rewarded students for eating a serving of fruit or vegetables per day in various different ways and found a 27% increase in the fraction of children eating at least one serving of fruit or vegetables when any incentive was offered. But they do not study the effect of the short-term rewards on the long-term behavior, which could potentially even go in the opposite direction through crowding-out of intrinsic motivation (Deci and Ryan, 1985).

Belot et al. (2016) conducted a field experiment in schools in England to test the effectiveness of two different incentive schemes on choosing fruit and vegetables: a piece rate and a tournament. Both schemes provided participants with a sticker for choosing a fruit or vegetable. In the piece-rate treatment, subjects who collected four stickers over the week received a prize. In the tournament

[2] For a meta-analysis on the effect of food labelling on food choices, see Cecchini and Warin (2016).

[3] There are other studies demonstrating that giving young adolescents information about the healthiness of food items can increase their consumption of healthy food (for a meta-analysis see, Dudley, Cotton & Peralta 2015).

[4] Material incentives have been used to promote healthy behaviors other than the one addressed in this paper. For example, Charness and Gneezy (2009) were the first to demonstrate that financial incentives can have long-lasting positive effects on individuals' willingness to exercise that persist even after such incentives are removed. John *et al.* (2011) and Volpp *et al.* (2008) provide evidence of the effectiveness of financial incentives for weight loss. However, this effect is not robust (Cawley and Price, 2013). Volpp *et al.* (2009) show that financial incentives significantly decrease the smoking rate during the intervention (when the incentives were in play), but not afterwards. Of course, monetary incentives are difficult to roll out on a large scale and daunting to maintain in the long run.

treatment, only the child with the largest number of stickers won the prize. Results show that there was little effect of the piece-rate scheme on choices yet a positive effect of the competition mechanism. However, choices were not significantly different in the long term once the incentives were removed.

List and Samek (2015) performed a field experiment in Chicago in which children were given a choice between a dried fruit cup (healthy item) and a cookie (unhealthy item). Cash incentives for healthy eating proved effective: the proportion of children choosing the healthy snack increased from 17% in the baseline to nearly 80% with monetary incentives. The authors find some evidence of short-run post-intervention effects in the form of behavior one week after removal of the incentives but do not provide any evidence on longer-term efficacy. Compared to List and Samek (2015), our paper also studies the role of parents as potential change agents.

Loewenstein et al. (2016) conducted a field experiment in elementary schools in Utah. During a three- or five-week reward period, children eating at least one serving of fruits or vegetables received a token worth 25 cents that could be redeemed at a school store, school carnival, or book fair. The authors find a strong positive effect of the reward on diet: the percentage of children who ate at least one serving of fruits or vegetables increased from 38 to 40% to 76–80%. However, two months after removal of the incentives, this percentage dropped to 48–54%: roughly 70% of the increase over the baseline dissipated over the two months following the intervention. In our paper, we are able to study the persistence of effects from different non-monetary incentives.

Finally, Belot et al. (2019) conducted a field experiment to examine the malleability of dietary habits. Low-income families in the UK participated in one of two treatments. In the "Meal" treatment, families received free groceries and were asked to cook five healthy meals per week. In the "Snack" treatment, families were asked to reduce snacking and eat at regular times. The two treatments were implemented for 12 weeks. Children in both treatments reduced their body mass index and sugar intake compared to children in the control group. However, there was no strong evidence that children's preferences changed in favor of healthier foods. One potential explanation for the patterns observed may be that the interventions had an impact on what the parents fed their children, rather than on children's preferences. In our study, we observe and target the latter.[5]

3. Experimental design and procedures

3.1. Experimental design

Our experimental design consists of four different treatments (in our case, four different information conditions):

- *Baseline*: Children participating in the experiment were presented with five different food trays. Each tray included five different food items of similar nutritional value, selected by a nutritionist.[6] Subjects had to pick four food items from any combination of the trays. They could choose items from the same or from different trays, and were able to select more than one item from the same tray and more than one unit of the same item.
- *Nutritionist Treatment (NT, hereafter)*: Similar to Baseline, except that a nutritionist explained the benefits of healthy eating on the first day of the experiment before participants made any decisions. The nutritionist did not encourage children to pick any particular item from the trays.[7] The same nutritionist gave the same talk at all schools in this treatment.[8]
- *Grades Treatment (GT, hereafter)*: Similar to Baseline, except that each of the five trays was labelled with a different "grade" that described its items' nutritional content. These grades were analogous to those used to mark children's academic schoolwork (Spanish grades at schools range from 0 to 10, so the five trays had assigned marks of 0, 2.5, 5, 7.5, and 10). Each item in a given tray had the same grade assigned. Children were shown the grades associated with each tray and told that these grades represented the nutritional value of the items in the tray.[9]
- *Parents Treatment (PT, hereafter)*: Similar to GT, except that parents received information about the average mark (linked to the nutritional composition) of their children. Children received exactly the same information as in GT. In addition (and before

[5] There have been other interventions that explore alternative ways to improve healthy consumption. Fletcher *et al.* (2010) find a moderate effect of soft drinks taxes on children's consumption. Griffith *et al.* (2018) show that the introduction of vouchers for fruit, vegetables and milk increases spending on fruit and vegetables. They find that the vouchers are more effective than the equivalent cash benefit.

[6] See Appendix A.1 for the nutritionist's justification of the allocation of food items to trays, and Appendix A.2 for a summary of the food nutritional information. The same nutritionist allocated grades to trays in *GT* and *PT*. Appendix C.1 presents the detailed composition of the trays and the corresponding grade associated with each one.

[7] A summary of the nutritionist's talk is reported in Appendix B. The reason why we chose the nutritionist to intervene only once was twofold. First, we aimed to design an experiment that could be easily implemented by policy makers later on. While having the grades system in place would be a one-time effort and everything could be done automatically, having a nutritionist come every day to give a speech and train all the participants would be harder and more expensive to implement. Second, having an external person come every day of the experiment to give a speech could produce a demand effect that would make our results harder to interpret.

[8] We chose to have the nutritionist talk only once because we felt that having the same talk would have felt artificial and repetitive. This decision came at the cost of potentially disadvantaging this treatment relative to other treatments in which the children received feedback once or twice per week. In hindsight, we would have had a slightly different talk by the nutritionist each week. We thank an anonymous reviewer for raising this issue.

[9] Students were not encouraged to earn high grades in the experiment, nor were they told their average grade after the decision (although they could easily compute it).

selecting the items), children were made aware of the fact that their parents would receive a weekly report about their average grade. Note that parents did not learn their children's choices but merely the average grade for the week.[10]

3.2. Procedures

The experiment was conducted simultaneously in 12 Spanish elementary schools participating in a European Union (EU) program aiming to encourage healthy eating at schools. Our experiment was run as an additional activity within the EU program in these schools, so that children would see their decisions as less artificial, mitigating potential concerns about the external validity of our findings. Schools participating in the EU program received, during the academic year and depending on the week, different types of fruits and vegetables to be distributed amongst the children.[11] Students received the corresponding food, and their only choice was whether to eat it. We felt that the children in the program would see the experiment as an alternative activity in which the main difference is that they could choose the food they preferred to consume.[12]

All selected schools were located in Jaén, a city in the South of Spain. They were chosen to produce a large and diverse pool of potential subjects, as measured by the demographic and socioeconomic characteristics of the students and their families, and all lie within few kilometers of one another.[13] When selecting the schools, we tried to avoid extreme socioeconomic conditions.[14] Each participating school was randomly assigned to one of the treatments described above, so that all participants from a school faced the same information conditions during the intervention.[15] Only one class from each school was randomly selected to participate. The total sample of 282 students from the fifth grade (aged 9 to 10) was almost evenly distributed over the 12 classes. We chose this age range because, at this age, children are old enough to understand the meaning and importance of grades, but still young enough to care about their parents' opinions (Larson et al., 1996; Allen et al., 2005).

On the first day of the intervention in each school, experimental subjects belonging to the same class were gathered in a room (Room A). Participants were not given any information about the experiment. Each student was then—independently and sequentially—asked to move to another room (Room B) in the company of one of the experimenters. In this second room, the subject was walked through a short orientation session—the details of which depended on the subject's treatment assignment. Next, the student picked the four food items he or she preferred. Finally, the student was taken to a third room (Room C), where he or she joined classmates who had already completed the task. This procedure ensured that the participants' food choices were not directly influenced by the choices of other participants who chose before them on the day.

An observer from the research team recorded the students' dietary choices and grades. To mitigate social-desirability effects, the bags used to store the food were transparent, which allowed the researcher to record the childrens' snacks without being directly involved in choices. The intervention was always conducted right before the main school break at 11 a.m., when children go outside to get exercise; they also eat a snack during the break to tide them over until lunch, which—in Spain—starts at 2:00 - 3:00 p.m.[16,17] To make the food-decision salient, the parents were instructed not to prepare any snacks for their children for the days of the intervention.[18] In this way, children chose the food they would eat during the break that day. A staff member presided over Room C, where children congregated after making their choices. This person observed that no food was wasted (although children could store food for later consumption). As a result, most of the items chosen in Room B were consumed in Room C.[19] We collected data twice a week for three weeks and so have (up to) six observations per subject.[20]

To ensure similar conditions across treatments, parents received the same information about the experiment, signed the same consent, and provided the same contact details before the experiment in each treatment. Parents in *PT* only received information about their child's average grade in a given week; they received no information about the choices of other children.

After the sixth day of the experiment, we gave post-experimental questionnaires to children, teachers and parents, in order to obtain information about participants' socioeconomic variables, self-control indicators, and eating habits (see Appendix D.1 to D.3 for the

[10] The reason why we provided only average grade over the week rather than daily grades was to not overwhelm parents with information.

[11] In 2018, 79,000 schools across Europe involving over 30 million children participated in this program.

[12] This experiment was run as a substitute of the EU program. This means that, while the experiment was conducted, during school time, children did not have access to any type of fruit or vegetable that was not included in the trays. We believe that decisions would be seen as less artificial in such an environment than if the experiment was run in schools in which no activity related to food was previously conducted.

[13] Appendix C.2 provides a map of the city and the location of each school.

[14] Note that every school that was approached agreed to participate in the study. This should help to mitigate potential concerns about some sort of self-selection bias. Riener *et al.* (2020) show, for a large-scale study in Germany, that there is practically no self-selection of schools into study participation.

[15] We are not too concerned about contamination across schools, since children typically went directly home from school and in any event were quite young and thus highly unlikely to socialize with children from other schools.

[16] Only a fraction of Spanish students (22%) eats lunch at the school canteen. For those who do, the schools provide a set menu for lunch.

[17] After school ends at 2:00 pm, the children have lunch. They do not eat between their 11 am snack and lunch. Children of this age usually have dinner around 8:30-9:00 pm, preceded by another snack around 5:30 pm.

[18] While we could not enforce that children didn't bring any food to school, we ensured that they did not have access to their backpacks from the moment that they participated in the experiment until after the school break was over.

[19] We did not find any differences across treatments in the amount of food saved for later consumption.

[20] We collected data on the same days of the week for each school. That is, if one school was allocated to the Monday-Wednesday (Tuesday-Thursday) schedule, this schedule would remain the same over the three-week period. The same procedure was used to collect the data on each day.

complete questionnaires).[21] To assess the students' socioeconomic backgrounds, we utilized a questionnaire administered by the Andalusian Government, which their parents completed. The set of variables includes i) the highest educational level achieved by the parents, ii) proxy variables for the economic level of the household, and iii) parents' involvement in the school-related activities (homework) of their child. For the self-control indicators, following Tsukayama et al. (2013), children answered questions related to their level of interpersonal and schoolwork impulsivity.[22] For eating habits, parents at home filled out a short survey designed by a nutritionist. We collected information regarding the weekly consumption of a variety of food items belonging to five different food groups (milk and dairy products, carbohydrates, proteins, fruit and vegetables, and fats and sugars). In the teacher questionnaire, we just gathered information regarding students' average performance in class, average attendance to the school, and a measure of self-control.

Table E.1 in Appendix E presents descriptive statistics for variables capturing children's dietary habits and background characteristics, built from the responses to the post-experimental questionnaires. Covariate balance checks reported in Tables E.2 and E.3 in the Appendix reveal no systematic imbalances in either the initial dietary habits or socioeconomic characteristics across the subjects assigned to the four treatment arms, indicating that randomization was successful.[23,24] It is worth noting, however, that while the groups are well-balanced in a statistical sense, the descriptive statistics suggest that they differ in a qualitative sense.[25] In order to try to account for these differences, some of the regression models reported in Section 5.3 incorporate control variables (Gelber and Zelen, 1986; Raab and Butcher, 2005). We also compare the results obtained from models with and without controls; the fact that the main substantive findings hold across specifications highlights the robustness of our results and reinforces our confidence the validity of our conclusions (Bruhn and McKenzie, 2009).

Another concern is that, since the data on children's dietary habits was conducted after the experiment, the lack of systematic differences in children's food consumption patterns (Tables E.2 and E.3) may be the result of "desirability bias". Given the nature of our treatments, parents may have been primed to report that their children's dietary habits were healthier than they really were.

To address this issue, Tables E.4 and E.5 in the Appendix replicate the covariate balance checks using additional data collected two years after the original experiment. In this robustness check we use the responses on dietary habits from 189 (new) children who were the same age as the original participants, and who were enrolled in the same schools that participated in the original experiment. The evidence from Tables E.4 and E.5 indicates again that there were no systematic imbalances in the dietary habits of subjects belonging to schools assigned to the different treatment arms, suggesting that there was no desirability bias.

Finally, four months after the end of the intervention and during the next academic year, subjects participated in a *surprise* session, which was conducted as an (unannounced) one-shot *Baseline* for all subjects.[26] The idea was to study whether the effect of the different information conditions persisted for some time after the intervention. We varied the data-collections procedures in the surprise session from those of the main experiment, so children did not have the feeling of repeating the same decisions they made during the experiment. Decisions in the surprise session were made in different room than in the main experiment. The trays were presented to participants in the same order as before, but the layout was modified. Note that in the surprise session children who participated in *PT* and *GT* saw that there were no grades associated with the trays (as opposed to all other days they participated). Thus, no information could be released to the parents from the surprise session.

4. Theoretical predictions and hypotheses

Let $X = \{x \in Z_+^L : \sum_{l=1}^{L} x_l = k\}$, for some integer $k > 1$, be the set of non-negative integer combinations of the L snacks available to the student. Let $\ominus \subset R^L$ be a set of real-valued parameters that measure the healthiness of the L available snacks. For example, θ_l might be equal to minus the calories of snack l, or it might take on the value of one if and only if l belongs to the healthiest quintile of snacks. The utility-from-consumption function $v^i : X \times \ominus \times R \rightarrow R$describes student i's direct preferences over snack bundles, which depend upon healthiness of the snacks and an additional parameter described below. We assume that a student i who chooses bundle $x^i \in X$ at any given date derives *direct utility*:

[21] The completion rate of the questionnaires for parents was 91.2%, and for children and teachers the rate was 91.5%.

[22] We used a reduced version of the Impulsivity Scale for Children (ISC) questionnaire (Tsukayama *et al.*, 2013).

[23] The covariate balance checks reported in Table E.2 account for the clustered nature of our randomization using the methodology developed by Hansen and Bowers (2008), based on *Fisher's* randomization inference. For robustness, in Table E.3 we fit a multinomial logit model for treatment assignment (e.g. Gerber *et al.* 2009), using Ibragimov and Muller's (2010) approach to modelling clustered data while accounting for the small number of schools in our sample (see also Esarey and Menger, 2019). The results in both tables suggest the absence of any systematic covariate imbalances across treatment arms.

[24] The dietary habits reported in Table E.1 capture the habits of the household, and not just those of experimental participants. As a consequence, we did not expect these habits to change after our intervention. This feature reconciles the absence of systematic imbalances in Tables E.2 and E.3 with the main results from the paper.

[25] We thank an anonymous reviewer for bringing this issue to our attention.

[26] Because parents received no advance notice of this surprise session, they would have been expected to pack snacks for the day, unlike in the previous sessions. This should not have affected decision by the children, who did not have access to their backpacks until after the school break and could not eat during class.

$$v\left(x^{i},\ \theta,\alpha^{i}\right)=t^{i}\left(x^{i}\right)+\alpha^{i}\sum_{l=1}^{L}\theta_{l}x_{l}^{i} \tag{1}$$

where $t^i(x^i)$ denotes student i's real-valued enjoyment of the *taste* of the bundle $x^i = (x^i_1, ..., x^i_L)$, $h(x^i,\ \theta) = \sum_{l=1}^{L}\theta_l x^i_l$ represents the *healthiness* of the bundle, and $\alpha^i > 0$ measures student i's concern for health.[27] Whereas the health value $h(x^i,\ \theta)$ is assumed to be additively separable in the quantities of the different snacks, x^i_l, the taste utility $t^i(x^i)$ is not; this allows the student to care about the composition of his bundle (e.g., he might not find four copies of his tastiest single snack to be the tastiest bundle). Student i knows his preference parameter α^i but may be uncertain about the health consequences $\theta = (\theta_1, \ldots, \theta_L)$ of the different snacks. A student who does not know θ forms beliefs encapsulated by expectations $\mathrm{E}[\theta]$.

In the *Baseline*, children make choices without knowing θ. Suppressing the student index i for parsimony, students solve:

$$\max_{x\in X} E_{\theta}[v(x,\theta,\alpha)] \Leftrightarrow \max_{x\in X}\left\{t(x)+\alpha\sum_{l=1}^{L}E[\theta_{l}]x_{l}\right\} \Leftrightarrow \max_{x\in X} v(x,\alpha E[\theta],1). \tag{2}$$

so that a student who maximizes the expectation over θ of utility simply maximizes the direct utility given the expected value of θ. Let $x^*(\theta,\alpha)$ denote the student's optimal choice as a function of the parameters θ and α; for simplicity, we take this optimal choice to be single-valued.[28]

Fact 1. *For each snack l, the optimal consumption level $x^*_l(\theta_l, \theta_{-l},\ \alpha)$ is a non-decreasing function of θ_l.*[29] The proof is in Appendix F.

Fact 2. The healthiness of the optimal choice $h(x^*(\theta,\ \alpha), \theta) = \sum_{l=1}^{L}\theta_l x^*_l(\theta;\ \alpha)$ is a non-decreasing function of α. The proof is in Appendix F.

The *Nutritionist Treatment* makes nutrition salient to the children, which we conceptualize as increasing the weight α that they assign to health when choosing their snacks. This gives us:

Prediction 1. : The *Nutritionist Treatment* gives healthier choices than the *Baseline*.

The *Grades Treatment* makes two conceptually distinct changes to the *Baseline*. First, it reveals information about θ. Second, it gives students a yardstick by which to compete. We examine each of these effects in turn.

Suppose that students know the nutrition of all foods other than Snack 1. Fact 1 implies that a student who learns good news about Snack 1 consumes no less of it after learning θ, whereas a student who learns bad news about Snack 1 consumes no more of it. Revealing nutritional information about Snack 1 has an ambiguous effect on the overall healthiness of the student's chosen bundle. To see how better information about health may worsen health outcomes, consider a case in which a student avoids tasty Snack 1 in the *Baseline* under the expectation that is very unhealthy (i.e., $E[\theta_1]$ is low). If Snack 1 turns out to be healthier than predicted, namely $\theta_1 > E[\theta_1]$, then the student may substitute away from healthier foods into Snack 1, which decreases the overall healthiness of the chosen bundle.

In certain cases, however, the effect is unambiguous.

Fact 3. If for each good l, (i) $E[\theta_l] = \overline{\theta}$, or if (ii) α is sufficiently large, then $h(x^*(\theta,\alpha),\theta) \geq h(x^*(E[\theta],\ \alpha),\theta)$. The proof is in Appendix F.

In the *Grades Treatment*, grades not only provide students with information about the healthiness of the different foods, but they also constitute a yardstick by which students can compete with one another. To capture the competitive angle of the *Grades treatment*, we enrich students' preferences to incorporate some concern about their grades relative to the grades of their peers. Let $r(h(x^i,\theta), (h(x^j,\theta))_{j\neq i})$ measure that part of i's utility that derives from a comparison of the healthiness of i's snacks relative to the healthiness of his or her peers' snacks. We assume that r increases in its first argument, $h(x^i,\ \theta)$, that it is continuous in each argument, and that it is, for each $j \neq i$, supermodular in $h(x^i,\ \theta)$ and $h(x^j,\ \theta)$: healthy eating by student j is a *strategic complement* (Bulow et al., 1985) to healthy eating by student i. Taking other students' snack choices $(x^j)_{j\neq i}$ as given, student i solves

$$\max_{x^{i}\in X}\left\{v\left(x^{i},\theta,\alpha^{i}\right)+\beta r\left(\theta\cdot x^{i},\left(\theta\cdot x^{j}\right)_{j\neq i}\right)\right\} \tag{3}$$

for $\beta > 0$. In a Nash equilibrium in this game, each student i maximizes his or her utility (3) by taking other students' equilibrium choices as given. Due to strategic complements in healthy eating, the game may have multiple equilibria. In one equilibrium, each student may eat healthy and be motivated by his or her peers' healthy eating; in another equilibrium, each student may eat less healthy and feel insulated by his or her peers' unhealthy eating.

Fact 4. In the most and least healthy Nash equilibria of the game induced by the Grades Treatment, the healthiness of students'

[27] It would be straightforward to generalize the model to make utility from consumption an increasing but not necessarily additive, function of t(x) and h(x,θ). Note too that α may incorporate some discounting of the future health benefits of healthy eating; since the decision problem is static, time-consistent and time-inconsistent discounting are observationally equivalent.

[28] Although the optimal choice will be multi-valued for a non-generic set of parameter values, we abstract from this detail to simplify exposition.

[29] Throughout, the subscript $-l$ refers to snacks $j \neq l$.

choices is non-decreasing in β.[30]

When students have very accurate initial beliefs about nutrition, then the *Grades Treatment* does not change their beliefs about θ; we show in the next section that students in our experiment do indeed have rather accurate beliefs about the nutrition of the various snacks without seeing any grades. Consequently, moving from the *Baseline* to the *Grades Treatment* corresponds to an increase in β, which increases healthiness by Fact 4.

Prediction 2. **:** The *Grades Treatment* gives healthier choices than the *Baseline*.

Finally, we conceptualize the *Parents Treatment* as introducing the same β parameter as the *Grades Treatment* while also increasing α relative to the *Grades Treatment*. The idea behind this conceptualization is that parents value their children's healthy eating more than the children do themselves, and that children seek to please their parents. By providing parents with information about children's eating, the *Parents Treatment* leads children to weight their parents' welfare more heavily in their decision-making, which in turn leads the children to up-weight healthiness in their own maximization problem.[31] This gives:

Prediction 3. **:** The *Parents Treatment* gives healthier choices than the *Grades Treatment*.

5. Results

5.1. Children's behavior on the first day

We start with an overview of the decisions made by the participants in each treatment. Table 1 presents a summary of the average grades (from food choices) and the proportion of healthy choices, defined as items assigned grades of 7.5 or above.[32] Note that children do not know anything on the first day of the experiment regarding what will happen in the experiment after they make their decision. In particular, they do not even know that they will have the opportunity to talk to their classmates right after making their decision. Hence, when running non-parametric tests, we can consider individual decisions to be independent from each other.

As seen in Table 1, the average grade in *NT* (6.38) is higher than in *Baseline* (5.63). This difference is close to statistical significance using individual data and a Mann-Whitney test ($Z = -1.89$, $p = 0.058$, two-tailed test).[33] So, the nutritionist's talk seems to initially increase the rate of healthy food choices of children. We note that this initial increase might have been better sustained had the nutritionist spoken more than once (see footnote 8). However, providing information in the form of simply having grades assigned to the food does not result in an increase in the average health quality chosen, since the *GT* average is suggestively *lower* than the Baseline average ($Z = -1.63$, $p = 0.103$).[34] Hence, these initial results support Prediction 1, but not Prediction 2.

Differences in grades are dramatically larger when information is released to the parents. Table 1 shows that the average grade in *PT* (7.57) is far higher than in any other treatment. A Mann-Whitney test on individual data shows significant differences for each pairwise comparison to the *PT* treatment ($p = 0.000$ for every comparison to *Baseline, NT* and *GT*).[35] These results support Prediction 3 and thus the idea that involving parents in the process by making them aware of their children's decisions is a very strong mechanism for encouraging healthy behavior.[36]

Turning to the proportion of initial healthy choices, results are the same as those using the average grades. As shown in Table 1, the percentage of healthy items chosen in *Baseline* is 47%. This percentage increases in *NT* (54%) and reaches its highest value in *PT* (70%). By contrast, the percentage of healthy choices is lowest in GT (40%).

Result 1. **:** *Providing information to children about the nutritional value of the food (in NT) has a moderate, but nevertheless significant, initial effect on choices. Assigning grades to the food trays (in GT) does not have a significant effect in comparison to the Baseline. However, releasing the information about the child's average grade to parents (in PT) has a very strong initial effect on subjects' decisions, greatly increasing the consumption of healthy food items.*

[30] The results in Echenique (2002) imply that any method of selecting from multiple equilibria that does not produce non-decreasing comparative statics in β relies on "unstable" equilibria, which themselves are unlikely to occur in our setting.

[31] Formally, suppose that a student in *GT* maximizes $u(x,\theta,\alpha,\beta)$, whereas his parents value nutrition more by having preferences $u(x,\theta,\alpha',\beta)$, for $\alpha' > \alpha$. When the student in *PT* gives his parents' utility the weight $\delta > 0$ by maximizing $u(x,\theta,\alpha,\beta) + \delta u(x,\theta,\alpha',\beta)$, this is equivalent to maximizing $u(x,\theta,\alpha'',\beta)$, for $\alpha'' = \frac{\alpha + \delta\alpha'}{1+\delta} > \alpha$.

[32] As a robustness check, we replicated this analysis for the case in which we define as healthy choices those items that were assigned a grade of 10 and for the case in which healthy choices are considered those with a grade of 5 and above. The main results (available upon request) remain qualitatively similar in both cases.

[33] Henceforth, all reported statistical tests in this paper are two-tailed. We round all p-values to the nearest third decimal place.

[34] These results hold also when accounting for the clustered nature of individual observations (i.e., children within schools). Cluster (school)-adjusted p-values from tests for differences in mean grades are $p = 0.099$ for the comparison *NT vs Baseline,* and $p = 0.023$ for the comparison *NT vs GT*. Difference between the grades in *GT* and *Baseline* are again not statistically significant after adjusting for clustering ($p = 0.312$).

[35] These differences are also significant after accounting for clustering: the cluster (school)-adjusted p-values are $p = 0.004$ for the *PT-Baseline* comparison, $p = 0.002$ for the *PT-GT* comparison, and $p = 0.046$ for the *PT-NT* comparison.

[36] This conclusion is also confirmed when we look at school-level data. As we show in Table E.6 in the Appendix, both the average grades and the proportion of healthy choices in each of the schools in *PT* are higher than in any school in the other three treatments.

Table 1
Average grades and healthy food choices on the first day, by treatment.

Treatment	Observations	Subjects	Average grade	Proportion of healthy choices
Baseline	73	73	5.63 (2.27)	0.47 (0.30)
NT	71	71	6.38 (1.78)	0.54 (0.26)
GT	65	65	5.22 (1.82)	0.40 (0.26)
PT	73	73	7.57 (1.59)	0.70 (0.24)

Notes: Standard deviations are reported in parentheses.

5.2. Dynamics during the intervention period

This section focuses on the dynamics of behavior across the four treatments. Fig. 1 represents the average percentage of healthy choices over time, i.e., over the six experimental days across the three weeks of the intervention. Like Table 1, Fig. 1 codes items assigned a grade of 7.5 or 10 as healthy choices.[37]

Fig. 1 shows a positive general trend in *GT* and *PT*, with the improvement mainly from Day 1 to Day 3. Again, the percentage of healthy choices in the first period in *PT* is much larger than in *Baseline, NT*, and *GT*. Apparently just knowing that their parents will view their food grades has a large effect on children's food choices. Lacking grades, the *Baseline* and *NT* treatments showed negative trends in healthy eating.[38] In Appendix G we offer an *ex-post* rationalization of the dynamics observed in Fig. 1, based on our static model from Section 4.

5.3. Analyzing children's behavior

Table 2 reports parameter estimates from probit models in which the dependent variable is a dummy taking the value 1 if subject i made a "healthy choice" (i.e., subject i's choice had an average grade of 7.5 or higher) in period t, and 0 otherwise. All specifications use cluster-robust standard errors at the school level to account for heteroskedasticity and intra-school correlation.[39]

Since the number of schools in our sample is quite small, conventional inference methods may be unreliable: large-sample assumptions do not hold, and standard errors may be biased downwards (Donald and Lang, 2007; Cameron et al., 2008). Hence, the table also reports p-values for Wald hypothesis tests computed according to Kline and Santos' (2012) score bootstrap procedure for clustered data, which adapts the wild bootstrap-t procedure developed by Cameron et al. (2008) to maximum likelihood estimators. This approach is useful for small sample sizes like ours because it does not rely on asymptotic approximations, and it has been shown to perform well with very small numbers of clusters (Kline and Santos, 2012).[40]

The explanatory variables in the first column of Table 2 are indicators for treatments *NT, GT*, and *PT*, taking *Baseline* as the benchmark. The probability of a healthy choice increases in all three treatments compared to *Baseline*, even accounting for only having few schools in our sample. This result suggests that participants generally responded to different information conditions. Moreover, and in line with the previous results, we observe that subjects' probability of choosing healthy food is significantly larger in *PT* than in *GT* ($\chi^2 = 17.832$, $p = 0.000$; test for equality of coefficients) and in *NT* ($\chi^2 = 22.770$, $p = 0.000$).[41] There is no statistically-significant difference between the effect of the grades-only treatment and the effect of the nutritionist, though ($\chi^2 = 0.504$, $p = 0.478$; test for equality of coefficients between *GT* and *NT*).

The treatment effects remain robust if we incorporate additional controls in column (2): *Male*, Average School GPA (a measure of academic performance), and two measures for impulsivity levels. For the latter, we built two indexes that measure two domain-specific impulsivity levels (Tsukayama et al., 2013): *Personal Impulsivity* and *School Impulsivity*, which we include as co-variates. *Average School GPA* is significantly and positively associated with the probability that subjects make healthy food choices. Among the impulsivity measures, only *Personal Impulsivity* is (marginally) significantly correlated with the probability that subjects choose healthy food items.

Column (3) replicates the analysis in column (2), but also takes into account the socioeconomic level of a child's family.

[37] When the analysis is replicated using children's average grades and a stricter definition of healthy choices (i.e., defining as healthy choices those items that were assigned a grade of 10), the main patterns emerging from Figure 1 remain unchanged. Results are available upon request.

[38] All trends are statistically significant ($Z = 4.179$, $p < 0.001$; $Z = 4.252$, $p < 0.001$; $Z = 2.723$, $p = 0.006$; and $Z = 4.995$, $p < 0.001$; Cochran-Armitage test for *Baseline, NT, GT*, and *PT*, respectively). We again caveat that the effects seen in the NT may be best interpreted as a lower bound on nutritional education, since the nutritionist spoke only at the first meeting.

[39] Since the explanatory variables included in the models in Table 2 are time-invariant, these specifications do not include individual or school fixed effects. For robustness, Table E.9 in the Appendix reports estimates from multi-level probit models including random effects to account for time-invariant individual heterogeneity, within-school correlation, and common temporal shocks. Multi-level models provide an alternative tool for dealing with clustered data (e.g., Primo *et al.*, 2007), and Bayesian inferential methods yield accurate estimates for such hierarchical models even with as few as 3 clusters (Gelman, 2006). The main results are like those presented in Table 2.

[40] Tables E.10-E.11 in the Appendix report estimates from linear regression models in which the dependent variable is subject i's grade in period t. Table E.10 resorts Cameron *et al.*'s (2008) wild bootstrap-t procedure for clustered data to account for the small number of schools in our sample, while Table E.11 presents estimates from multi-level regression models. The key substantive results from these additional specifications are similar to those in Table 2.

[41] Recall that our theoretical model does not rank PT or GT relative to NT.

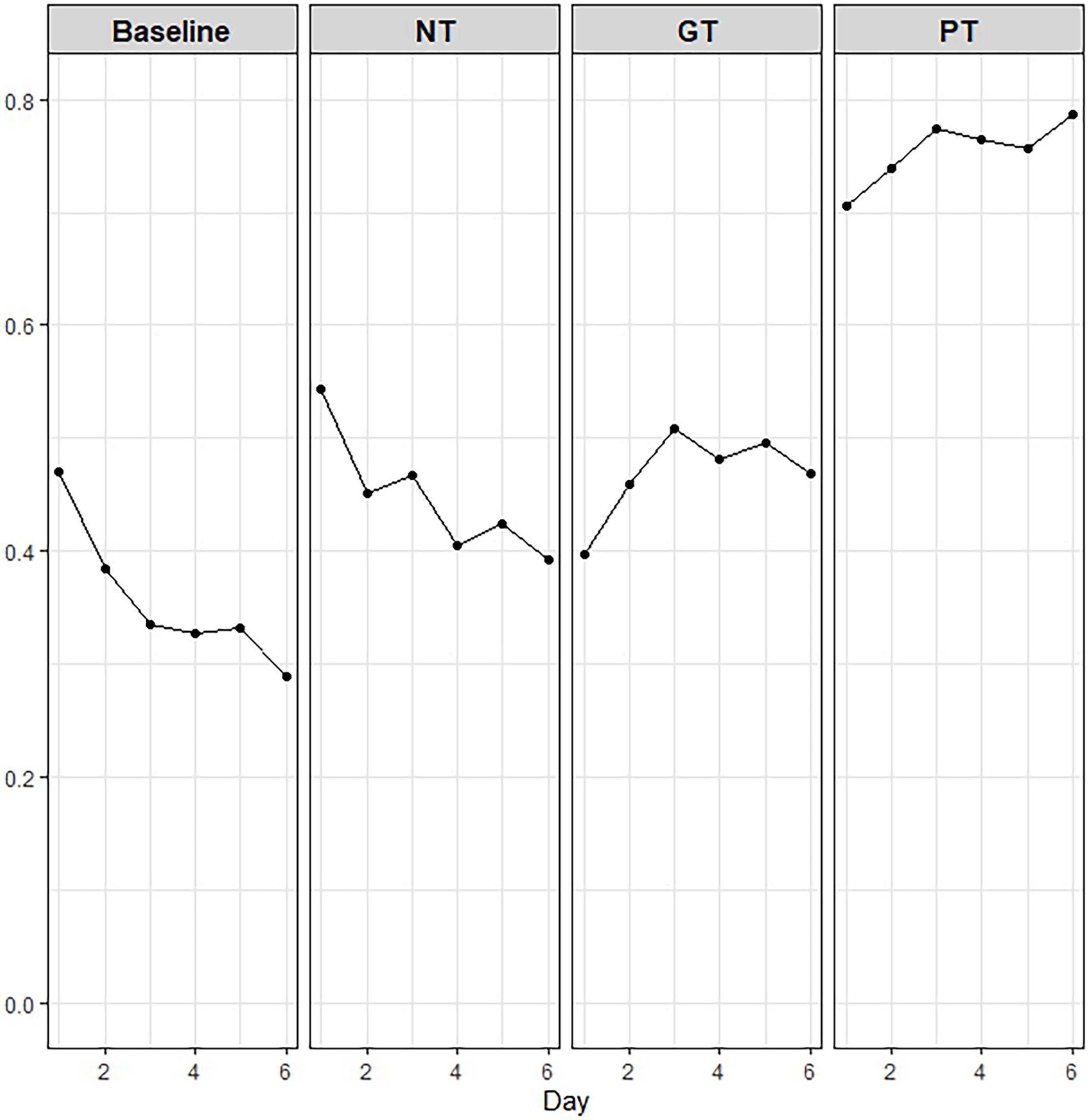

Fig. 1. Percentage of healthy choices over time.

Specifically, *Parents hold University degree* is a binary variable that takes value 1 if at least one of the parents achieved graduate or postgraduate education and 0 otherwise; and *Household Commodities* is a composite index of the number of durable goods (laptop and desktop computers, tablets, electronic books, cars) available in each student's home.[42,43] We observe that a higher educational level of

[42] Using consumption-based measures of household welfare or resources – instead of relying on self-reported income – is a common practice since income is more likely to be under-reported (e.g., Meyer and Sullivan, 2011).

[43] A concern regarding the specifications in columns (2) and (3) is that the variables measuring students' impulsivity and households' socio-economic characteristics were obtained from the response of subjects, teachers, and parents to surveys conducted post-treatment. While it is highly unlikely that the treatments affected subjects' impulsivity or fixed family characteristics, it is in principle conceivable that the survey responses (or unobservables related to these characteristics) were influenced by the experimental intervention. That said, the fact that treatment-effect estimates remain significant across columns (1) to (3) – see also Tables E.9 - E.11 in the Appendix – mitigates this concern.

Table 2
Probit regression on the probability of choosing healthy food items.

	(1)	(2)	(3)
Constant	−1.34***	−1.69***	−1.66***
	(0.08)	(0.40)	(0.48)
Wild bootstrap-t *p*-value	(0.00)	(0.00)	(0.00)
NT	0.27**	0.33**	0.33**
	(0.12)	(0.13)	(0.16)
Wild bootstrap-t *p*-value	(0.04)	(0.05)	(0.05)
GT	0.21*	0.33**	0.44**
	(0.12)	(0.15)	(0.17)
Wild bootstrap-t *p*-value	(0.06)	(0.05)	(0.03)
PT	1.39***	1.37***	1.40***
	(0.11)	(0.12)	(0.15)
Wild bootstrap-t *p*-value	(0.02)	(0.01)	(0.00)
Male		0.10	0.16*
		(0.08)	(0.09)
Wild bootstrap-t *p*-value		(0.55)	(0.27)
Average school GPA		0.11***	0.09**
		(0.04)	(0.04)
Wild bootstrap-t *p*-value		(0.03)	(0.08)
Personal Impulsivity		−0.13*	−0.15**
		(0.07)	(0.07)
Wild bootstrap-t *p*-value		(0.09)	(0.09)
School Impulsivity		−0.11	−0.07
		(0.08)	(0.09)
Wild bootstrap-t *p*-value		(0.19)	(0.24)
Parents hold University degree			0.03
			(0.11)
Wild bootstrap-t *p*-value			(0.32)
Household Commodities			0.05
			(0.11)
Wild bootstrap-t *p*-value			(0.47)
# Observations	1611	1347	1129
Pseudo-R^2	0.16	0.17	0.17

Notes: Maximum likelihood estimation. Cluster-robust standard errors clustered by school in parentheses (first line below the coefficients). The *p*-values for two-sided Wald tests – computed according to Kline and Santos' (2012) score bootstrap method that accounts for small number of clusters (schools) – are also reported in parentheses (second line below the coefficients). ***, **, and * denote significance at $p = 0.01$, 0.05, and 0.10, respectively, based on the more conservative, bootstrapped Wild tests.

parents does not significantly affect their child's grade, and neither does the household's economic level. The estimates for *PT, GT* and *NT* remain statistically significant after controlling for these additional covariates.

We also estimated additional specifications adding interaction terms between the treatment dummies and the proxies for the socioeconomic level of a child's family. The rationale behind the inclusion of these interaction terms is that it may be reasonable to expect the positive effect of the treatments – and of *GT* and *PT* in particular – on subjects' healthy food choices to be especially marked when parents are highly educated and care for their children's nutrition, as well as when children come for wealthier households. This intuition is confirmed by the results in Table E.12 of the Appendix. The probability of a healthy choice continues to be significantly higher in *NT, GT* and *PT* than in the *Baseline* after interacting the treatment indicators with *Parents hold University* degree, and these treatment effects increase with parents' education. The estimate for *NT* becomes statistically indistinguishable from zero once we incorporate interactions between the treatments and *Household Commodities,* but the average effects of *GT* and *PT* on children's probability of choosing healthy food items remain statistically significant; the effect of *PT* also increases with the number of durable goods available in students' homes.

5.4. Longer-run effects – the surprise session four months later

A key issue of any behavioral intervention is the persistence of its effects over time. While some previous work has found strong effects from paying children to choose the healthier option, there is little evidence that the effects persist much after the payments cease. A mechanism that produces enduring effects is more practical than one that does not, particularly if cash payments are not required indefinitely. This section analyzes the effect on children's decisions once any additional information has been removed. As explained in the experimental design, four months after the end of the intervention and within the next academic year, we ran a surprise session. In this case, all subjects participated in a one-shot *Baseline*, so none of the previous incentives that might be triggered

Table 3
Average grades and healthy food choices in surprise session, by initial treatment.

Initial Treatment	Observations	Subjects	Average grade	Proportion of healthy choices
Baseline	33	33	5.04 (1.70)	0.41 (0.25)
NT	68	68	5.60 (1.77)	0.47 (0.26)
GT	65	65	5.94 (1.91)	0.53 (0.27)
PT	45	45	7.43 (1.65)	0.69 (0.21)

Notes: In the surprise session, no additional information was provided. Std. deviations in parentheses.

through different information conditions were in play.

Table 3 presents a summary similar to Table 1, including the average individual grades and the proportion of healthy choices made by individuals in the surprise session.[44] Even after removing additional information and four months after the intervention period, the proportion of healthy choices in each of *NT* (47%), *GT* (53%) and *PT* (69%) are larger than in *Baseline* (41%). A two-sided cluster (school)-adjusted chi-square test of proportions at the individual level shows that the differences are significant for *PT* versus *Baseline* ($p = 0.064$), but not when we compare *GT* or *NT* to *Baseline* ($p = 0.115$ and $p = 0.208$, respectively). Comparing choices in *PT* to *NT* and *GT*, we find significant differences ($p = 0.067$ and 0.043 for two-tailed cluster-adjusted chi-square tests of proportions for the comparisons *PT* vs. *GT* and *PT* vs. *NT,* respectively).[45] These results indicate that providing information to the parents has a strong and persistent effect.

The children made similar choices in the surprise session as they did on average during the intervention. Differences between the average of the first six days and the surprise session are not statistically significant in any treatment, even with two-tailed Wilcoxon signed-rank tests at the individual level ($Z = 1.342, p = 0.180$; $Z = 1.603, p = 0.109$; $Z = 0.534, p = 0.593$; and $Z = 1.342, p = 0.180$, for *Baseline, GT, NT*, and *PT*, respectively).[46] In Appendix H we provide some suggestive evidence on habit formation that might explain this result.

Result 2. ***: Participants' behavior in the surprise session (without any information beyond that in the Baseline condition) four months after the experiment was quite similar to their average behavior in the three-week intervention period when the different information conditions were in place, showing a remarkable degree of behavioral persistence.***

6. Conclusion

Poor diet and obesity have been linked to a variety of contemporary health problems, with concomitant economic consequences. Perhaps this is most important for children, where there are long-lasting health effects. Data from a variety of sources indicate that children still fail to meet recommendations for the daily consumption of fruit and vegetables. Childhood obesity rates in the U.S. tripled from 1971 to 1974 to 2011–2012.

We conduct a field intervention designed to evaluate the effectiveness of different means of influencing children's diets. Previous work has shown positive effects of contemporaneous material benefits for healthy eating (Belot et al., 2016), but at best limited enduring effects. Our design shows that non-material incentives (in the form of information provided to the subjects) can lead children to make healthier food choices at school. We align children's appraisal of food choices with their appraisal of schoolwork by introducing a system in which food items are graded based on their nutritional value. This provides students a yardstick to assess the nutritional value of food and by which to compete over healthy eating.

Critically, we also involve parents as change agents, providing them with information regarding the food choices of their children. While providing information about grades and advice given once by a nutritionist have some value, involving the parents in the decision process generates by far the biggest boost in healthy eating. This provides us with very strong results *that are undiminished four months after our intervention was completed.*

Why might healthy eating persist after withdrawal of the incentives? Charness and Gneezy (2009) find that students induced to exercise through financial incentives continue to exercise after the expiration of these incentives because they learn to enjoy exercise. In our context, students may come to learn that they enjoy consuming healthy foods; some evidence for this comes in the finding that students tend to eat the same foods in the surprise treatment that they had started eating in the main experiment. Here there may be a lesson for future design: if children can get accustomed to eating healthy snacks, then they may come to crave those snacks in the future.

Of course, if parental oversight is so effective, one may wonder why so many parents fail to ensure healthy eating at home or fail to

[44] The lower number of observations in the surprise sessions mainly reflects a school in Baseline (Santo Domingo II) and a school in PT (Gloria Fuertes II) that decided not to participate in the surprise sessions. As we show in Table E.7 of the Appendix, removing these two schools does not lead to systematic co-variance imbalances.

[45] If we ignore school-level clustering, the two-tailed chi-square tests of proportions for the comparisons between *PT vs. GT* and *PT vs. NT* both yield $p = 0.000$ (although recall that our theory model does not make a prediction on the latter)*;* differences are also significant for *GT vs. Baseline* ($p =$ 0.029), but not for *NT vs Baseline* ($p = 0.352$). Table E.8 in the Appendix reports the average grades and the proportion of healthy choices by school and shows that these are higher in each of the schools in *PT* than in any school in the other three treatments.

[46] The conclusions remain the same when we look at the average grades, and when using *t*-tests and permutation tests.

provide their children with healthy lunch packages for school. One possible explanation is that parents in our experiment are involved in an indirect way (they just receive and monitor weekly grades reports) and need not engage in costly daily negotiations with their children over healthy eating. Children may prize the autonomy from maximizing preferences that attend to both taste and health. This shared effort could be key in reducing the burden on parents, hence making the intervention more effective.

In order not to over-generalize these findings, we should point to some specific characteristics of the study that might influence the results. First, our study does not collect prior baseline data on healthy habits or food consumptions in the households. Some of the results may be moderated by household characteristics. For example, it could be that the effect of *PT* may be stronger on children whose parents are more concerned about healthy habits and pay more attention to their children. Although our schools have similar socioeconomic levels, and the explanatory variables collected in our questionnaires do significantly predict children's choices, unmeasured household characteristics could produce heterogenous treatment effects. Further research is needed to properly address this question. Second, because our analysis assumes that nutrition is additively separable across snacks, it does not evaluate the balance in children's diets. For future studies, it would also be interesting to contemplate the dietary composition of students' entire diets when evaluating the effects of different incentive schemes.

To sum up, our approach involves little financial cost, requires only monitoring from parents or peers, and has proved highly effective in both the short and medium run. It is obvious that more research would be helpful since our study is limited mainly to middle-class families in Spain. If policy-makers were able to establish good dietary habits in early childhood, this would make great inroads on the current set of health problems from poor nutritional habits and obesity.

Data availability

Data will be made available on request.

Acknowledgements

We received helpful comments from three referees, Kelly Bedard, Michèle Belot, Francisco Lagos, Sofia Monteiro, Heather Royer, Anya Samek, and seminar participants at the American University of Sharjah and UCSB. Assistance provided by David Fuentes, Alberto Ramos, Nuria Losada and Maria Eugenia Mata is greatly appreciated. Financial support from the Max Planck Society, the American University of Sharjah, the University of Cologne (through the Kelsen Prize) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866 is gratefully acknowledged. This study is registered in the AEA RCT Registry and the unique identifying number is: AEARCTR-0003370.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.euroecorev.2023.104562.

References

Aladino Study (Estudio Aladino), 2019. Spanish Government. https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/consumo/Documents/2020/300920_Nutricion-infantil_estudio_Aladino.pdf.

Allen, J.P., Porter, M., McFarland, F., Marsh, P., McElhaney, K., 2005. The two faces of adolescents' success with peers: adolescent popularity, social adaptation, and deviant behavior. Child Dev. 76, 747–760.

Belot, M., James, J., Nolen, P., 2016. Incentives and children's dietary choices: a field experiment in primary schools. J. Health Econ. 50, 213–229.

Belot, M., Berlin, N., James, J., Skafida, V., 2019. The Formation and Malleability of Dietary Habits: A field Experiment With Low Income Families. Mimeo.

Bollinger, B., Leslie, P., Sorensen, A., 2011. Calorie posting in chain restaurants. Am. Econ. J. Econ. Policy 3, 91–128.

Bulow, J., Geanakoplos, J., Klemperer, P., 1985. Multimarket oligopoly: strategic substitutes and complements. J. Political Econ. 93, 488–511.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-base improvements for inference with clustered standard errors. Rev. Econ. Statistics 90 (3), 414–427.

Cawley, J., Price, J., 2013. A case study of a workplace wellness program that offers financial incentives for weight loss. J. Health Econ. 32 (5), 794–803.

Cecchini, M., Warin, L., 2016. Impact of food labelling systems on food choices and eating behaviours: a systematic review and meta-analysis of randomized studies. Obesity Rev. 17 (3), 201–210.

Charness, G., Gneezy, U., 2009. Incentives to exercise. Econometrica 77 (3), 909–931.

Deci, E.L., Ryan, R.M., 1985. Intrinsic Motivation and Self-Determination in Human Behavior. Plenum, New York.

DeCosta, P., Moller, P., Frost, M.B., Olsen, A., 2017. Changing children's eating behaviour – A review of experimental research. Appetite 113, 328–357.

Donald, S., Lang, K., 2007. Inference with difference-in-differences and other panel data. Rev. Econ. Statistics 89 (2), 221–233.

Downs, J.S., Wisdom, J., Wansink, B., Loewenstein, G., 2013. Supplementing menu labeling with calorie recommendations to test for facilitation effects. Am. J. Public Health 103 (9), 1604–1609.

Dudley, D., Cotton, W., Peralta, R., 2015. Teaching approaches and strategies that promote healthy eating in primary school children: a systematic review and meta-analysis. Int. J. Behav. Nutrition & Phys. Activity 12, 28.

Echenique, F., 2002. Comparative Statics by adaptive dynamics and the correspondence principle. Econometrica 70 (2), 833–844.

Esarey, J., Menger, A., 2019. Practical and effective approaches to dealing with clustered data. Political Sci. Res. Methods 7 (3), 541–599.

Fichera, E., von Hinke, S., 2020. The response to nutritional labels: evidence from a quasi-experiment. J. Health Econ. 72, 102326.

Fisher, J., Birch, L., 1999. Restricting access to palatable foods affects children's behavioral response, food selection, and intake. Am. J. Clin. Nutrition 69 (6), 1264–1272.

Fletcher, J., Frisvold, D., Tefft, N., 2010. The effects of soft drink taxes on child and adolescent consumption and weight outcomes. J. Public Econ. 94, 967–974.

Galloway, A.T., Fiorito, L.M., Francis, L.A., Birch, L.L., 2006. Finish your soup: counterproductive effects of pressuring children to eat on intake and affect. Appetite 46 (3), 318–323.

Global Burden of Disease (2016). The lancet. Available at http://www.thelancet.com/gbd, accessed 02 January 2021.
Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 1 (3), 515–533.
Gerber, A., Karlan, D., Bergan, D, 2009. Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. Am. Econ. J. Appl. Econ. 1 (2), 35–52.
Griffith, R., von Hinke, S., Smith, S., 2018. Getting a healthy start: the effectiveness of targeted benefits for improving dietary choices. J. Health Econ. 58, 176–187.
Haire-Joshu, D., Tabak, R, 2016. Preventing obesity across generations: evidence for early life intervention. Am. Rev. Public Health 37, 253–271.
Hansen, B.B., Bowers, J., 2008. Covariate balance in simple, stratified and cluster comparative studies. Statistical Sci. 23 (2), 919–936.
Ibragimov, R., Muller, U.K., 2010. t-Statistic based correlation and heterogeneity robust inference. J. Bus. Econ. Statistics 28 (4), 453–468.
Jansen, E., Mulkens, S., Jansen, A., 2007. Do not eat the red food!: prohibition of snacks leads to their relatively higher consumption in children. Appetite 49 (3), 572–577.
Jansen, E., Mulkens, S., Emond, Y., Jansen, A., 2008. From the garden of Eden to the land of plenty. Fruit and sweets intake leads to increased fruit and sweets consumption in children. Appetite 51 (3), 570–575.
John, L., Loewenstein, G., Troxel, A., Norton, L., Fassbender, J., Volpp, K., 2011. Financial incentives for extended weight loss: a randomized, controlled trial. J. Gen. Intern. Med. 26 (6), 621–626.
Just, D.R., Price, J., 2013. Using incentives to encourage healthy eating in children. J. Human Resour. 48 (4), 855–872.
Kim, S., Moore, L., Galuska, D., 2014. Vital signs: fruit and vegetable intake among children – United States, 2003–2010. Morbidity and Mortality Weekly Report 63 (31), 671–676.
Kline, P., Santos, A., 2012. A score based approach to wild bootstrap inference. J. Econom. Methods 1, 23–41.
Larson, R., Richards, M., Moneta, G., Holmbeck, G., Duckett, E., 1996. Changes in adolescents' daily interactions with their families from ages 10 to 18: disengagement and transformation. Dev. Psychol. 32, 744–754.
List, J.A., Samek, A.S., 2015. The behavioralist as nutritionist: leveraging behavioral economics to improve child food choice and consumption. J. Health Econ. 39, 135–146.
Loewenstein, G., Price, J., Volpp, K., 2016. Habit formation in children: evidence from incentives for healthy eating. J. Health Econ. 45, 47–54.
Meyer, B., Sullivan, J., 2011. Further results on measuring the well-being of the poor using income and consumption. Canadian J. Econ. 44 (1), 52–87.
Micha, R., Peñalvo, J.L., Cudhea, F., Imamura, F., Rehm, C.D., Mozaffarian, D, 2017. Association between dietary factors and mortality from heart disease, stroke, and Type 2 diabetes in the United States. J. Am. Med. Assoc. 317 (9), 912–924.
Mora, T., Lopez-Valcarcel, B., 2018. Breakfast choice: an experiment combining a nutritional training workshop targeting adolescents and the promotion of unhealthy products. Health Econ. 27 (2), 306–319.
Morizet, D., Depezay, L., Combris, P., Picard, D., Giboreau, A., 2012. Effect of labeling on new vegetable dish acceptance in preadolescent children. Appetite 59, 399–402.
Noble, E., Kanoski, S., 2016. Early life exposure to obesogenic diets and learning and memory dysfunction. Curr. Opin. Behav. Sci. 9, 7–14.
Ogden, J., Cordey, P., Cutler, L., Thomas, H, 2013. Parental restriction and children's diets. The chocolate coin and Easter egg experiments. Appetite 61, 36–44.
Pelchat, M., Pliner, P., 1995. Try it. You'll like it. Effects of information on willingness to try novel foods. Appetite 24 (2), 153–165.
Primo, D.M., Jacobsmeier, M.L., Milyo, J., 2007. Estimating the impact of state policies and institutions with mixed-level data. State Politics and Policy Quarterly 7 (4), 446–459.
Raab, G.M., Butcher, I., 2005. Randomization inference for balanced cluster-randomized trials. Clin. Trials 2 (2), 130–140.
Riener, G., Schneider, S.O., Wagner, V., 2020. Addressing Validity and Generalizability Concerns in Field Experiments. Max Planck Institute for Research on Collective Goods. Discussion Paper 2020/16.
Rigal, N., Rubio, B., Monnery-Patris, S., 2016. Is harsh caregiving effective in toddlers with low inhibitory control? An experimental study in the food domain. Infant Behav. Dev. 43, 5–12.
Roberto, C., Larsen, P., Agnew, H., Baik, J., Brownell, K., 2010. Evaluating the impact of menu labelling on food choices and intake. Am. J. Public Health 100 (2), 312–318.
Rothman, R., Housam, R., Weiss, H., Davis, D., Gregory, R., Gebretsadik, T., Shintani, A., Elasy, T., 2006. Patient understanding of food labels: the role of literacy and numeracy. Am. J. Preventing Med. 31 (5), 391–398.
Sorhaindo, A., Feinstein, L., 2006. What is the relationship between child nutrition and school outcomes? Wider benefits of learning. Research Report No. 18. Institute of Education, University College London.
Thorndike, A., Sonnenberg, L., Riis, J., Barraclough, S., Levy, D, 2012. A 2-phase labeling and choice architecture intervention to improve healthy food and beverage choices. Am J Public Health 102 (3), 527–533.
Thorndike, A., Sonnenberg, L., Riis, J., Barraclough, S., Levy, D, 2014. Traffic-lights labels and choice architecture: promoting healthy food choices. Am. J. Preventing Med. 46 (2), 143–149.
Tsukayama, E., Duckwoth, A., Kim, B., 2013. Domain-specific impulsivity in school-age children. Dev. Sci. 16 (6), 879–893.
Volpp, K.G., John, L.K., Troxel, A.B., Norton, L., Fassbender, J., Loewenstein, G., 2008. Financial incentive-based approaches for weight loss: a randomized trial. J. Am. Med. Assoc. 300 (22), 2631–2637.
Volpp, K.G., Troxel, A.B., Pauly, M.V., Glick, H.A., Puig, A., Asch, D.A., Audrian-McGovern, J., 2009. A randomized controlled trial of financial incentive for smoking cessation. N. Engl. J. Med. 360 (7), 699–709.
Vyth, E., Steenhuis, I., Heymans, M., Roodenburg, A., Brug, J., Seidell, J., 2011. Influence of placement of a nutrition logo on cafeteria menu items on lunchtime food choices at Dutch work sites. J. Am. Diet Assoc. 111, 131–136.
Weinreb, L., Wehlrer, C., Perloff, J., Scott, R., Hosmer, D., Sagor, L., Gundersen, C., 2002. Hunger: its impact on children's health and mental health. Pediatrics 110 (4), e41.
Whitaker, R., Phillips, S., Orzol, S., 2006. Food insecurity and the risks of depression and anxiety in mothers and behavior problems in their preschool-aged children. Pediatrics 118 (3), e859–e868.
Williams, J., Buoncristiano, M., Nardone, P., Rito, A.I., Spinelli, A., Hejgaard, T., Kierkegaard, L., Nurk, E., Kunešová, M., Musić Milanović, S., García-Solano, M., Gutiérrez-González, E., Brinduse, L.A., Cucu, A., Fijałkowska, A., Farrugia Sant'Angelo, V., Abdrakhmanova, S., Pudule, I., Duleva, V., Yardim, N., Gualtieri, A., Heinen, M., Bel-Serrat, S., Usupova, Z., Peterkova, V., Shengelia, L., Hyska, J., Tanrygulyyeva, M., Petrauskiene, A., Rakhmatullaeva, S., Kujundzic, E., Ostojic, S. M., Weghuber, D., Melkumova, M., Spiroski, I., Starc, G., Rutter, H., Rathmes, G., Bunge, A.C., Rakovac, I., Boymatova, K., Weber, M., Breda, J, 2020. A snapshot of European children's eating habits: results from the fourth round of the WHO European childhood obesity surveillance initiative (COSI). Nutrients 12 (8), 2481.
Wisdom, J., Downs, J., Loewenstein, G., 2010. Promoting healthy choices: information versus convenience. Am. Econ. J.: Appl. Econ. 2 (2), 164–178.
World Health Organization, 2009. Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks. World Health Organization, Geneva.