Improving healthy eating in children: Experimental evidence*

GARY CHARNESS,[1] RAMÓN COBO-REYES,[2] GABRIEL KATZ,[3] ÁNGELA SÁNCHEZ,[4] AND MATTHIAS SUTTER[5]

[1] UNIVERSITY OF CALIFORNIA SANTA BARBARA

[2] AMERICAN UNIVERSITY OF SHARJAH

[3] UNIVERSITY OF EXETER

[4] NYU ABU DHABI

[5] MAX PLANCK INSTITUTE BONN, UNIVERSITY OF COLOGNE AND UNIVERSITY OF INNSBRUCK

Abstract

We conduct a field experiment to study whether and how it might be possible to incentivize 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. We also involve parents and classmates as change agents, providing them with information regarding the food choices of their children or friends. We find parents' involvement in the decision process to be crucial in boosting a healthy behavior, with very strong results that survive months after our intervention was completed.

JEL-Codes: C93, I12

Keywords: Healthy eating, children, incentives, field experiment

May 5, 2019

* We received helpful comments from Michelle Belot, Heather Royer, Kelly Bedard, Francisco Lagos, Sofia Monteiro, Any Samek, and seminar participants at the American University of Sharjah and UCSB. Assistance provided by David Fuentes, Nuria Losada and Maria Eugenia Mata is greatly appreciated. Financial support from the Max Planck Society and from the American University of Sharjah is gratefully acknowledged. This study is registered in the AEA RCT Registry and the unique identifying number is: AEARCTR-0003370

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 (GBD, 2016), poor diet is linked to one in five deaths worldwide, with low intake of healthy foods being the leading risk factor for mortality. Moreover, 45.4% of all cardio-metabolic deaths in the U.S. were associated with suboptimal intakes dietary factors (Micha *et al.*, 2017).

A poor diet has been shown not only to have consequences for the adult population but to have long-lasting implications also for children; affecting their physical well-being as well as contributing to the development of dental caries, cognitive impairments and diabetes (Noble and Kanoski, 2016). It may also weaken their immune systems (Sorahindo and Feinstein, 2006), hinder their growth, or undermine their intellectual performance (Weinreb *et al.*, 2002; Whitaker *et al.*, 2006).

Recent data on children's eating patterns are far from encouraging. It has been shown that, despite the 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). A report from the National Cancer Institute (2018) states that 60% of children did not eat enough fruit to meet daily recommendations in the period 2007-2010, and 93% of children did not eat enough vegetables.[1] On a different dimension, the National Health and Nutrition Examination Survey (NHANES) reported in 2011-2012 that over 30% of U.S. children between 2-19 years old were overweight. This percentage was tripled in comparison to the survey carried out in 1971-1974. Since dietary habits are consolidated in childhood, it is imperative that improvements are made during this period (Haire-Joshu and Tabak, 2016).

[1] In Europe only 23.5% of the children eat the recommended daily amount of fruit and vegetables (Lynch et al., 2014)

Hence, designing, implementing and evaluating interventions aiming to 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.

We conduct an innovative field experiment at 12 different schools in Spain to study whether and how it might be possible to influence or incentivize school children to make healthier food choices. In total, we have 300 children, aged nine to ten. In particular, we consider four treatments in our design, with each treatment conducted at three different schools. In all treatments, children were presented with five different food trays with an assortment of choices; they were required to make four choices from whichever trays. In the *Baseline* treatment, nothing else was done. In the *Nutritionist* treatment, a nutritionist gave a talk at the beginning of the first day that explained the benefits of healthy eating. In the *Grades* treatment, there was a clearly-labeled "grade" associated with each of the trays, with healthier foods receiving higher grades. Finally, in the *Parents* treatment, parents received weekly information about the average food grade received by their children; children were aware that parents would receive this information. The intervention lasted for three weeks. After four months, we returned for a surprise visit to examine long-term effects after the removal of any incentives.

Our theoretical model predicts that the *Grades* and *Parents* treatments will generate competition between participants and, based on the positive effect of competition on children choices of fruits and vegetables found in Belot *et al.* (2016), we hypothesize that these two treatments will lead to more healthy choices than the *Baseline*. However, we do not expect the information released by the nutritionist to significantly change the behavior of children, so we should not observe big differences between the *Nutritionist* treatment and the *Baseline*.

Our results show that there is a modest improvement in healthy eating in both the *Grades* and *Nutritionist* treatments, although there are different trends over time. For example, the proportion of healthy choices is 36% in the *Baseline*, 46% in the *Grades* treatment, and 45% in the *Nutritionist* treatment. The increase in the *Parents* treatment is quite dramatic, with 74% (more than double) of the choices being healthy foods. We show that this is not due to outliers. Perhaps most critically, we also show that these effects persist over time even after the removal of incentives. In "surprise" sessions conducted four months later (and without any stimulus), the respective proportions of healthy choices are 41%, 54%, 47%, and 69%, depending on the form of the incentives that had been in force four months before. We observe an average change of a *positive* 2.5 percentage points in the rate of healthy choices made compared to the rate when the intervention was in force. It certainly seems that there are benefits to these interventions, even after removal, and particularly from getting the parents involved in the process.

The contribution of this paper is threefold. First, this paper proposes a sustainable and costless way of improving healthy choices in children. We show that engaging parents as change agents and providing them (with the knowledge of their children) with information about children's food choices is quite effective. In the previous literature, the role of parents has been primarily restricted to two strategies to influence children's food intake: restriction and pressure. Evidence has shown that pressure to eat a target food influences the preference for the food negatively, and that restriction may increase the desire and subsequent intake of the restricted item (DeCosta *et al.*, 2017). In our study, we incorporate parents into the decision process in a non-invasive way. Parents only receive information about the average grade obtained by their children, but they do not have exact information regarding which specific items they chose. So, although parents can still try to influence or stimulate their children's decisions, the final choice is on children, who have the flexibility to choose the food they prefer without any specific restrictions.

Second, and perhaps the most critical contribution of this paper, is that the observed improvements in diet are fully present four months after the intervention and once the incentives have been removed. Our results are even more striking since the surprise visits were conducted not only four months after the first intervention but also within a different academic year. In a sense, our approach of incentivizing children seems to generate some sort of good habits or learning, shedding some light on how to achieve long-term impacts. As far as we know, the only paper that finds a post-intervention effect is that by List and Samek (2015), discussed in more detail in the next section; however, they were only able to test for this one week after the end of the intervention, so that it is unclear whether lasting habits were formed.[2]

Finally, this paper also contributes from a methodological perspective. We provide useful (grading) information to children who are familiar with the grading system used to assess their schoolwork, and who will thus be able to clearly understand what it means for their food choices to be satisfactory or not – i.e., to "fail" or "pass" – given the marks they receive. This feature relates our paper to those studies that focus on choice architecture and nudging. Most of these studies generally use labels that either just try to make the food attractive to subjects (Morizet *et al.*, 2012; Pelchat and Pliner, 1995) or they provide 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

[2] In a different dimension, Charness and Gneezy (2009) in their exercise study find that good habits are sustainable for at least some period after a successful intervention.

data. There is no evidence that any color code interventions have had longer-term effects. Our information system has several advantages over the color code. The grade system provides more flexible and accurate information about the nutritional value of the food while retaining simplicity. This information makes it easier for people to balance consumption when they choose more than just one item. Finally, the proposed system makes the general information about choice and consumption easier to understand and use for other potential agents involved in the process.

We feel that our intervention, particularly the treatment involving parents, offers great promise for successfully battling the problems of poor childhood diet and obesity that has been expanding in the developed world. Changing the eating habits for school children may be the most critical problem to tackle; if sustainable healthy eating habits are established early, we should expect less obesity in the adult population as well. In addition, it may well be the case that parents become inspired by the successful dietary change made by their children and that they themselves adopt a healthier diet. All in all, we feel that this is a very promising avenue.

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

As was discussed briefly in the introduction, this paper relates mainly to two streams of the literature. First, our experiment links to studies that analyze the effect of parental control on food choices of children, although our approach is radically different from what has been done previously. Second, this paper also connects to those that have investigated the effect of providing food information on healthy decisions.

Regarding parental control, there are two main strategies that have been used by parents to influence children's behavior: restriction and pressure. Restriction has been found to have negative consequences in terms of eating behavior; prohibition leads to increased desire and consumption when the forbidden food becomes available (Fisher and Birch, 1999; Jansen *et al*., 2007; Jansen *et al*., 2008). Ogden *et al*. (2013) also found that the restricted group was more preoccupied with the target food. Thus, we are not aware of any studies indicating that restriction is an effective strategy for effecting dietary change.

Galloway *et al.* (2006) tested the effect of pressure to eat on food intake in children. In the study children had to eat soup under two conditions: In the pressure condition, children were reminded to 'Finish your soup, please', four times during the session. In the no-pressure condition, children were not pressured to finish their soup. They find that the increase in intake was higher in the no-pressure condition. Rigal *et al.* (2016) studied the effect of harsh vs. gentle instructions on food acceptance. Children were exposed to baby corn under two conditions: gentle ("You may eat that food. Try to taste it. It's good") or harsh ("You have to eat that food"). Results show that change of intake was higher in the gentle condition.

Our study, although related to the previous literature, offers a completely different approach of parents' involvement in children's decision process. Parents neither can restrict the food options (since choices are made by children in school) nor can they put pressure on children when they are making their decision. In our setting, although parents can still try to influence or stimulate their children's decisions (given that they only receive weekly information about the average grade obtained by their children), the final choice (made in school) is on children, who have the flexibility to choose the food they prefer without any specific restrictions.

Ample attention has been also paid to the role of the provision of 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.

Along the same line, Roberto *et al.* (2010) studied the effect of calorie labels on food choices in a restaurant. They propose three treatments in their experiment: i) a baseline without any information, ii) a calorie label in the menu, and iii) a menu with calorie label and a label stating the recommended daily calorie intake. They find a reduction on calories consumed during the dinner in the restaurant for the calorie and calorie-plus-recommended-intake treatments compared to the baseline. However, calories consumption in the evening after the study dinner was very similar in the baseline treatment and in the calorie plus recommended intake treatment. Furthermore, calorie intake after the study dinner was higher in the calorie treatment than in baseline. So, although there was a short-term effect of the calorie labels on consumption, results show that calorie labels did not work right after they were no longer in place.

There is no robust positive effect of calorie labels on healthy choices (see Downs *et al.*, 2013 and Roberto *et al.*, 2010), which could reflect the fact that nutritional information is not always easy to understand, and many nutrition labels require a high level of literacy and numeracy to interpret (Rothman *et al.*, 2006). To tackle this problem, scholars have proposed an alternative way of providing information about the food that would be easier to understand, showing more promising results. For example, Vyth *et al.* (2011) investigated whether labeling foods with the Choices nutrition logo (a logo introduced in several large catering organizations in the Netherlands in 2006) had an effect on food choices. The experiment was conducted in work site cafeterias in the Netherlands. They find only a significant effect of the logo on fruit sales. However, no significant differences were found in sales for the other items (soup, bread, salad and snacks).

In the same vein, Thorndike *et al.* (2012) introduced a color-code system in a Massachusetts General Hospital cafeteria over 3 months. They used the labels red, yellow and green to label unhealthy, less unhealthy and healthy items, respectively. They find that the sales for red items decreased by 9.2% and for green ones increased by 4.5%. These results come mainly from the effect of color labels on beverages choices: red beverages decreased by 16.5% and green beverages increased by 9.6% during the study. Results are less conclusive, though, when authors perform a difference-in-differences analysis between the intervention cafeteria and two comparison sites. Authors compute the differences in sales between the period of time in which there were no labels in the cafeteria and the period that had the color-label system in place. These differences were computed for both the intervention site and the comparison sites. Results from the between-groups differences are not conclusive: for some items, choices in the intervention site were healthier than in the comparison sites, while for other items they were small and not significantly different and could be even less healthy.

Thorndike *et al.* (2014) study the long-term effect of the color labels system proposed in Thorndike *et al.* (2012). In a second phase of the study, in addition to implementing the traffic light system, Thorndike *et al.* (2012) rearranged the positioning of foods to make healthy items more accessible and unhealthy items harder to get. After 24 months of keeping the labels and the new layout of the cafeteria, red items decreased from 24% to 20% and green items increased from 41% to 46%. So, this paper shows some long-term effect of the incentives on healthy choices, perhaps driven in part by differences in the physical layout (the long-term effect reported was for the labels and "pro-healthy" layout together). Critically, however, they keep the incentive in place for the whole duration of the study (unlike our design), meaning that they cannot speak to the potential persistence of effects after the end of the interventions.

Related to food information, literature focusing on children proposed an alternative type of label, which aims to simply enhance the attractiveness of one particular food. For example, Morizet *et al.* (2012) study the effect of descriptive labels on food choices for children aged 8 to 11. Results show that there is a short-term effect of the labels in the proportion of children choosing the unfamiliar dish. However, this effect was only significant for one of the two items.

Finally, although not directly related to our paper, a now-standard approach in behavioral economics would be to provide material incentives aiming to promote healthy behavior.[3] These incentives have been shown to have an effect in the short-term, but the long-term effect is not yet conclusive. For example, Just and Price (2013) conducted a field experiment at fifteen elementary schools in Utah. They assign students from the different schools to different incentive schemes in which they change the reward for eating a serve of fruit or vegetables per day (a quarter or a prize) and the time the child receives the reward (immediately or in one month). There was a 27.3% 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).

[3] Material incentives have been used to promote healthy behaviors others than the one addressed in this paper. For example, Charness and Gneezy (2009) demonstrate that extrinsic 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. Volpp *et al.* (2009) show that financial incentives significantly decrease the smoking rate, although only during the intervention. Of course, one problem with monetary incentives is that they are difficult to roll out on a large scale.

Belot *et al.* (2016) conducted a field experiment in schools in England to test the effectiveness of two alternative incentive schemes on choosing fruit and vegetables: a piece rate and a tournament. In both schemes, participants got a sticker for choosing a fruit or vegetable. In the piece-rate treatment, subjects who collected four stickers over the week would get a prize. In the tournament treatment, only the child with the largest number of stickers would get the prize. Results show that there was little effect of the piece rate scheme on choices but there was 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) ran a field experiment in Chicago in which children were given a choice between a dried fruit cup (healthy item) and a cookie (unhealthy item), and they were allowed to select only one item. There were four treatments: i) a gain-frame incentive (the child received a small prize for consuming a fruit cup), ii) a loss-frame incentive (the child received a small prize but then it was taken away for not consuming a fruit cup), iii) a 3-min educational message about the benefit of fruits vs. cookies, and iv) a loss-frame incentive combined with the educational message. Paying cash directly for performance was effective: The proportion of children choosing the healthy snack increased from 17% in the baseline to nearly 80% when the incentives were introduced. Perhaps the more interesting question is whether effects continue when the flow of money is cut off. While there is some evidence of post-intervention effects, the only available data are from one week after the removal of the incentives and can therefore only be considered suggestive evidence.

Finally, Belot *et al.* (2019) conducted a field experiment to examine whether dietary habits are malleable. 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. Results provided evidence that children in both treatments reduced their body mass index compared to children in the control group. They also found some evidence that sugar intake was reduced in both treatments. However, Belot *et al*. (2019) did not find strong evidence that children's preferences changed in favor of healthier foods. They argued that one potential explanation for the patterns observed is that the interventions had an impact on what the parents fed their children, rather than on children's preferences.

3. Experimental design and procedures

3.1 Experimental design

Our experimental design consists of four different treatments:

- *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.[4] 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*, with the difference that on the first day of the experiment and before children made their decisions, a nutritionist gave a short talk explaining the benefits of eating in a healthy way. Note that the nutritionist gave a general talk for everyone and did not lead children to pick any particular item from the trays.[5]

[4] See Appendix A for a detailed explanation of how food items were allocated to trays.
[5] A summary of the nutritionist's talk is reported in Appendix B.

- *Grades Treatment* (*GT*, hereafter): Similar to *Baseline*, with the difference that students saw labels with a "grade" associated with each of the five trays before making their choice. The grades corresponding to each tray depended on their nutritional content, and were analogous to those used to mark children's academic schoolwork (in Spanish schools grades range from 0 to 10, so the five trays had assigned marks of 0, 2.5, 5, 7.5, and 10). All the items included in a given tray had the same grade assigned. These grades were the only pieces of information that children in this treatment received. [6]
- *Parents Treatment* (*PT*, hereafter): Similar to *GT*, with the difference that parents received information about the average mark (linked to the nutritional composition) of their child. Before selecting the items, children were aware of the fact that their parents would receive a report about their "performance". Note that parents did not receive exact information about the choices of their children, but rather just the average grade received over the past week.

3.2 Procedures

The experiment was conducted in 12 Spanish schools currently 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 not perceive their decisions as artificial, mitigating potential concerns about the external validity of our findings. Schools participating in the EU program receive, throughout the academic year and depending on the week, different types of fruits and vegetables to be distributed among children. Students just receive the corresponding food, and their only choice is whether to eat it. We felt that children participating in that program

[6] See Appendix C for the detailed composition of the trays and the corresponding grade associated to each one.

would see the experiment as an alternative activity in which the main difference is that they had the chance of choosing the food they prefer to consume.

The selected schools were chosen to guarantee a large and diverse pool of potential subjects, as measured by the demographic and socio-economic characteristics of the students and their family background. Tables E.1 and E.2 in Appendix E present descriptive statistics and tests of differences in means for all the pre-treatment characteristics for which information was obtained from post-experimental questionnaires administered to participants, their parents, and their school teachers (see below). As seen in these tables, there was little variation in the characteristics of the individuals assigned to the different treatments along most of the background covariates. To account for any remaining imbalances, we control for relevant covariates in our econometric (individual-level) analysis.

Each participating school was randomly assigned to one of the treatments described above. This means that all participants from the same school faced the same incentive during the intervention.[7] One class was randomly selected from each school to participate in the experiment. On average, there were 24 children in each class, so we have a total sample of 282 students between the ages of 9 and 10.

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 other classmates who had already completed the task. This procedure

[7] We are not very concerned about contamination across schools, since children typically went directly home from school and in any event were quite young to have a social life with children from other schools.

ensures that the decisions of subsequent classmates (who remained either in Room A or C at the time the participant was choosing her lunch in Room B) were not directly influenced by the decisions made by previous individuals.

A member of the research team, who was present during the decision process, recorded the students' dietary choices and grades. Note that in order to minimize interaction with the other participants, children made their decisions alone. The bags used to keep the food were transparent so the researcher could see what children picked, without being directly involved during the decision.

The intervention was always conducted right before the main school break.[8] To make the food decision salient, in every treatment the parents were instructed not to prepare any snack for their children for the days of the intervention.[9] In this way, children chose the food they would eat during the break that day.

A staff member of each school was present in Room C where children would gather together after making their choices. This person made sure that no food was wasted (although children could keep the food for later consumption). As a result, most of the items chosen in Room B were consumed in Room C.

To analyze the dynamics and long-term effects of the incentives, we collected data twice a week over three weeks, for a total of six observations per subject.[10] Importantly, to keep the same conditions across treatments, *the consent letter was the same in each treatment* and the parents of all the participants received exactly the same information about the experiment and were asked to provide their contact details before the experiment began. Note that parents in *PT* only received information about their child's average grade. They did

[8] This break is at 11am. Typically kids go outside during this break, play some games and have a snack to help them keep going until the lunch break that – in Spain – only starts at 2:30-3:00pm.
[9] While we could not enforce that children didn't bring any food with them to school, we did insure 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.
[10] The procedures to collect the data in subsequent days were the same as in the first day, i.e., participants would be located in different rooms and decisions would be made privately.

not receive any other information about other children's performance. The information was released once a week and included the average grade of the particular child for each of the two sessions ran during that week.

After the sixth day of the experiment children were administered a questionnaire asking them to identify their closest friends in class. It is conceivable that participants would talk to each other after the experimental session and discuss their choices. Hence, information about classmates' decisions might affect a child's choices over the course of the intervention. Consequently, we tried to elicit each subject's social network in order to account for these potential interdependencies.

We also conducted post-experiment questionnaires with children, teachers and parents, in order to obtain information about participants' socioeconomic variables and self-control indicators. Regarding the socioeconomic background of the student, we utilized a questionnaire administered by the Spanish Government. 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.[11] For the self-control indicators, following Tsukayama *et al.* (2013), we included questions related to interpersonal impulsivity and also to schoolwork impulsivity.[12] Finally, in the questionnaire administered to the teachers, we gathered information regarding students' average performance in class, average attendance to the school, and a measure of self-control.

Finally, four months after the end of the intervention and during the next academic year, all subjects participated in a surprise session, which was conducted as a one-shot *Baseline* for everyone and which was unannounced. The aim of this was to study whether the

[11] See Appendix D.1 for the complete survey.
[12] See Appendix D.2 for the complete questionnaire.

effect of the different incentives remained after some time and once incentives had been removed.

4. Theoretical predictions and hypotheses

In this section, we offer a simple model that captures the main features of our experimental design and that allow us to state our main hypotheses.

Consider i's utility function to be:

$$u_i(x_i) = -(x_{it} - x_{pref})^2 - \delta_{it}\alpha_i(x_{it} - x_{norm})^2 + \gamma_{it}\beta_i\left(g_{it} - \bar{g}_{j,t-1}\right)$$

where x_{it} represents the amount of healthy food consumed by subject *i* in period *t* and $x_{it} - x_{pref}$ captures the difference between *i's* healthy food consumption in period *t* and *i*'s preferred healthy food consumption.

Basically, what the utility function proposes is a satiation point. Subject *i* wants to consume exactly his or her preferred amount of healthy food. Deviating from that preferred level of healthy consumption reduces i's utility. Consider, for example, $x_{pref} = 0$, so that *i*'s preferences are such that subject *i* wants to consume only junk food. So, the more healthy food subject *i* consumes, the less junk food she will consume, and the lower her utility will be.

In what follows below, we define the following:

- $x_{it} - x_{norm}$ captures the difference between *i*'s consumption in period *t* and *i*'s expectation of the social norm regarding healthy consumption. Note that here we assume that subjects experience some disutility from deviating from what they consider to be socially "good",

- δ_{it} represents the erosion factor of α_i over time. This factor considers that subject i will care more about the social norm at the beginning but over time he or she starts potentially caring less about behaving as the society dictates,[13]
- $g_{it} - \bar{g}_{jt-1}$ represents the difference between i's average grades and the observed average grade of other people in i's group in the previous period. This difference captures a sort of competitive preferences in which subject i will increase her utility if she is getting a higher grade than her colleagues in class. As explained in the experimental design, grades are a function of consumption and they will play a role only in *GT* and *PT,*
- γ_{it} represents the erosion factor of β_i over time.

To compute the optimal consumption in *Baseline*, given that grades are not announced and the optimal consumption of healthy items is not explicitly defined, we assume that $\beta_i = 0$. Hence, the optimal consumption of healthy items in *Baseline* would be:

$$x_{it}^{*} = \frac{2x_{pref} + 2\delta_{it}\alpha_i x_{norm}}{2 + 2\delta_{it}\alpha_i}$$

NT differs from the *Baseline* in that the socially-accepted amount of healthy food is more accurately defined. In this case, it will be the nutritionist who states what is the desirable consumption of healthy items. However, we do not expect this accurate information to be very different from what people may presume, so we should not observe big differences in choices between the two treatments.

Hypothesis 1: *There will be a similar proportion of healthy items picked in NT and in Baseline.*

[13] It is also possible that over time a child learns better about the social norm, so the norm itself would change.

In *GT*, as noted in the experimental design, food is graded according to its healthiness. Given that grades are directly correlated with food choices, we replace g_{it} with x_{it} and $\bar{g}_{j,t-1}$ with $\bar{x}_{j,t-1}$ – the average proportion of healthy choices by people other than i in period t-1. In this case we allow for $\beta_i \geq 0$. The optimal consumption in *GT* would be

$$x_{it} = \frac{2x_{pref} + 2\delta_{it}\alpha_i x_{norm} + \gamma_{it}\beta_i}{2 + 2\delta_{it}\alpha_i}$$

Following Belot *et al.* (2016), who find a positive effect of a competition mechanism on children choices of fruits and vegetables, we assume that $\beta_i > 0$ even though our environment does not incentivize competition directly. A strictly positive β will lead to a higher proportion of healthy choices in *GT* compared to *Baseline*.

Hypothesis 2: *There will be a larger proportion of healthy food chosen in GT than in Baseline.*

Regarding subjects' behavior in *PT*, we conjecture that the presence of parents will make grades even more important since it allows for the possibility of getting social recognition not only at school but also at home. Experimental literature has found status seeking to be strongly related to competition among individuals (Huberman *et al*., 2004; Rustichini and Vostroknutov, 2008; Charness *et al*., 2014). Based on these findings, we assume that having the grades released to the parents would make β_i larger in *PT* than in *GT*, leading to a higher proportion of healthy items chosen in *PT*.

Hypothesis 3: *There will be a larger proportion of healthy food chosen in PT than in GT.*

Regarding the dynamics of the experiment, we assume that the erosion factor in our model will play a role, reducing the effect of incentives over time while the incentives are still in place, i.e., over the six experimental days across the three weeks of the intervention. We base our assumption on the lack of a long-term effect found in Belot *et al.* (2016). This will lead to our fourth hypothesis.

Hypothesis 4: *The proportion of healthy food chosen in all treatments will decrease over time during the intervention period, while incentives are still in place.*

5. Results

This section is structured as follows. We begin by comparing the average behavior of children in each treatment. We then examine the dynamics and evolution of decisions over time during the intervention period and persistence effects in the longer run. Finally, we explain the main determinants of subjects' choices.

4. 1. Children's behavior

We start with an overview of the decisions made by the participants in each treatment. Table 1 presents a summary of the average individual grade, and the proportion of healthy choices made by individuals. For the second analysis, we define as healthy choices those items that were assigned a grade of 7.5 or above.[14]

Table 1. Average behavior

Treatment	Observations	Subjects	Average grade	Proportion of healthy choices
Baseline	400	73	4.77	0.36
NT	412	71	5.50	0.45
GT	373	65	5.59	0.46
PT	432	73	7.88	0.74

[14] 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. Results remain qualitatively similar in both cases (see Tables E.3 and E.4 in Appendix E).

As seen in Table 1, the average grades in *GT* (5.59) and in *NT* (5.50) are both larger than in *Baseline* (4.77). Furthermore, the grades in every session of *GT* and *NT* were higher than the grades in any session of the *Baseline*. If we pool the six observations for *GT* and *NT*, a Wilcoxon ranksum test using session-level data finds a significant difference at $p = 0.024$, two-sided test.[15,16] This is the highest significance possible with this number of observations.[17]

So, it seems that simply providing information to children, either by having grades assigned to the food or by having a nutritionist explaining children the benefits of eating in a salubrious manner, does help to improve healthy food choices in this part of our study. Moreover, the two ways of providing information seem to have a very similar effect on subjects' decisions.[18] These results support Hypothesis 2 but not Hypothesis 1.

But differences in grades are dramatically larger when information is released to the parents. Results from Table 1 show that the average grade in *PT* (7.88) is significantly larger than in all the other treatments. Once again, the Wilcoxon ranksum test on school-level data gives the highest possible significance level for all pairwise comparisons with the *PT* treatment. Tests using individual-level data give very strong results for the comparison between *PT* and each of the other three treatments: $Z = -18.630$, $p = 0.000$ versus *Baseline;* $Z = -15.682$, $p = 0.000$ versus *GT;* and $Z = -16.228$, $p = 0.000$, versus *NT.* These results support the idea that involving parents in the process by making them aware of their children's decisions is a very strong mechanism to spur a healthy behavior. Hence, we find strong support for Hypothesis 3.

---

[15] The lowest possible *p*-value for a one-sided test (justified by Hypothesis 2) with only three observations in each treatment is $p = 0.050$ and that is what we obtain in a pairwise test of *GT* versus *Baseline*. Since we have no directional hypothesis for *NT* versus *Baseline*, we must use a two-sided test and get the maximum possible significance level of $p = 0.100$

[16] We round all *p*-values to the nearest third decimal place.

[17] If we assume that there is no interaction among the students, we can use individual-level data, which gives significance at $p = 0.000$.

[18] Differences are not significant when we compare average grades in *GT* and *NT* ($Z = 0.218$, $p = 0.414$, one-tailed Mann-Whitney test).

Turning to the proportion of healthy choices, results are very similar to those using the average grades. As shown in Table 1, the percentage of healthy items chosen in the *Baseline* is 36%. This percentage increases in *NT* (45%) and *GT* (46%) and more than doubles with *PT* (74%) compared to the *Baseline* rate.[19] These findings support the previous results, showing that there is a moderate effect on choices in *NT,* a positive although non-significant effect in *GT*, and that children radically change their behavior when parents are informed in *PT*.

> **Result 1:** *Providing information to children regarding the value of the food has a moderate, but nevertheless significant effect on choices. Releasing the information about the child's average grade to her parents has a very strong effect on subjects' decisions, largely increasing the consumption of healthy food items*.

We now investigate a potential explanation for the reported positive effect that releasing the grade information to the parents has on children's choices. We argue that the average increase in healthy choices is generated by the effect that the incentive has on the majority of the population, however we could also observe the same results if there were a stronger effect affecting only a fraction of the subjects. To study to what extent the majority of subjects is affected by the incentive, Figure 1 plots the proportion of subjects who fall into each of four choice categories. We compute the average grade of each individual and then allocate this to one of the following categories, with the average grade: *i)* at or below 2.5, *ii)* between 2.51 and 5, *iii)* between 5.01 and 7.49, and *iv)* at or above 7.5.

[19] Once again, conservative Wilcoxon ranksum tests on school-level data give the maximum significance when we compare *Baseline* and *NT* (p = 0.10, two-sided test) and *Baseline* and *PT* (p = 0.05, one-sided test). However, the difference is not quite significant for the comparison *Baseline* and *GT* (Z =-1.527, p = 0.063, one-tailed test). All differences using the *Baseline* as a reference level are significant with individual-level data: Z =-2.471, p = 0.007 versus *GT*; Z = -4.621, p = 0.000 versus *NT*; and Z = -15.665, p = 0.000 versus *PT,* all one-tailed tests. Again, these findings remain similar using alternative definitions of "healthy choice" (see Tables E.3 and E.4 in the Appendix).

Figure 1. Proportion of subjects in each choice category

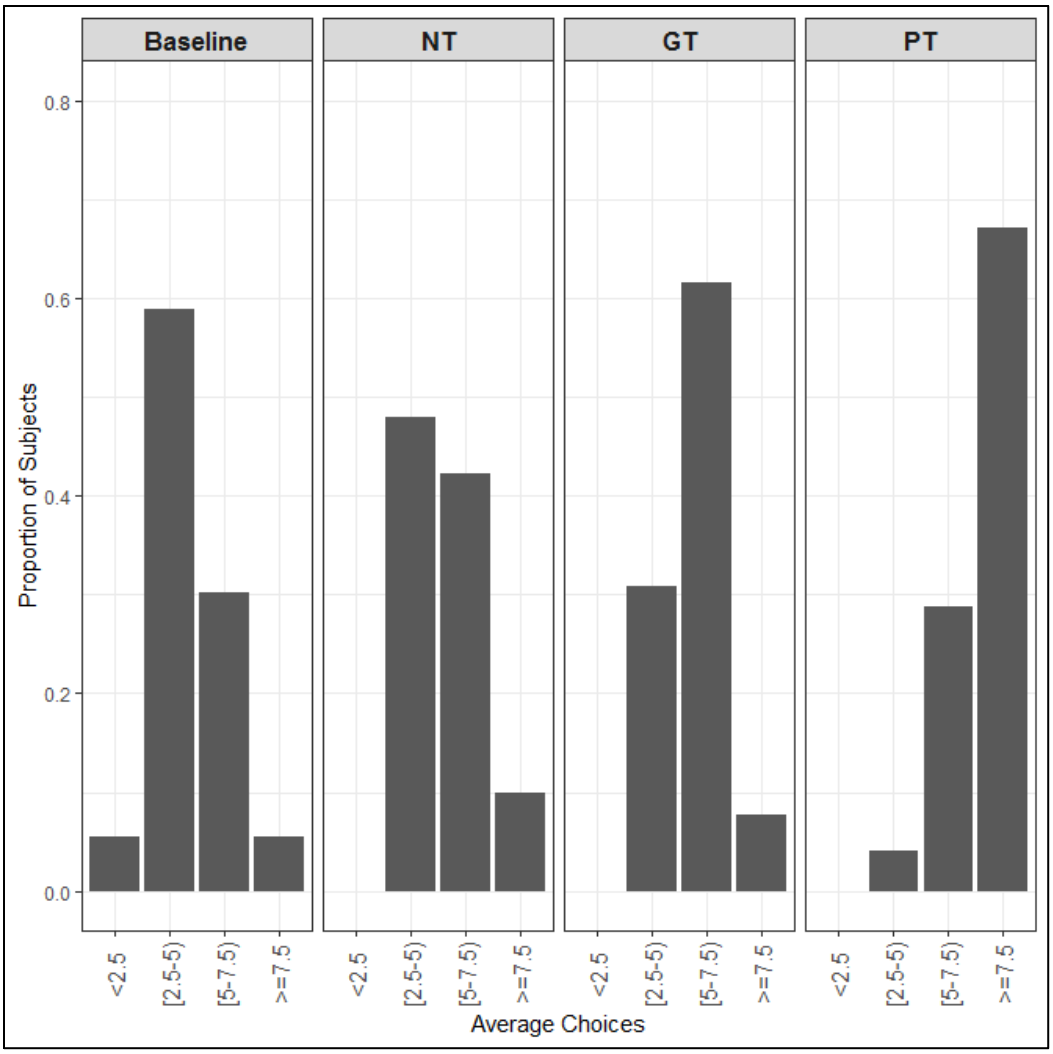


Figure 1 reveals that the distribution of subjects substantially changes in *PT* compared to all the other incentive schemes. Almost 70% of the population in *PT* had an average grade above 7.5. This percentage is significantly larger than in *Baseline* (5.5%), *GT* (12.20%) and *NT* (9.85%).[20] Moreover, the distributions in *GT* and *NT* are very similar (two-sample Kolmogorov-Smirnov test: D = 0.21, p = 0.190, two-tailed test). These results suggest that the improvement in healthy choices found in *PT* comes mainly from the effect that the

[20] All differences are statistically significant at the 5% level (one-tailed test) with conservative ranksum tests on school-level data using *PT* as the reference group. These differences with respect to *PT* are more significant using individual-level data: Z = -7.718, p = 0.00 versus *Baseline;* Z = -7.115, p = 0.000 versus *GT*; Z = -7.023, p = 0.000 versus *NT*; all one-tailed tests.

incentive had on the majority of participants rather than from a very strong effect on just a subset of the population.

> **Result 2:** *A majority of participants react to the incentives in PT, leading to the large improvement on healthy choices*.

4.2. Dynamics during the intervention period

This section focuses on the dynamics in the four treatments. Figure 2 represents the average percentage of healthy choices over time, i.e., over the six experimental days across the three weeks of the intervention. Note that, similar to Table 1, Figure 2 considers as healthy choices items that were assigned a grade of 7.5 or 10.[21]

We observe two clear patterns in our data. While Figure 2 shows a positive trend in *GT* and *PT*, this trend is clearly negative in *Baseline* and *NT*.[22] So, the percentage of healthy choices increases over time in *GT* and *PT* and it drops in *Baseline* and *NT*. This result only partially supports Hypothesis 4.

One intuition for this result is that the grades introduce some sort of competition among children. Note that even if the decisions were made individually in an isolated room, children were gathered together after choosing the food. This would mean that they had information about others' choices, which would allow them to easily compute others' grades. This could influence subjects' decisions in subsequent periods, which would lead to an improvement in choices over time. Competition has been shown to have strong effects on behavior.[23]

---

[21] Figure E.1 in the Appendix provides the same analysis considering the average grade and the percentage of people with an average grade >7.5 as variables.

[22] All trends are statistically significant ($Z = 2.68$, $p = 0.007$; $Z = -2.72$, $p = 0.006$; $Z = 2.68$, $p = 0.007$; and $Z = 4.99$, $p < 0.001$; Cochran-Armitage test for *Baseline, GT, NT*, and *PT*, respectively).

[23] Teachers informed us that kids in class were comparing their grades and talking about how to improve the next day, so we do think that competition played a role here, as it appears to have in Belot *et al.* (2016). Examples of the effect of competition include Bornstein, Gneezy, and Nagel (2002), who find that competition between groups increased contributions in the minimum-effort game, Fershtman, Gneezy, and List (2012), who find that just having the notion of competition in the air led to social preferences vanishing, and Charness and Holder (2019), who show that competition between groups to have their charitable contributions matched led to

Figure 2. Percentage of healthy choices over time

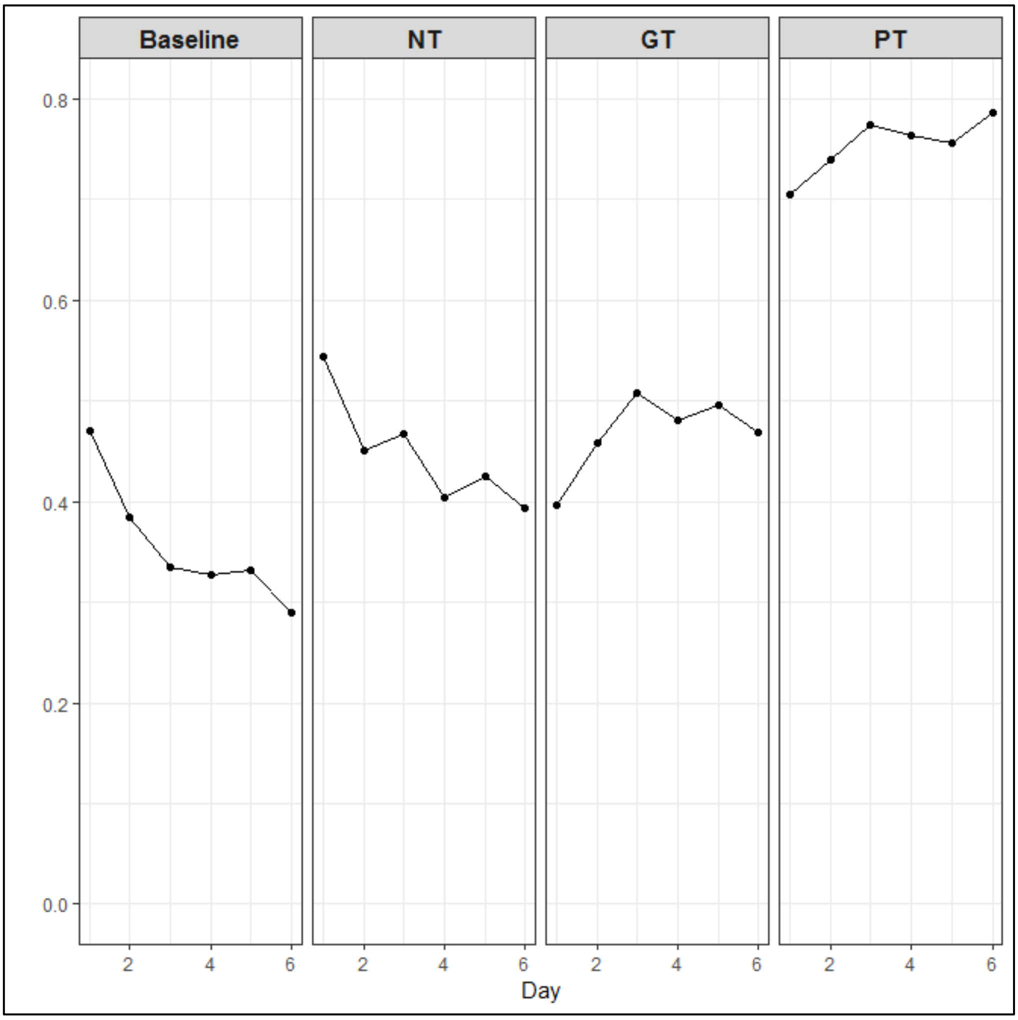


Figure 2 also reveals that in *PT* the percentage of healthy choices keeps increasing, reaching almost 80% of 'good' items chosen on the last day of the three-week intervention. So, it seems that, on top of the potential competition created by the grades, keeping parents informed about their children's 'performance' further affected participants' behavior. However, in *Baseline* and *NT*, subjects do not have any reference point for competition. This fact combined with the lack of incentives for choosing the healthy option leads to a reduction in the proportion of healthy items chosen over time.

a substantial increase in charitable contributions. Majolo and Marechal (2017) find greater within-group cooperation when groups of children were competing with other groups.

To study the extent to which children react by picking healthier items solely due to the knowledge that their average grade will be released to their parents, we look at students' decisions in the first period in each treatment. Because all parents received exactly the same information regarding the experiment in all treatments, children's behavior in period 1 in *PT* is not affected by their parents' direct intervention but rather only by the expectation of their information to be released to their parents in the future.

As Figure 2 shows, the percentage of healthy choices in the first period is 70.52% in *PT*, which is larger than in *Baseline* (46.68%), *GT* (39.68%), and *NT* (54.37%).[24] These results indicate that simply knowing that their parents will view this information has a large effect on the food choices made by the students.

4.3. Longer-term effects after the intervention period and the removal of incentives

A really critical issue is whether the effects of the intervention persist over time. Some previous work (e.g., List and Samek, 2015) has found that there are strong effects from paying children to choose the healthier option, but no study has found effects that persist much after the payments cease. A mechanism that produces enduring effects is eminently more practical than one that does not. This section analyzes the effect on children's decisions several months after the incentives are 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 the previous incentives were not in play.[25]

[24] Differences are statistically significant for the comparison *PT* versus *Baseline* ($Z = 4.584$, $p = 0.000$, one-tailed ranksum test), for the comparison *PT* versus *GT* ($Z = 6.061$, p = 0.000, one-tailed ranksum test), and for the comparison *PT* versus *NT* ($Z = 3.562$, p = 0.000, one-tailed ranksum test). Note that, given that in the first period all subjects are independent, all the tests here are done at the individual level, rather than at the school level; tests with school-level data indicate significance at the 5%-level, as before.

[25] We did not inform the parents or the children about the surprise session in advance and so did not warn parents not to pack snacks. However, we followed the same procedure as before and children did not have access to their backpacks until the break was over, so we could make sure that the snacks they ate during the break were the ones they chose in the experiment.

Table 2 presents a summary similar to the one reported in Table 1, which includes the average individual's grade, and the proportion of healthy choices made by individuals in the "surprise session".[26]

Table 2. Average behavior without incentives in the surprise session (4 months later)

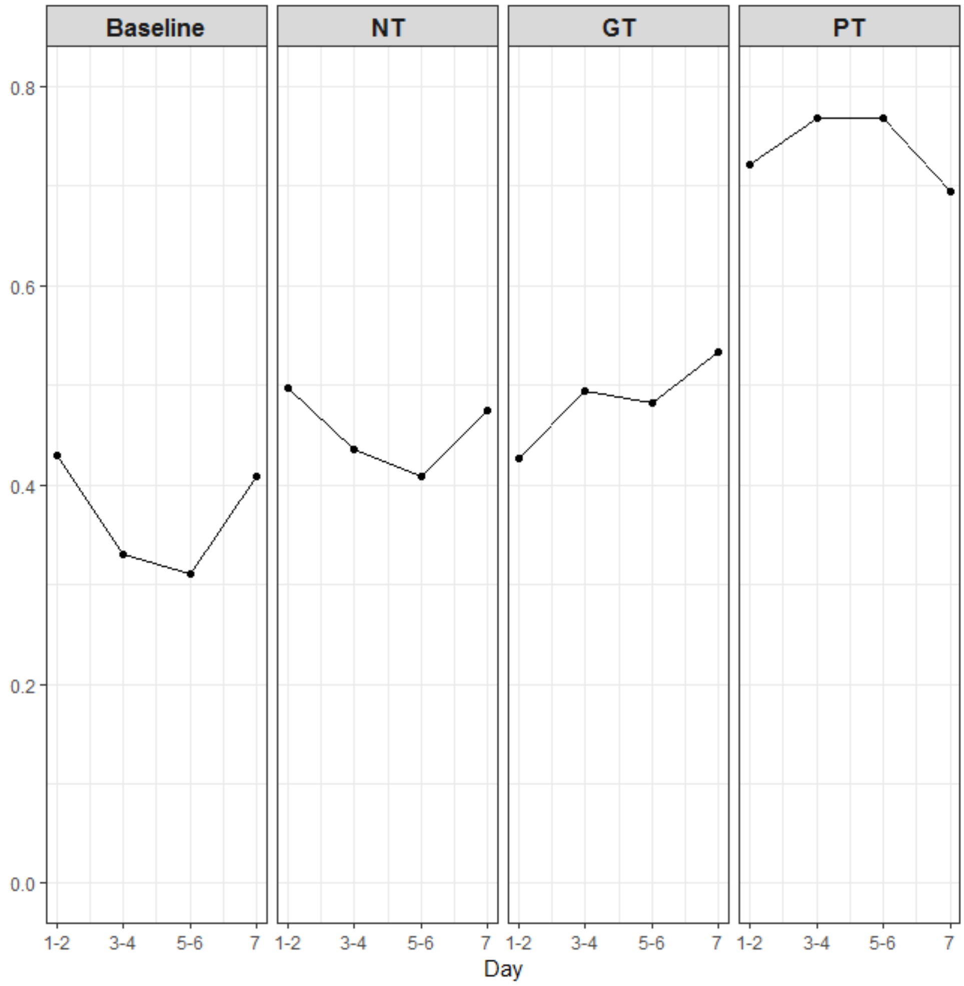
Initial Treatment	Observations	Subjects	Average grade	% healthy choices
Baseline	33	33	5.04	0.41
NT	68	68	5.64	0.47
GT	65	65	5.93	0.53
PT	45	45	7.42	0.69

As shown in Table 2, even after removing the incentives and four months after the intervention period, the proportion of healthy choices in each of *GT* (53%), *NT* (47%) and *PT* (69%) are larger than in *Baseline* (41%). A test of proportions at the individual (child's choice) level shows that while differences are significant for the comparison *GT vs. Baseline*, and *PT vs. Baseline* ($p = 0.025$, and $p = 0.000$, respectively); they are not significant when we compare *NT* and *Baseline* ($p = 0.259$). Comparing choices in *PT* to *NT* and *GT*, we also find differences to be statistically significant ($p = 0.000$ for one-tailed test of proportions for the comparisons *PT vs. GT* and *PT vs. NT*).[27] These results lead to two main conclusions. First, there is a clear residual positive effect of *GT* on healthy choices, but the effect of the incentive weakens considerably in *NT*. Second, releasing information to the parents in *PT* has a stronger effect and, even with the incentive no longer being present, leads to considerably healthier choices four months after the initial intervention.

[26] The lower number of observations in the surprise sessions is mainly due to the fact that one school in *Baseline* and one school in *PT* decided not to participate in the surprise sessions. The rest of the missing data reflects children who were not present the day of the surprise session.

[27] The results are similar if we take the average choices at the school level and conduct one-tailed Wilcoxon Mann-Whitney tests (with *p*-values ranging from 0.04 to 0.06).

Figure 3. Percentage of healthy choices in the first stage (days 1-6) and on day 7 (surprise session 4 months later)



We next focus on the comparison of subjects' decisions in the "surprise session" to their average behavior in the first part (first six days) of the experiment. Figure 3 compares the average percentage of healthy choices in the first stage and the "surprise" session. We see that children did not behave differently on day 7 (the surprise session) compared to their average behavior during the first stage (i.e., days 1-6). In *Baseline*, the percentage of healthy choices improves from 36% in the first stage of the experiment to 41% in the "surprise" session; in GT this improves from 46% to 53%; in NT it improves from 45% to 47%; and, finally, in PT the percentage drops from 74% and 69%. Differences between the average of the first six days and the "surprise" session are not statistically significant in any treatment even with one-tailed Wilcoxon signed-rank tests (Z = -1.342, p = 0.179; Z = -1.342, p =

0.179; Z = -1.069, p = 0.285; and Z = 1.342, p = 0.179, for *Baseline, GT, NT*, and *PT*, respectively).[28] So, it appears there is a carry-over effect for all treatments, which would lead to healthier choices in the cases in which the incentives were effective in the first place.

> ***Result 3:*** *Participants' behavior in the "surprise session" of the experiment was quite similar to their average behavior in the three weeks intervention period when the different incentives were in place, showing a remarkable degree of persistence. This indicates significantly healthier behavior even four months after the incentives have been removed.*

4.4. Individual behavior

To better understand the mechanisms underlying the group-level patterns reported above, this section presents individual-level analyses of choices. The first column of Table 3 reports the marginal effects from a panel Probit in which the dependent variable is the probability that subject i chooses a healthy food item (grade 7.5 or above) in period t. As explanatory variables, we use dummies for each treatment (*NT*, *GT*, and *PT*, taking *Baseline* as the benchmark), a gender dummy (for being male), 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 covariates. Given that childhood self-control can predict (among other outcomes) physical health, and substance dependence (Moffitt et al, 2011), we considered that this variable – or its obverse, impulsivity – would be an interesting control in a different dimension of health-related choices. Moreover, we include information about the minimum grade of food choices of a subject i's network members in period t-1 (*Minimum Grade*$_{-i,t-1}$) in order to account for any peer effects. Finally, we control for the socio-economic level of a child's family by taking into account parents' education – whether or not at least one parent holds a graduate or postgraduate degree – and a family's consumption level. The latter is

[28] Conclusions remain the same when we look at the average grades. Average grades are 5.04 versus 4.77 in Baseline, 6.18 versus 5.59 in GT, 5.64 versus 5.50 in NT, and 7.44 versus 7.88 in PT. Differences are not statistically significant in any case.

measured as a composite index of the number of durable goods (laptop and desktop computer, tablet, electronic books, cars) available in each student's home.

Our models also include subject and period random effects to account for time-invariant individual heterogeneity and common temporal shocks affecting all subjects, along with session and group random intercepts to control for potential session effects (Fréchette, 2012) and contemporaneous correlation between same-group members (Poen, 2009).

The probability of choosing a healthy item increases in all three treatments compared to *Baseline*. This result suggests that participants generally responded to incentives, and reinforces the group-level evidence reported in Table 1. Moreover, and also in line with the group-level results, we observe that the probability of a healthy choice is significantly larger in *PT* than in *GT* (F = 68.862, $p = 0.000$; test for equality of coefficients) and *NT* (F = 164.90, $p = 0.000$). We also see that the effect of grades is stronger than the effect of the nutritionist in boosting healthy decisions, controlling for other variables (F = 12.953, $p = 0.000$; test for equality of coefficients between *NT* and *GT*). Also, individuals' indexes of personal and schoolwork impulsivity negatively affect healthy decisions. So, children with more self-control make healthier choices.[29]

Finally, results also indicate a peer effect on children's decisions. The minimum grade obtained by any member belonging to the social network of the child in the previous period positively affects the probability of making a healthy choice. The intuition behind this result could be that when the child observes that somebody in her social network chose some "bad" items, this gives that child the justification to also choose an unhealthy item in the next period.[30]

[29] Replacing self-reported impulsivity with an index reported by teachers at schools leads to similar results.
[30] We also run regressions that consider the average grade of the friends, instead of the minimum, as a covariate. We find this variable to also be significant and positive. The variance of a child's grades, on the other hand, is negative but not significant.

Table 3. Probit regression on the probability of choosing healthy food

	Data from intervention period			Four months after intervention
	(1)	(2)	(3)	(4)
Constant	5.12*** (0.36)	5.01*** (0.41)	5.39*** (0.57)	6.66 *** (0.93)
GT (dummy)	1.22*** (0.18)	1.22*** (0.23)	1.51** (0.64)	1.37*** (0.51)
NT (dummy)	0.62*** (0.16)	0.79*** (0.22)	-0.10 (0.61)	0.71 (0.48)
PT (dummy)	2.79*** (0.21)	2.17*** (0.28)	1.86*** (0.63)	2.55*** (0.51)
Male	0.13 (0.11)	0.16 (0.12)	0.14 (0.11)	-0.13 (0.62)
Personal Impulsivity	-0.24*** (0.09)	-0.27*** (0.09)	-0.23*** (0.09)	-0.49* (0.27)
School Impulsivity	-0.22** (0.10)	-0.16 (0.12)	-0.21** (0.11)	-0.16 (0.23)
Minimum Grade$_{-i,t-1}$	0.17*** (0.04)	0.20*** (0.04)	0.16*** (0.040	
Parents hold University degree		0.38 (0.32)		
Consumption			-0.08 (0.14)	
GT x Parents hold University degree		0.60 (0.49)		
NT x Parents hold University degree		-0.23 (0.38)		
PT x Parents hold University degree		1.00** (0.39)		
GT x Consumption			-0.07 (0.18)	
NT x Consumption			0.21 (0.18)	
PT x Consumption			0.24 (0.16)	
# Observations	1,122	943	1,122	170
Log Likelihood	-2,265.85	-1,879.96	-2,262.56	-331.88

Notes: Standard errors in parentheses. ***, **, and * denote significance at p = 0.01, 0.05, and 0.10, respectively. *Parents hold University degree* in column (2) is binary variable that takes value 1 if at least one of the parents achieved graduate or postgraduate education and, 0 otherwise. *Consumption* in column (3) represents the consumption level of the household as a proxy of economic level of the family; it is a composite index of the number of durable goods (laptop and desktop computer, tablet, electronic books, cars) available in each student's home.

Columns (2), and (3) repeat the specification in column (1), but take into account the socioeconomic level of a child's family. Column (2) focuses on the parents' educational level, adding to the set of explanatory variables a binary variable that takes value 1 if at least one of the parents achieved graduate or postgraduate education and, 0 otherwise. We also add the interaction between *Parents hold University degree* and the three treatment dummies, allowing us to assess whether the educational level of the parents affected children's decisions differently depending on the incentives scheme. We observe that a higher educational level for parents does not significantly affect the probability of the children choosing healthy in *Baseline*. However, this educational level of the parents increases the probability of a healthy choice in *PT*.

Column (3) uses the consumption level of the household as a proxy of economic level of the family, and its interaction with the treatment dummies instead of the parents' educational level. We find no effect of the economic level of the household on healthy choices in any treatment.[31]

Finally, Column (4) repeats the specification in column (1) for the data in the surprise session. Note that all subjects participated in the *Baseline* in the surprise session, so the treatment dummies in column (4) capture the persistence from the original treatments. We report that the positive effect found for *GT* and *PT* still remains in the long run.

In order to check the robustness over time of the reported treatment effect, Figure 4 plots the differences in the probability of choosing healthy foods between the *Baseline* treatment and *GT*, *NT* and *PT* for each period (i.e., the daily treatment effects).[32] Results show that differences between *PT* and the *Baseline* are positive and statistically significant

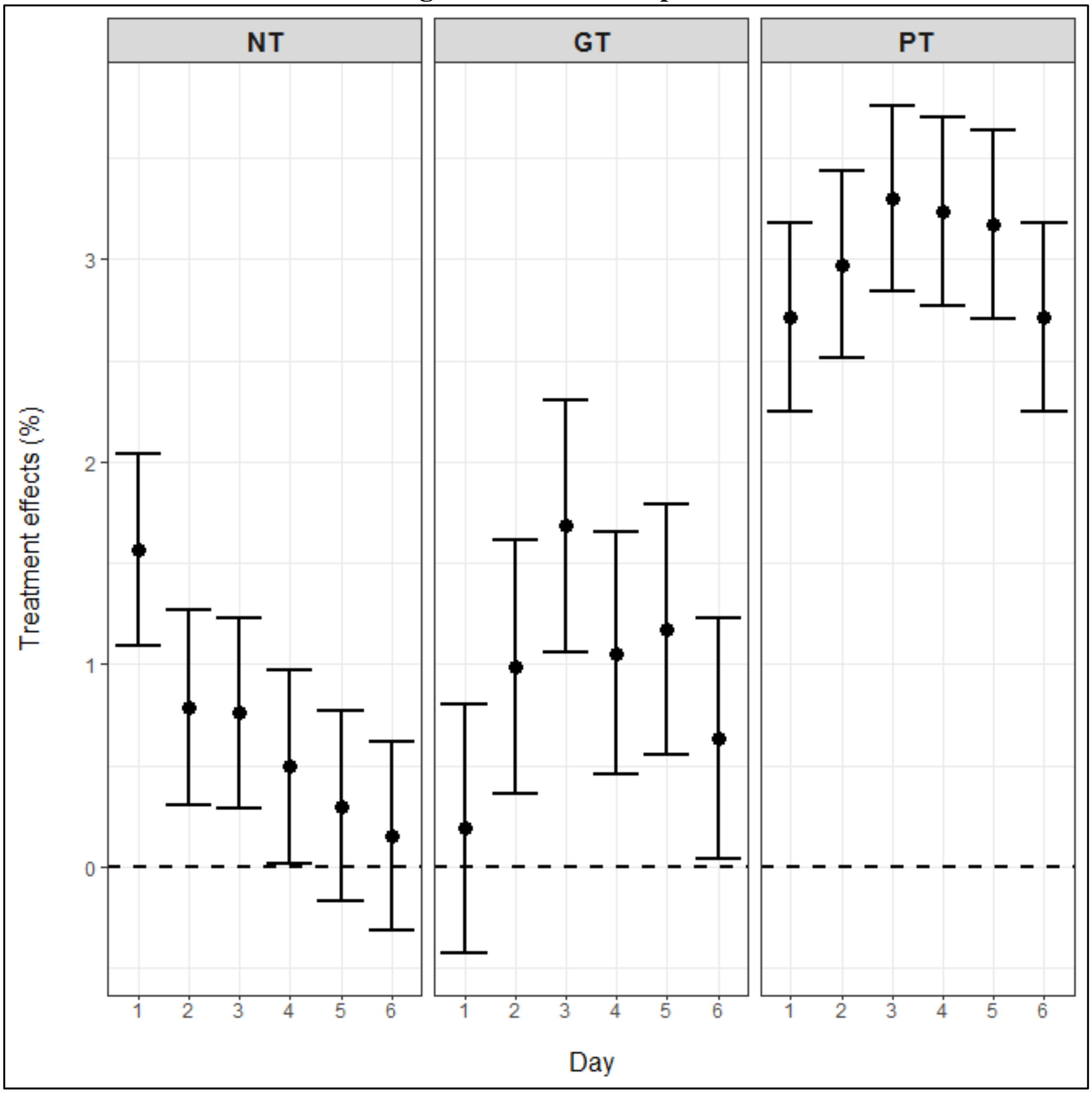
[31] It is true that children from lower-socioeconomic background have been shown (e.g., Muller et al, 2005) to be much less responsive to interventions aimed at tackling obesity and overweight. However, note that our population is quite homogeneous, so a lower socioeconomic background would still be a middle class family (medium-low). In the same way, a high-level family would be medium-high. We do not have extreme cases in our population.

[32] The per-period effects in Figure 4 are based on the parameter estimates reported in Table E.5 in the Appendix.

throughout the experiment, and much larger than what we see in the panels referring to treatments GT and NT.

The differences between *GT* and the *Baseline* become significant from the second period onwards. In the case of *NT*, the probability that the average student chooses a healthy food item remains significantly higher than in the *Baseline* treatment for the first 4 periods; in the last two periods, though, differences between *NT* and the *Baseline* become statistically indistinguishable from zero. These findings support the idea that the reported positive effect of the treatment dummies – specifically *GT* and *PT* – on the probability of choosing a healthy item comes from a stable effect that affects decisions in every period rather than being just a strong initial effect that vanishes over time.

Figure 4. Daily differences (from the *Baseline* treatment) in the rate of healthy choices during the intervention period



Notes: For each period (day), the figure plots the difference in the probability of making healthy food choices in *NT* (left panel), /*GT* (middle panel) and *PT* (right panel) vis-à-vis the *Baseline* treatment, in percentage points. Solid circles represent point estimates; vertical lines give the 95% confidence intervals.

6. Discussion and conclusion

A critical contemporary problem is that of poor diet and obesity. This has been linked to a variety of health problems, with concomitant economic consequences. Perhaps this is most important for children, where there are long-lasting health effects; addressing this issue at the root should offer the best hope for the future. Data from a variety of sources has indicated that children's eating patterns are far from encouraging, since children still fail to meet recommendations for the daily consumption of fruit and vegetables. Childhood obesity rates in the U.S. tripled from 1971-1974 and 2011-2012.

We conduct a field intervention designed to test whether it is possible to beneficially affect the diets of children. Previous work has shown positive effects with contemporaneous benefits, but there is little or no research that shows a more enduring effect. We conduct a field experiment to study whether and how it might be possible to incentivize children (without paying cash) 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.

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 from nutritionists 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 (the consumption of healthy foods roughly doubles) *that are entirely undiminished four months after our intervention was completed*.

We feel that our results are extremely promising, since they involve little or no financial cost and require only monitoring from parents. It is obvious that more research would be helpful, since our study is limited 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 resulting from poor nutritional habits and obesity.

References

Belot, M., James, J., and Nolen, P. (2016) “Incentives and children's dietary choices: A field experiment in primary schools,” *Journal of Health Economics* 50: 213-229.

Belot, M., Berlin, N., James, J., and Skafida, V. (2019) “The formation and Malleability of Dietary Habits: A field experiment with Low Income Families”. Mimeo

Borgmeier, I., and Westenhoefer, J. (2009) “Impact of different food label formats on healthiness evaluation and food choices of consumers: a randomized-controlled study” *BMC Public Health*, 9: 184.

Bornstein, G, Gneezy, U., and Nagel, R. (2002) “The effect of intergroup competition on group coordination: an experimental study” *Games and Economic Behavior*, 41(1): 1-25

Centers of disease prevention. source: National Center for Chronic Disease Prevention and Health Promotion, Division of Nutrition, Physical Activity, and Obesity https://www.cdc.gov/media/releases/2014/p0805-fruits-vegetables.html

Charness, G., and Gneezy, U. (2009), “Incentives to Exercise.” *Econometrica 77*(3): 909-931.

Charness, G., and Holder, P. (2019) “Charity in the Laboratory: Matching, Competition, and Group Identity” *Management Science*, 65(3), 1398-1407

Charness, G., Masclet, D., and Villeval M.C. (2014). “The dark side of competition for status”. *Management Science,* 60(1): 38–55.

Deci, E. L., & Ryan, R. M. (1985). “Intrinsic motivation and self-determination in human behavior”. New York: Plenum.

DeCosta, P., Moller, P., Frost, M.B., and Olsen, A. (2017), “Changing children’s eating behaviour – A review of experimental research,” *Appetite* 113: 328-357.

Downs, J. S., Wisdom, J., Wansink, B., & Loewenstein, G. (2013). Supplementing Menu Labeling With Calorie Recommendations to Test for Facilitation Effects. *American Journal of Public Health*. *103*(9), 1604-1609.

Fershtman, C., Gneezy, U., and List, J (2012) “Equity aversion: Social Norms and the Desire to Be Ahead” *American Economic Journal: Microeconomics*, 4(4): 131-44

Fisher, J. and Birch, L. (1999). Restricting access to palatable foods affects children's behavioral response, food selection, and intake. *American Journal of Clinical Nutrition*, 69(6), 1264:1272

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

GBD. 2016. *Global Burden of Disease 2016*. *The Lancet*. Available at http://www.thelancet.com/gbd, accessed 02 January 2018.

Haire-Joshu, D., and Tabak. R. (2016) “Preventing obesity across Generations: Evidence for early life intervention,” *American Review of Public Health* 37: 253-271.

Huberman, B., Loch, C. and A. Önçüler (2004), “Status as a Valued Resource,” *Social Psychology Quarterly*, 67(1), 103-114

Jaime, P.C., and Lock, K. (2009), “Do school based food and nutrition policies improve diet and reduce obesity?,” *Preventive Medicine* 48(1): 45-53.

Jansen, E., Mulkens, S., and 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., and 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.

Just, D.R., and Price, J. (2013), "Using incentives to encourage healthy eating in children," *Journal of Human Resources* 48(4): 855-872.

Kim SA, Moore LV, Galuska D, et al. (2014), "Vital signs: fruit and vegetable intake among children – United States, 2003–2010." *MMWR. Morbidity and mortality weekly report*, 63(31): 671– 676.

List, J.A., and Samek, A.S. (2015), "The behavioralist as nutritionist: Leveraging behavioral economics to improve child food choice and consumption," *Journal of Health Economics* 39: 135-146.

Lynch, C., Kristjansdottir, A., Te Velde, S., Lien, N., Roos, E., Thorsdottir, I., et al. (2014), "Fruit and vegetable consumption in a sample of 11-year-old children in ten European countries – the PRO GREENS cross-sectional survey," *Public Health Nutrition* 17(11): 2436-2444.

Majolo, B., and Marechal, L (2017) "Between-group competition elicits within-group cooperation in children" *Nature Scientific Reports*, 7

Micha R, Peñalvo JL, Cudhea F, Imamura F, Rehm CD, Mozaffarian D. (2017), "Association Between Dietary Factors and Mortality From Heart Disease, Stroke, and Type 2 Diabetes in the United States," *JAMA*, 317(9): 912–924.

Moffitt, T., Arseneault, L., Belsky, D., Dickson, N., Hancox, R., Harrington, H., Houts, R., Poulton, R., Roberts, B., Ross, Malcolm, B., Sears, R., W., and Caspi, A. (2011) "A gradient of childhood self-control predicts health, wealth, and public safety" *Proceedings of the National Academy of Sciences*, 108 (7): 2693-2698

Morizet, D., Depezay, L., Combris, P., Picard, D., and Giboreau, A. (2012) " Effect of labeling on new vegetable dish acceptance in preadolescent children" *Appetite*, 59: 399-402

Noble, E., and Kanoski, S. (2016), "Early life exposure to obesogenic diets and learning and memory dysfunction," *Current Opinion in Behavioral Sciences* 9: 7-14.

Ogden, J., Cordey, P., Cutler, L., and Thomas, H. (2013). "Parental restriction and children's diets. The chocolate coin and Easter egg experiments." *Appetite*, 61: 36-44

Patall, E.A., Cooper, H., and Robinson, J.C. (2008), "The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings," *Psychological Bulletin* 134(2): 270-300.

Pelchat, M., and Pliner, P. (1995) "Try it. You'll like it. Effects of information on willingness to try novel foods." Appetite, 24(2): 153-165

Ridder, M.A., Heuvelmans, M.A., Visscher, T.L, Seidell, J.C., and Renders, C.M. (2010), "We are healthy so we can behave unhealthily: A qualitative study of the health behaviour of Dutch lower vocational students," *Health Education,* 110(1): 30-42.

Rigal, N., Rubio, B., and Monnery-Patris, S. (2016). "Is harsh caregiving effective in toddlers with low inhibitory control? An experimental study in the food domain." *Infant Behavior and Development*, 43: 5-12

Roberto, C., Larsen, P., Agnew, H., Baik, J., and Brownell, K. (2010), "Evaluating the impact of menu labelling on food choices and intake" *American Journal of Public Health*, 100(2): 312-318

Rothman, R., Housam, R., Weiss, H., Davis, D., Gregory, R., Gebretsadik, T., Shintani, A., and Elasy, T. (2006) "Patient understanding of food labels: The role of literacy and numeracy," Am J Prev Med. 31(5): 391–398.

Rustichini, A. and A. Vostroknutov (2008), "Competition with Skill and Luck," Mimeo, University of Minnesota.

Sorhaindo, A., and 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 AN, Sonnenberg L, Riis J, Barraclough S, Levy DE (2012), "A 2-phase labeling and choice architecture intervention to improve healthy food and beverage choices," *Am J Public Health* 102(3): 527–33.
Thorndike AN, Riis J, Sonnenberg L, Barraclough S, Levy DE (2014), "Traffic-lights Labels and Choice Arquitecture: Promoting healthy food choices" *American Journal of Preventing Medicine*, 46(2): 143-149
Tsukayama, E., Duckwoth, A., and Kim, B. (2013) "Domain-Specific Impulsivity in School-Age Children" *Developmental Science*, 16(6): 879-893
Usual Dietary Intakes: Food Intakes, U.S. Population, 2007-10. Epidemiology and Genomics Research Program website. National Cancer Institute. http://epi.grants.cancer.gov/diet/usualintakes/pop/2007-10/. Updated *April 24, 2018*. Accessed *September 6, 2018*.
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." *Journal of the American Medical Association* 300 (22), 2631–2637.
Volpp, K.G., Troxel, A.B., Pauly, M.V., Glick, H.A., Puig, A., Asch, D.A., and Audrian-McGovern, J. (2009), "A randomized controlled trial of financial incentive for smoking cessation," *New England Journal of Medicine* 360(7): 699-709.
Vyth EL, Steenhuis IHM, Heymans MW, Roodenburg AJC, Brug J, Seidell JC. (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., and Gundersen, C. (2002), "Hunger: its impact on children's health and mental health," *Pediatrics* 110(4): e41.
Whitaker, R., Phillips, S., and 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.
Wisdom J, Downs J, and Loewensteing G. (2010) "Promoting healthy choices: information versus convenience." American Economic Journal: Applied Economics, 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.

Appendix A: Justification of trays

The diet we follow is one of the most important factors in determining our health. Because of this, it is important to follow a healthy diet, in accordance with the needs of each life stage. This type of diet, in addition to providing numerous health benefits, will also help to prevent the possible development of illnesses.

In this study, four trays of prepared foodstuffs which had not undergone any culinary treatment were used. The composition of each one of them was carried out with foodstuffs in accordance with the target population.

Trays with 0 points and 2.5 points:

The content of these two trays is ultra-processed products which each share low-quality ingredients. Among these ingredients, we find refined flours, non-virgin vegetable oils, added sugar and salt. All of these are associated with illnesses such as high blood pressure, diabetes, being overweight, depression, obesity and cancer, among others.

Refined flours are those in which the refining process has removed the bran from them and the germ from the whole grain. In this way, fibres, vitamins, minerals, anti-oxidants and phytosterols are lost; and, as such, the nutritional value is lowered. This occurs with sliced pan bread which, in addition, has sugar added to it.

The same refining process occurs with vegetable oils. Among the oils that we highlight in these products is palm oil. This is not a healthy oil as it is related to, as we mentioned before, an increase in cardiovascular risk, visceral fat. It is also related to certain types of cancer.

Considering that the WHO (World Health Organization) recommends that the ingestion of sugar in children does not exceed 15 grams per day, we can see how the consumption of one Phoskitos or 2 Bollycaos would exceed these recommendations and risk the health of a school-going child. The same occurs with sweets, Nocilla sticks, cereals, milkshakes and similar products.

In the case of cold meats, although their refined sugar and salt content is lower, they are still highlighted owing to their high content of salt and bad quality fats. Processed meats are related to some types of cancer and other pathologies.

As such, all of these products have a heightened calorie density and very few nutrients. Therefore, they only provide empty calories, what we refer to as “low nutritional density”.

The foodstuffs present in tray 0 do not contain interesting ingredients, they have empty calories but, most of all, they are in the lowest level because not only are the responses they cause in the organism non-beneficial but their consumption may result in serious health problems.

In the following, tray 2.5, we find foodstuffs that are not recommended either. However, in certain circumstances, they may be used as they contain some raw materials that do provide valuable nutrients such as protein in dairy products, minerals and vitamins in cereals and sources of protein and iron, such as chorizo or sausage. However, the advantages of these nutrients are outweighed by the disadvantages of the consumption of these dairy products which contain sugars, such as processed cereals. Bread is not a foodstuff of interest, nor is processed meat, because, despite being at a higher level because of that mentioned, they continue to “fail” as foodstuffs.

Tray with 5 points:

This tray is found in the middle of the trays in this study, containing properties of the two previous ones and the two last ones.

It should be highlighted that here foodstuffs are found that could be interesting such as milk which provides nutrients such as beneficial fatty acids, vitamins such as A, D, C and those from the B group, in addition to minerals such as calcium, phosphorus and potassium.

Bread provides fiber by being whole. By being less refined, it preserves its properties.

With respect to the yoghurts of this tray, it is noted that they are high in previously-mentioned sugars, this being the reason for not including them in subsequent trays.

Finally, these types of cheese and turkey only contain parts of the foodstuff's quality, as different substances have been added which, over a longer period of time, may have repercussions for our health.

In the tray with 5 points we have foodstuffs which contain nutrients of interest, they are preparations with more nutritional value but the advantages of their consumption continues to be improved by a wide range of foodstuffs that are different to the ones in them. Babybel cheese is not the best in its range, milk is not an essential foodstuff, but it is better if it is full-cream milk and does not contain sugars like milkshakes. Flavored yoghurts are not recommended; it is better to have natural ones without sugar and a turkey sandwich is not a preparation of nutritional interest.

<u>Tray 7.5 and 10 points:</u>

These trays are formed by minimally-processed dairy products, fruits, dried fruits, whole rye bread, jamón serrano and virgin olive oil.

They are characterized by a high content of total fiber, which relieves hunger, regulates intestinal motility and bacterial micro-flora substrates of the colon. They prevent illnesses such as constipation, diverticulosis, inflammatory bowel disease and irritable bowel syndrome, cancer, diabetes mellitus, atherosclerosis and obesity, among others.

Dried fruit has a good protein content, similar to that of legumes. They also have a high fat content which provides energetic and healthy properties due to the mono-saturated and poly-saturated fatty acids that they are made of. We would recommend 3 to 7 rations a week. Nuts and almonds are the best choices.

Fruits are a real dietary jewel. They are characterized by having a high content of simple sugars but, unlike ultra-processed foodstuffs, these, by being found inside the natural source, do not cause any of the previously-mentioned health problems.

Fruits, in general, do not contain fats. There are exceptions, such as avocados, olives and coconuts; where fat is a key component and very beneficial.

The vitamins that are present in fruits and vegetables make them very important, such as Vitamin C, which is more easily available in raw fruits, as once it is subject to thermal treatments, it decreases. On the other hand, Vitamin D in lactose is a very important vitamin at this moment as it helps growth due to its role in the absorption of calcium, as well as the metabolism of this one and phosphorous.

We include hard foodstuffs that are appropriate for reinforcing teeth, providing Magnesium, Calcium and Fluoride in varied and balanced diets. In addition, water is the essential drink, providing schoolchildren with Fluoride in addition to other minerals.

The dairy products of this group are strongly recommended as they help to battle different infections that may arise, helping to regenerate intestinal flora and to maintain a good system of defenses. It should be noted that full-cream products are better than skimmed ones.

In the tray that is deemed to have a score of 7.5, we already have a dairy product that would be better than the previous ones; natural yoghurt. Peanuts are very high in vegetable protein. Fresh fruit is always a good recommendation.

Finally, this tray has a score of 10 points as we find ingredients of high nutritional value, such as olive oil, the main foodstuff of our Mediterranean diet. It is rich in vitamins A, D, E and K. This foodstuff provides us with a wide range of benefits of which the following stand out: protection from vascular illnesses, expels grassy residues from the liver, anti-oxidant action, improvement of the digestive system, reducing, additionally, constipation, improvement of metabolic functions and cerebral development, stimulating the absorption of certain minerals such as calcium and controlling the level of glucose in blood among others. These last highlighted ones are key for this age.

We also find ourselves with our dear sandwich. However, this time it is more complete than in previous baskets and with better ingredients: whole rye bread, jamón serrano (and not processed meat such as chorizo or turkey), virgin olive oil and let's not forget that it would be ideal to add fresh tomato and even vegetables.

Dried fruits such as nuts and almonds with great nutritional value, fresh fruit and, of course, water, the best drink we can possibly have.

Appendix B. Summary of nutritionist's talk

Talks aimed at schoolchildren have as their main objective to help the child to differentiate real food and ultra-processed foods (pastries, snacks…). Established this differentiation, relates the state of health with the consumption of one or the other food, focusing on the following aspects of the ultra-processed ones (the information is transmitted using an easy-to-understand language adapted to their ages):

• Ultra-processed products have a high caloric density and that produce an increase in weight.

• Ultra-processed products generate a feeling of hunger shortly after their intake, so the tendency will be to eat again.

• Ultra-processed products produce pleasurable stimuli at the neurological level, so our brain will always prefer them over real food.

• Ultra-processed products have very intense and enhanced flavors compared to real food, so they will be the main choice for eating, leaving healthy foods aside.

Exposed the problem, the benefits of real food in health are listed, specifically a high consumption of vegetables (fruits, vegetables):

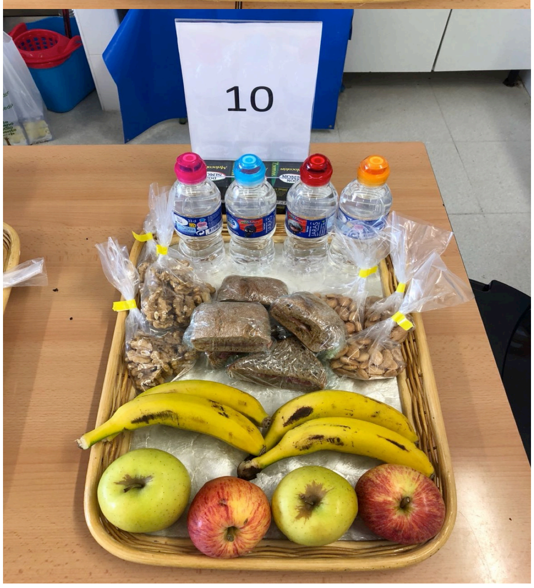
• In general, they produce or maintain an adequate state of health.

• They generate satiety, which prevents it from being chopped continuously between hours.

• They provide essential nutrients such as vitamins and minerals.

• The intake of vegetables in the diet provides a great source of fiber, which allows good intestinal health.

Appendix C. Composition of food trays



Appendix D.1. Socio-economic survey

1. This questionnaire has been filled by:

    a. Father

    b. Mother

    c. Both

    d. Other

2. The father has a University degree:

    a. Yes

    b. No

3. The mother has a University degree:

    a. Yes

    b. No

4. How many of the following items does the family own?

    Please, indicate for each item

	0	1	2	3	4	5	6	7 or more
Computers								
Laptops								
Tablets								
Video games								
Electronic books								
Cars or motorbikes								
Bathrooms in the house you are current living								
Number of properties (on top of the one you are currently living)								

5. Please, indicate your level of participation in your child's activities

    Please, indicate for each activity

	Never	Some days	Often	Every day
We encourage him/her to study				
We ask about his/her homework				

We check he/she does his/her homework				
We ask about his/her day in school				
We help with his/her homework				

Appendix D.2. Impulsivity questionnaire

Interpersonal Impulsivity. How often…

…do I interrupt other people?

...do I say something rude?

...do I lose temper?

...do I talk back when upset?

1 = almost never, 2 = about once per month, 3 = about 2 to 3 times per month, 4 = about once per week, and 5 = at least once per day

Schoolwork Impulsivity. How often…

...do I forget something needed for school?

...do I cannot find something because of mess?

...do I not remember what someone said to do?

1 = almost never, 2 = about once per month, 3 = about 2 to 3 times per month, 4 = about once per week, and 5 = at least once per day

Appendix E. Additional Results

Table E.1 – Descriptive statistics of relevant baseline covariates across treatment groups

	Baseline	NT	GT	PT
Both parents went to University[a]	0.032	0.046	0.000	0.080
Both parents are economically inactive[a]	0.081	0.031	0.028	0.000
Household Income[b]	26.22	35.04	28.69	30.61
Parents' involvement in children's school success[c]	3.5375	3.5621	3.4736	3.4583
Average grade (academic)[d]	7.24	7.097	7.21	7.84
Frequency of child's school absenteeism[e]	0.0526	0.0138	0.093	0.00
Proportion of male subjects[a]	48.68	56.338	53.488	52.702

Notes: [a]As proportion of all subjects allocated to each treatment

[b] Proxied by the percentage of households that own the following goods: house, car, desktop computer, notebook computer, and tablet.
[c]Index based on parents' self-reported attention to children's class attendance and homework and parent's overall involvement in their children's scholastic work.
[d] Coded on a scale from 1 to 10.
[e] Coded as 1 if frequent, 0 otherwise. Based on the teacher's judgment.

Table E.2 - Tests for differences in individual characteristics across treatment groups

	GT - Baseline	NT – Baseline	PT – Baseline	GT-NT	GT-PT	NT-PT
Both parents went to University	$Z = -0.401$, $p = 0.688$	$Z = -1.110$, $p = 0.266$	$Z = 0.957$, $p = 0.339$	$Z = 1.340$, $p = 0.181$	$Z= 1.758$, $p = 0.078$	$Z = 0.642$, $p = 0.532$
Both parents are financially inactive	$Z= -1.017$, $p = 0.309$	$Z = -1.226$, $p = 0.220$	$Z = 2.000$, $p = 0.045$	$Z = -0.061$, $p = 0.951$	$Z= -1.171$ $p = 0.242$	$Z = -1.22$, $p = 0.222$
Household Income	$Z = 0.450$, $p = 0.653$	$Z =1.803$, $p =0.070$	$Z= 0.693$, $p =0.488$	$Z = -1.100$, $p = 0.271$	$Z = 0.288$, $p = 0.773$	$Z = -0.637$, $p = 0.524$
Involvement in children's school success	$Z = -0.778$, $p = 0.436$	$Z = 0.380$, $p = 0.704$	$Z = -1.007$, $p = 0.313$	$Z = -1.165$, $p = 0.244$	$Z = -0.167$, $p = 0.867$	$Z = -1.420$, $p = 0.156$
Frequency of child's school absenteeism	$Z = 0.841$, $p = 0.400$	$Z = -1.299$, $p = 0.194$	$Z= -1.990$, $p = 0.046$	$Z = 2.004$, $p = 0.045$	$Z = -2.658$, $p = 0.078$	$Z = -1.010$, $p =0.311$
Average grade (academic)	$Z = -0.099$, $p = 0.921$	$Z = -0.497$, $p = 0.619$	$Z = 1.461$, $p = 0.144$	$Z = 0.350$, $p = 0.726$	$Z = 1.609$, $p = 0.110$	$Z = 1.967$, $p = 0.049$
Proportion of male subjects	$Z = 0.501$, $p = 0.616$	$Z = 0.925$ $p = 0.355$	$Z = 0.490$, $p = 0.624$	$Z = -0.295$, $p = 0.768$	$Z = -0.0817$, $p = 0.935$	$Z = -0.437$, $p = 0.661$

Note: Covariate balance checks based on the test statistics proposed by Hansen and Bowers (2008) to assess baseline differences between individual covariates across treatment groups.

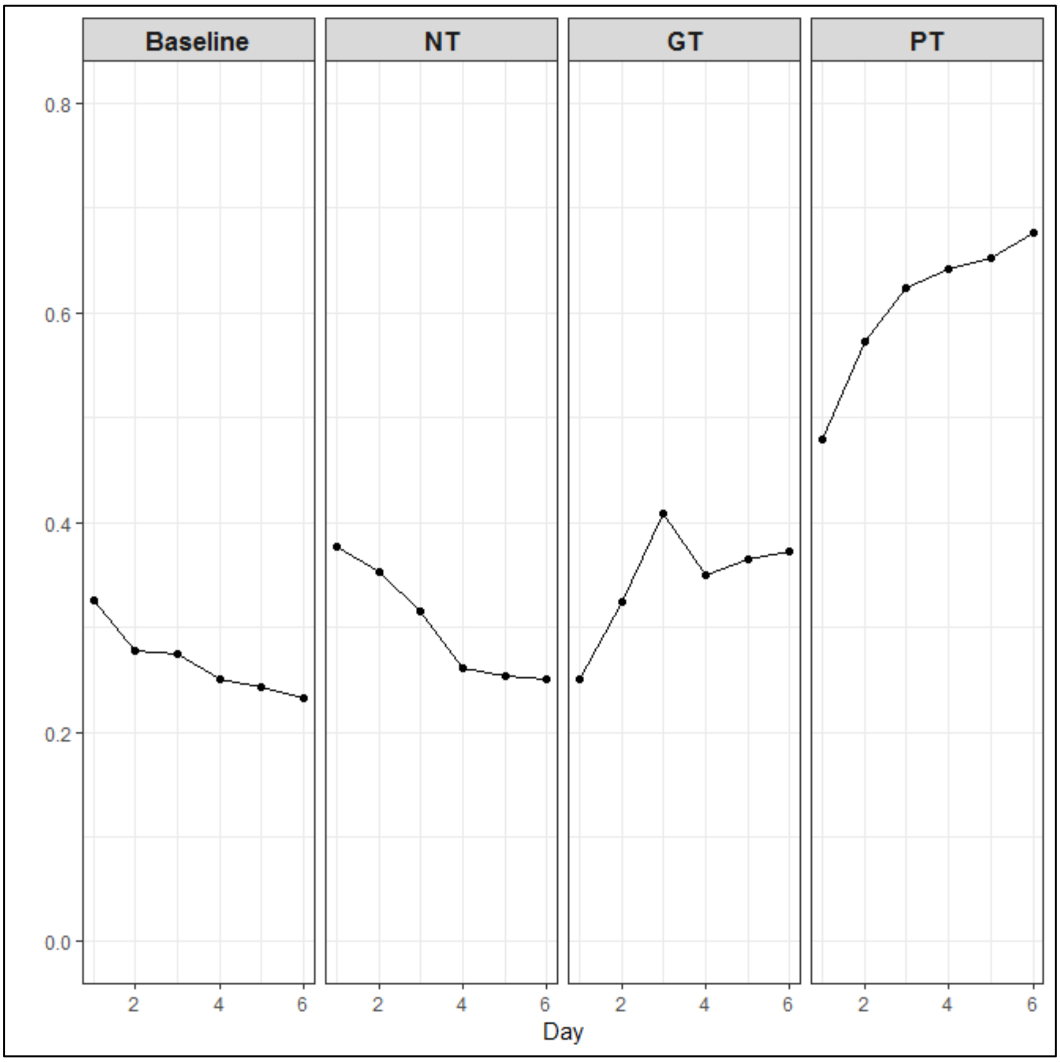
Table E.3 – Proportion of healthy choices and differences across treatments defining "healthy choice" as those with a grade equal to 10

Treatment	Proportion of healthy choices	School-level Wilcoxon ranksum tests vis-à-vis Baseline (two-sided)	Individual-level Wilcoxon ranksum tests vis-à-vis Baseline (two-sided)
Baseline	0.27	-	-
NT	0.30	Z = 1.091, p = 0.276	Z = 2.471, p = 0.014
GT	0.34	Z=1.964, p = 0.050	Z = 4.621, p < 0.001
PT	0.60	Z= 1.964, p = 0.050	Z = 15.665, p < 0.001

Table E.4 – Proportion of healthy choices and differences across treatments defining "healthy choice" as those with a grade greater or equal than 5

Treatment	Proportion of healthy choices	School-level Wilcoxon ranksum tests vis-à-vis Baseline (two-sided)	Individual-level Wilcoxon ranksum tests vis-à-vis Baseline (two-sided)
Baseline	0.49	-	-
NT	0.58	Z = 1.964, p = 0.050	Z = 4.756, p < 0.001
GT	0.59	Z = 1.964, p = 0.050	Z = 6.330, p < 0.001
PT	0.85	Z = 1.964, p = 0.050	Z = 17.959, p < 0.001

Figure E1. Robustness: Percentage of healthy choices over time, using a stricter definition of "healthy choice"



Notes: The figure replicates the analysis summarized in Figure 2 of the main text, but defining as "healthy choices" those with grades equal to 10 only.

Table E.5. Probit regression on the probability of choosing healthy food

	(1)
Constant	6.12*** (0.30)
Male	0.07 (0.10)
PT x Period1	2.81*** (0.24)
PT x Period2	3.07*** (0.24)
PT x Period3	3.39*** (0.24)
PT x Period4	3.33*** (0.24)
PT x Period5	3.27*** (0.29)
PT x Period6	2.81*** (0.25)
GT x Period1	0.29 (0.32)
GT x Period2	1.09*** (0.33)
GT x Period3	1.78*** (0.33)
GT x Period4	1.15*** (0.31)
GT x Period5	1.27*** (0.32)
GT x Period6	0.73** (0.31)
NT x Period1	1.66*** (0.25)
NT x Period2	0.88*** (0.25)
NT x Period3	0.86*** (0.25)
NT x Period4	0.60** (0.25)
NT x Period5	0.40 (0.25)
NT x Period6	0.25 (0.25)
Minimum Grade$_{-i,t-1}$	0.16*** (0.04)
Personal Impulsivity	-0.33*** (0.08)
School Impulsivity	-0.22** (0.10)