Changing Eating Habits – A Field Experiment in Primary Schools[1]

Michèle Belot (University of Edinburgh)

Jonathan James (University of Bath)

Patrick Nolen (University of Essex)

August 2013

Abstract

We conduct a field experiment in 31 primary schools in England to test whether temporary incentives to eat fruit and vegetables cause children choose and consume more healthy items. The intervention consists of rewarding children with stickers and little gifts for a period of four weeks for choosing a portion of fruit or vegetables at lunch. We compare the effects of two incentive schemes (piece rate and tournament) on choices and consumption over the course of the intervention. Furthermore we examine the effect of the interventions immediately after the incentives are removed and six months later to see if the temporary incentives had any lasting effect on dietary choices. We find that the two interventions, in general, had positive effects on choice and consumption and that the tournament works better overall; the piece rate scheme caused some students who missed the threshold for a bigger prize to become discouraged later in the week. However, we find that the treatment effects vary dramatically by age, gender and socioeconomic background. Furthermore, other than for children from poorer socio-economic backgrounds, we find little evidence of sustained long term effects.

Keywords: Incentives, Health, Habits, Child nutrition, Field experiment

JEL-codes: J13, I18, I28, H51, H52

[1] The authors thank the Esmée Fairbairn Foundation for their financial support (grant number 10-2206). We also thank participants to the seminars at Aarhus, Bath, Bristol, Ghent, Glasgow, Edinburgh, Essex, Queen's University Belfast, conference participants to the Workshop "The Determinants of Dietary Choices: An Interdisciplinary Approach" in Edinburgh (March 2013), the Royal Economic Society annual conference (Royal Holloway, 2013), the Florence Workshop on Behavioural and Experimental Economics (2013), "Field Days: Experiments Outside the Laboratory" (2013) and the International Conference in Public Policy (Grenoble, 2013). Finally, thank all the local authorities, schools, dinner ladies, parents and children who helped with the implementation of the experiment.

1. Introduction

Poor nutrition leads to numerous premature deaths and is a primary cause behind the recent surge in obesity which is not only one of the leading causes of death but is also contributing to the rising cost of health care in many developed countries.[2] According to the World Health Organization (2009) poor nutrition is related to three of the five highest risks for morality in the world: high blood pressure; high blood glucose; and overweight and obesity. To improve nutrition, reduce health care costs, and help prevent premature deaths, policy makers have been pushing information interventions, such as "5-a-day" campaigns, to encourage people to develop better eating habits. However, the effectiveness of information only campaigns have been questioned and policy makers are now considering if rewarding individuals for eating healthier is a better approach.[3]

Despite the push by policy makers, it is actually an open question if rewarding individuals for eating healthier will work or play any long lasting role in solving the problems caused by poor nutrition. Indeed some research shows that rewarding children for eating healthy items can lead to those items being less preferred.[4] While unable to look at individual level effects, recent work by Just and Price (2013a) has shown that schools where short term rewards are given for eating healthy items does lead to an increase in the proportion of children consuming a serving of fruits or vegetables at lunch time. Two weeks after the incentive is removed, however, there is no lasting change in the amount of fruits and vegetables consumed at the project schools. The lack of longer term effects could be due to the intervention period being too short or the incentives not being large enough.

Recent research in education (see Angrist and Lavy (2009), Angrist, Lang, and Oreopoulos (2009), Kremer, Miguel, and Thornton (2009) or Frey (2011)), smoking cessation (see Volpp et. al (2009) and Giné et. al. (2011)), and exercise (see Charness and Gneezy (2009) and Acland and Levy (2011)) has shown that incentives can induce individuals to engage in positive behaviour. Furthermore, the research on exercise has even shown that the habits developed during the incentive period can carry over to the post-intervention period.

[2] See Bhattacharya and Sood (2011) for an overview of the costs of obesity.

[3] See Havas et. al. (1998) for a review of one "5-a-day" program or Cliska et. al. (2000) for a review of many community based interventions. Some information interventions had significant effects but the size of the change was not always deemed clinically significant. Furthermore much of the criticism has been around the cost incurred for the minimal change in behaviour. See the NICE report for a review of why the UK government is considering using incentives to encourage healthy eating.

[4] See Birch et. al. (1982), Birch et. al. (1984), and Newman et. al. (1992) for examples.

However, unlike with the exercise research, the role of positive incentives on nutrition has not yet been examined at the individual level. Given the potential for external incentives to crowd out intrinsic motivation (see Gneezy and Rustichini (2000) for an example) and the fact that health research has shown negative effects of rewarding healthy eating, it is of utmost importance to examine the effects of incentivizing individuals to choose healthy items; especially given the recent push by policy makers.

Another body of work directly relevant to our study is the literature in Public Health evaluating the effects school-based health interventions. School-based programmes have the potential to be a very effective way of intervening since they enable large numbers of children to be targeted simultaneously. Although several such programmes have been developed in recent years, the increases in consumption they have achieved have been minimal, (Domel et al, 1993; Foerster et al, 1998; Nicklas et al, 1998; Perry et al, 1998; Baranowski et al, 2000; Reynolds et al, 2000). It would appear either that children's consumption of fruit and vegetables is particularly resistant to change, or that the methods employed to influence their eating patterns so far have not been well directed. However, there are some successful interventions that do not rely on rewards or use rewards only as part of a larger multifaceted program. For instance, the "Food Dudes" programme gives children the opportunity to watch 'role models' in the form of heroes of video cartoons eating fruit and vegetables. Role models that have been shown to be effective with children include cartoon characters (Harris & Baudin, 1972; Woolner, 2000), peers (Birch, 1980; Greer et al, 1991; Dowey, 1996; Hendy & Raudenbush, 2000; Hendy, 2002), mothers, unfamiliar adults (Harper & Sanders, 1975), and teachers (Hendy & Raudenbush, 2000).

In light of this policy push, the goal of this paper is to examine the effects of incentives on eating healthily at the individual level. We are particularly interested in whether there are heterogeneous effects and, if there are, which subgroups respond positively, negatively, or not at all. The individual level effects are of particular importance because, while, on average, the number of healthy items consumed at an intervention school may increase, vulnerable groups may eat worse due to the intervention. For example, as summarized in Belot and James (2011), boys tend not to respond to information treatments while girls respond positively. However, boys tend to have worse eating habits. In terms of societal welfare, one may not want to implement a policy if the increase in the proportion of healthy items

consumed is driven by an increase in consumption by those already eating healthily while those eating poorly decrease their consumption.

Given our focus on heterogeneous effects we examine two incentive schemes: an individual-based scheme (piece-rate) and a tournament. The previous literature on health and exercise has used individual-based rewards; the tournament incentive has not yet been used in the exercise or health interventions looking at incentivizing positive behaviour. However, work on gender and competition has shown that boys tend to be more competitive than girls (see Gneezy et. al. (2003) and Gneezy and Rustichini (2004)). Therefore, while boys might not respond to health interventions in general, we may find that boys respond to the tournament even if they do not respond the individual-based rewards. Using two schemes also allows us to examine one aspect of why the individual-based rewards in Just and Price (2013) may not have found effects beyond two weeks: does competition or a group based award work better than an individual-based reward? Keeping the intervention period constant, one could vary the incentive scheme and find longer term effects.

As in the research on exercise and smoking cessation, besides changing choices during the intervention period, the short-term effects could lead to changes in longer term habits. Indeed work on habit formation in exercise has shown that changes in behaviour remain after the invention has been removed, though, it is unknown just how long these changes last.[5] With regards to healthy eating there is evidence that habits form early on in life and track into adulthood (see Kelder et. al. (1994); Resnicow et. al. (1998); and Singer et. al. (1995)). Given this, we did not only monitor choices and consumption of students during the intervention period but also in the week after the intervention stopped and six months later. This allows us to see if any effects we observe during the intervention period last into the future. Furthermore we followed students in year two (roughly six years of age) and year five (roughly nine years of age) to examine if there were age effects.

Therefore, considering the habit formation literature and the closely related topic of incentivising positive behaviours, our paper has two major contributions: it examines the heterogeneous effects of incentivizing students to adopt a positive behaviour; and it examines how the type of incentive scheme matters during and after the treatment. We are also able to examine why these different incentive schemes have different effects. We

[5] See Acland and Levy (2011) and Charness and Gneezy (2009) for examples.

examine how these contributions fit into the habit formation literature by looking at how different subgroups respond both immediately after the incentive have been removed and six months later using a field experiment involving 31 schools across England.

We find that incentivizing students to choose healthy items has an overall positive effect on choice and consumption of healthy items at lunchtime in treated schools. However, the overall effect masks significant differences by age; in the piece-rate scheme year two students eat less when incentivized while there is a large positive effect for year five students. Besides stark heterogeneous differences in the effect, there are some differences by incentive type; overall the tournament works better while the positive effects of the piece-rate scheme disappear after two weeks. In particular, we find that the piece-rate scheme has a discouragement effect caused by the threshold requirement of having to eat four healthy items; students know when they missed the threshold and then actively choose *not* to have healthy items during the remainder of the week. This discouragement effect is not present in the tournament scheme where any number of healthy items could, in principle, result in an additional reward. The differences in the treatment effects also show up when the incentive scheme is removed: students assigned to the tournament treatment are more likely to continue eating healthily in the week immediately following the treatment. Six months later we do find some positive effects of the treatment for those that were affected during the incentive period.

Our results are broadly consistent with both the economic literature on habit formation and the health literature on how food preferences develop. Present-biased (hyperbolic) preferences, such as those discussed in Laibson (1997) and O'Donoghue and Rabin (1999), can explain the persistence of bad habits despite an individual being fully aware of the effects of poor nutrition. Individuals may over-weigh the initial costs of switching to being a healthier eater or under-weigh the longer term benefits. Thus an intervention providing an immediate benefit might be needed to get people to overcome the difference in perceived costs and benefits. Likewise the health literature discusses how neophobia (the predisposition to reject novel food) is non-monotonic with age and how children may learn not to prefer a food if they are incentivized to eat it.[6] Neophobia should be decreasing over

[6] See Birch (1999) for a good summary of the development of food preference and neophobia. Birch and Marlin (1982), Birch et. al. (1987, 1998), Sullivan and Birch (1990), and Cooke et. al. (2003) also provide strong evidence about the role of overcoming neophobia through repeated exposure to a new food or flavours.

the age of our sample meaning that the stark age effects – that year five students respond more to both incentive schemes – are consistent with the literature.

The results presented in this paper are important for policy makers and health officials trying to fight problems associated with poor nutrition. It shows that positive incentives do work in encouraging healthy dietary choices and that the results of a short term intervention can have long-lasting effects but that a "one-size-fits-all" reward scheme will not likely work. The heterogeneous effects suggest that health incentives need to be evaluated at the individual level and, consequently, different policies may have to be developed for different subgroups. Furthermore, increasing the length of time an intervention is taking place is not the only way policy makers can increase the likelihood that positive behaviours are adopted: for instance, tournaments could have an effect when individual-based schemes do not.

The remaining part of the paper is structured as follows. In Section 2 we present the experimental design. In Section 3, we present a simple conceptual framework and hypotheses that guide the analysis of the results. We present the results in Section 4 and conclude in Section 5.

2. Experimental Design

We recruited schools in a three step process.[7] First we approached all 150 Local Education Authorities (LEAs) in England to ask if they would be interested in participating; 22 responded positively. Second, we provided more information about the project to LEAs that responded and set-up meetings with them to answer questions and discuss how to recruit schools. We indicated to LEAs that we were interested in testing and comparing the effectiveness of incentives schemes in increasing choice and consumption of healthy items at lunchtime and that the interventions were specifically designed to target students who were generally considered unresponsive to health interventions. After the meetings 12 LEAs agreed to let us approach their schools and provided a list of at least three schools that would consider being involved. Finally we approached all 46 schools suggested by the LEAs; 31 of them agreed to participate.

[7] A companion paper, Belot and James (2013), documents the selection process of which schools choose to participate in this experiment. In particular they find only mild selection suggesting that the schools in the sample are representative of primary schools in England.

We recruited students from year two (aged 6 and 7) and year five (aged 9 and 10) in participating schools. Parents were provided with information about the project, asked to fill out a questionnaire, and were required to give consent to have data collected about their child. As agreed with the schools, all students in years two and five were included in the project. However, data about choice and consumption of healthy items were only recorded for children whose parents gave permission. Therefore, we have data on 626 students for the main part of the analysis.

Randomisation

We randomly allocated schools to one of three groups: control; tournament; or piece-rate. To make sure the groups were balanced we stratified our sample based on key observable features. We were particularly careful to make sure that, *ex ante*, the average school in each group had roughly the same number of students and looked the same in terms of observable characteristics.

Schools were grouped by their local education authority (LEA) and, if there were at least three schools in an LEA, one was randomly assigned to each of the three treatment arms. When there were more than three schools, the others were randomly assigned to treatment arms such that the overall sample was balanced based on observables. For the purpose of balancing the three groups we used the following characteristics: (i) proportion of female pupils; (ii) number of pupils; (iii) number of pupils in class groups (year 2 and year 5); (iv) proportion of students eligible for free school meals; (v) proportion of students eating free school meals; (vi) per pupil expenditure; (vii) per pupil expenditure on catering; (viii) percent of students achieving level 4 in both English and Mathematics; (ix) average point scores of students on level 4 exams; (x) average percent of students absent on a given day; (xi) percent of students absent from the level 4 exams; (xii) school type (religious or comprehensive); (xiii) whether a school was involved in the "Food for Life" Programme; (xiv) Ofsted School Categorization; and (xv) Ofsted Health Categorization (*OfHealth*).

The variables listed above were used so that the average school in each treatment arm was similar in ways that could have influenced whether the treatment scheme worked: socio-economic background; school quality; student quality; and school type. Variables (i), (ii), and (iii) relate to the demographic characteristics of the schools involved. Variables (iv) and (v)

relate to the economic background of the students. Variables (vi) and (vii) relate to the financial expenditure at the school level. Variables (viii) – (xi) relate to the quality of the student body at each school. Variable (xii) denotes if a school has a religious affiliation. Variable (xiii) denotes whether the school voluntarily chose to be part of the "Food for Life" programme which involves schools agree to teach students about healthy eating.[8] Variable (xiii) is the overall classification of the school based on its Ofsted results: 1 = outstanding; 2 = good; 3 = requires improvement; and 4 = inadequate. Variable (xiv) relates to the extent to which the pupils adopt a healthy lifestyle.

Using a random number generator, schools were assigned to one of the three treatment arms based on the LEA restrictions.[9] We then checked to make sure the three treatment arms were balanced based on all 15 observable characteristics. Indeed there were no *ex ante* differences between the control, piece-rate, and competition groups based on the 15 variables listed above.

Treatments

In our experiment we decided to incentivise choosing healthy items instead of eating healthy items. We did this for a few reasons. First, the health literature highlights how making rewards contingent on consumption of a particular food can cause children to have a lower preference for that item (see Birch et. al. (1982, 1984) and Newman and Taylor (1992) for examples). We wanted to minimize the potential for negative effects on healthy eating. Second, we wanted the experiment to be something that was relevant to policy and simpler to implement. Rewarding for choice removes any subjective judgement of the monitor to decide what constitutes an adequate amount of food consumed to be rewarded. Furthermore, schools can require students to take a healthy item at lunch but are unlikely to be able to force them to eat the item. Therefore the results of our study are likely to be more relevant to policies that are being considered at the school level now. Indeed the results of our study are especially relevant to determine if providing (or requiring a student to take) a healthy item at lunchtime has any follow through effect on consumption behaviour. Third, we

[8] See http://www.foodforlife.org.uk/ for further information about the programme.
[9] In schools with less than 75 pupils all students were invited to participate in the experiment, not just those in years 2 and 5. This was taken into account when we checked if the three treatment arms were balanced; especially with regards to variable (iii).

also wanted the program to involve minimal costs. Monitors were already people working in the school and with the students at lunch time. While we could have considered a multi-component approach such as "Food Dudes," combining such as aspects as bringing in role models, monitoring choices for each type of healthy item chosen, etc. this would have required a larger investment of resources and likely been too expensive for many schools to adopt.[10] Finally, rewarding for choice rather than actually consuming an item negates the possibility of cheating. For example, if rewards were based eating the pupil has an incentive to dispose of the fruit or vegetable; the student may hide it, give it to a friend or try to mislead monitors regarding actual consumption.

We examine the effect of providing incentives on choosing and eating healthy items using two different interventions: a piece-rate scheme and tournament scheme. In both schemes the pupils were given a sticker for choosing or bringing in a healthy item at lunch.[11] The piece-rate scheme was chosen because it is similar to many of the other individual based incentive schemes used in the healthy eating and habit formation literature (for instance, see Charness and Gneezy (2009) or Just and Price (2013)). The tournament was chosen because the literature on gender and competition suggests that boys respond more to competition than girls (see Gneezy and Rustichini (2004) and Gneezy et. al. (2003)). Given that boys tend not to respond to traditional healthy eating interventions, the tournament was seen as an incentive scheme that could get boys to respond. However, gender differences in competition can vary by task (see Iriberri and Rey-Biel (2013)). Therefore if the task of choosing a healthily item is viewed as a 'favouring females' then even the tournament scheme might not get boys to choose or consume healthy items.

In both schemes students received a sticker every day they choose or brought in a healthy item at lunchtime. Then, at the end of the week (Friday afternoon), each student had the opportunity to pick a larger prize depending on the incentive scheme in which the student was enrolled. In the piece-rate scheme, if a student collected four stickers in the week she or he was allowed to choose an additional reward such as an item of stationery or a small toy from a reward box.[12] If the student had three or less stickers, though, the student could not

[10] See Horne et. al. (1995, 1998)

[11] Examples of the stickers can be seen in the appendix. All students were given a list of fruits and vegetables that would be rewarded if they were included in packed lunches; the list is included in appendix B.

[12] See appendix for pictures of some of the rewards from which students were allowed to choose.

pick an additional award and the stickers did not count to earning the prize next week. In the tournament scheme, students were assigned to a random groups of four, and the student with the most stickers in each group was able to select an additional reward from the reward box. In the case of a tie all students with the highest number of stickers in the group were eligible for an additional reward. The groups were revealed at the end of the week so students would not engage in strategic behaviour, such make choices based on other group member's actions or absenteeism. For example, if a pupil was absent on Monday then the others in their group would know that that pupil could only collect a maximum of four stickers. The groups were changed each week so the children could not anticipate with whom they would be competing and unused stickers did not carry over to the next week.

Timing

Before the interventions began a background survey was sent to the parents that covered information on age, gender, ethnicity, primary language, height, weight, and typical dietary habits. Then, starting the second week of October in 2011, we monitored what students ate at lunch in all 31 schools. Lunch monitors[13] recorded if a student chose a healthy item or brought a healthy item in with a packed lunch and if the student consumed none, some or more than half the item. On Friday that week students took a food knowledge test and a "spot-the-difference" test.[14] The food knowledge test required the student to identify seven pictures of different items (e.g. celery or snickers bar) and mark if each item was healthy or not. The "spot-the-difference" test was designed to test a student's concentration and required a student to compare two sets of 30 dice that were arranged in a six-by-five square. There were five differences between the two sets of dice; the student was asked to circle the five differences. Students had 10 minutes to complete each test.

The students went on half-term break for one week after the baseline data was collected. Upon returning to school the students were reminded of the project and students were monitored for the next five weeks. At control schools, the lunch monitors continued to monitor students in the same way they did during the week in October: they collected data on whether a student choose or consumed a healthy item. At the tournament and piece-rate

[13] Lunch monitors were lunch ladies who worked in the cafeteria or school assistants who were present at lunch time and sat with the students as usual during the lunch period.
[14] Examples of both can be seen in the appendix.

schools students were incentivized to choose a healthy item for a period of four weeks.[15] Each day a student choose or brought in a healthy item with a packed lunch[16] the student received a sticker. Furthermore, as discussed above, at the end of each week, students would get a large prize based on the type of incentive scheme in which they were enrolled.

On the fourth Friday of the treatment, the students completed another food knowledge and "spot-the-difference" test and were reminded that it was the last day of incentives. The following week, immediately after the treatment, the choices and consumption of students were still monitored. This allows us to see if there was any effect on choice and consumption after the incentives were removed. To examine the longer term effects we of the incentives we also went back to schools six months later, in June 2012, and monitored the choice and consumption of the same students. Unfortunately, the week we chose did not work for all schools so only 21 of the original 31 schools participated in the six month follow-up.[17]

3. Conceptual Framework and hypotheses

The idea of using incentives to change dietary choices and potentially longer term habits relies on three insights from the health literature: (1) tastes and preferences of children can be changed by exposure to new foods; (2) children are relatively impatient and may suffer from projection biases; and finally (3) current consumption has an effect of future consumption (habit formation).

Given this framework we can set out the hypotheses that our experiment was designed to test. As in other work on incentivizing positive behaviour, we expect that our experiment will have a positive effect on choice and consumption of healthy items.

Hypothesis 1: Students will choose more healthy items when they are rewarded for taking a healthy item at lunchtime.

[15] Just and Price (2013a) incentivized students for a period of 2-3 weeks and found no longer run effects. Therefore, we chose to incentivize students for a longer period of time; 1-2 weeks longer.
[16] With the questionnaire and again at the start of the five weeks of monitoring, the parents of all students received lists of what items would count as healthy if they were included with packed lunches.
[17] Fortunately, we find no evidence of selective drop-out, i.e. the results using the data from the first 6 weeks do not change if we exclude the 10 schools for which we have no data in June.

By providing a reward for choosing a healthy option, the benefit of taking a healthy item at lunchtime will have increased for each student. Therefore we would expect that, while the incentive scheme is running, students are more likely to choose a healthy item. This would be consistent with the work by Gneezy and Charness (2009) and Just and Price (2013a). Furthermore, the effect is likely to differ by subgroups. Gender differences exist in regards to consumption of healthy items (see Belot and James (2013)) and students from poorer families may value the prizes more. The health literature highlights ages effects with regards to food preferences and tastes (see Birch (1999) and the references therein); suggesting that there is likely to be differences in the effect of the incentive by age. Furthermore, students who eat packed lunches are particularly likely to be effected by the incentives because recent surveys have found that 99% of packed lunches failed to meet the government set nutritional standards (see Evans et. al. (2010) for more details); in the baseline, 28% of our students brought in a packed lunch every day and 41% brought in a packed lunch at least four days.

Hypothesis 2: Students will consume more healthy items when they are rewarded for taking a healthy item at lunchtime.

The behavioural literature has shown us that the default option can effect choice made by individuals (see Keller et. al. (2011), Choi et. al. (2003), and Johnson and Goldstein (2003) for examples) and even help reduce calorie consumption (Wisdom et. al. (2010)). As a result health initiatives at schools have started to require students to have a healthy item on their plate.[18] By incentivizing students to take a healthy item our experiment is likely to have a follow-through effect on consumption. Furthermore, unlike previous studies, our students have no incentive to lie or cheat regarding the amount of the healthy item they consumed; the rewards are only based on choice. This means that we can estimate the causal effect of how an increase in having a healthy item on one's lunch tray effects consumption. As with choice, there is reason to expect that the effect on consumption will vary with gender, age, and socio-economic background. Therefore we will pay particular attention to subgroup effects.

[18] See Dillon and Lane (1989) for an evaluation of the differences between offering and serving a fruit or vegetable and Just and Price (2013b) for the effect of requiring schools to serve healthy items.

Hypothesis 3: Students will choose and consume more healthy items after the incentive is removed than before.

Given how food preferences develop, if students have been eating more healthy items during the intervention period they may have developed a preference for healthy items or developed a habit of eating healthy items at lunch time. Becker and Murphy (1988) and Becker (1992) develop a model of habit formation where the marginal utility of today's consumption is correlated with historical consumption. Therefore a small change in today's behaviour – caused by an exogenous increase in the benefit of consuming a healthy item for instance – could lead to long term changes in consumption. To measure this we look at the proportion of healthy items a student chooses in the week immediately following the intervention and six months later.

4. Results

4.1 Randomisation and pre-intervention summary statistics

We begin by comparing our treatment and control schools in the baseline period. Table 1 presents the means of the outcome variables and other covariates by control and treatment groups. The final two columns show the p-values for treatment and control differences. There are no statistically significant differences for the overall sample suggesting that, based on observables, the randomization worked as expected. Furthermore, even though they are insignificant, the size of the differences (in most cases) are less than one standard deviation, suggesting that the control and treatment groups are close to being observationally equivalent in the baseline.

4.2 Descriptive figures

We start with presenting three simple graphs, showing how our three main variables of interest changed over time on average. We will examine one main variable for choice, which we will label as "choice," and it will equal one if a student choose a healthy item on a given day. For consumption we will focus on two variables: "try" which will equal one if the student eats at least some of the healthy item she choose that day; and "eat more" which will equal one if the student eats more than half of the healthy item she chose that day. For the initial part of the analysis, we will look at the weekly mean outcome for a given pupil since

the largest prize in each incentive scheme is based on weekly consumption. Specifically, we construct the following variables:

$$Mean(outcome)_{isz} = \frac{\sum_{day=1}^{5} Outcome_{i,s,z,day}}{\sum_{day=1}^{5} Present_{i,s,z,day}}$$

Where $outcome_{i,s,z,day}$ is the outcome of interest for pupil *i* in school *s* in week *z* on a specific *day* of the week. *Present* indicates whether the child was present at lunch or not (this allows us to take account of possible absences).

Figures 1, 2 and 3 show how the three variables of interest changed over time.[19] Figure 1 shows the trend in 'choice' for the control, piece-rate, and tournament schools. We see that the percentage of healthy choices is consistently higher in both types of treatment schools than in the control schools during the interventions. However, the differences disappear immediately after the incentive is removed (week six) and six months later, we even observe a relative decrease in the competition schools. Figure 2, which shows the trends for 'try,' shows a much larger effect of the treatment schemes and a distinct pattern in piece-rate schools. There is a sharp increase in week two, both in the tournament and piece rate treatments. But while this increase is sustained for the remaining weeks in the tournament schools, it reverses slightly in the piece rate schools after two weeks. Six months later, we find higher rates of trying in the competition schools, followed by the piece rate schools and finally the control schools. Figure 3, which is for 'eat more' shows roughly the same pattern as Figure 2: there is a sharp increase in week two, which is sustained over the period of the intervention, in week six (immediately after the intervention) and six months later. We also note an increase in the control schools six months later; this could be due to a seasonal effect or other interventions that have taken place in the schools.

4.3. Intervention Effects

We begin by reporting the average treatment effects for the three main outcome variables of interest: 'choice,' 'try,' and 'eat more'. The treatment lasted for 4 weeks. Therefore, in the first instance, we present the results separately for each treatment by week.

[19] In the piece rate scheme an interesting threshold to keep in mind is 80% which corresponds to the threshold to pass to receive an additional reward at the end of the week.

This allows us to see if and how the intervention may vary over the intervention period; students may become tired of the prizes and stop responding after the first few weeks.

We estimate the following linear probability model:

$$Mean(outcome)_{isz} = \beta_0 + \sum_{z=2}^{5} \beta_z Week_z + \beta_{comp} Comp_s + \beta_{Piece} Piece\ rate_s$$
$$+ \sum_{z=2}^{5} \beta_{z,Comp}\, Comp_s \times Week_z + \sum_{z=2}^{5} \beta_{z,Piece}\, Piece\ rate_s \times Week_z + \alpha_i + u_{ist},$$

Where $outcome_{isz}$ is the outcome of interest for pupil *i* in school *s* in week *z,* $Week_{2,\dots5}$ are week dummy variables, *Comp* and *Piece rate* are dummy variables indicating whether school *s* is part of the competition or treatment group, and the unobserved error term is $\alpha_i + u_{ist}$. The parameters to be estimated are β_0, β_{Comp}, β_{Piece}, and β_z, $\beta_{z,Comp}$, $\beta_{z,Piece}$ (for *z=1,…5).*

Since our dependent variables are bounded upwards (at 1), we estimate this specification on the whole sample and on the sample of students most likely to respond: those who did not have a mean outcome equal to one in the baseline (referred to later as "Less than 100%"). We are particularly interested in the latter group because those who are not choosing or consuming a healthy item every day is the subgroup that could most benefit from the intervention – they could be encouraged to eat healthier – whereas the group that choose and consumed a healthy item every day is already eating as healthy as possible in our framework.

In the sample of all pupils, there was little effect of the piece rate scheme on all three outcomes of interest but there was a positive effect in the tournament scheme. Column [1] shows that competition increases the probability that a student choose a healthy item by 5.9 percentage points in the first week of the intervention. Thereafter the difference-in difference estimates are positive but no longer significant. There are no significant effects for the piece-rate scheme in column [1]. Column [2] we can see a similar but clearer pattern 'try:' there is an 8.5 percentage point increase in the probability of a student trying a healthy item in the first week of the intervention; a 9.5 percentage point increase in the second week of the intervention; and a 7.5 percentage point increase in the third week of the intervention. Likewise there is no significant effect for the piece-rate scheme, though, the point estimates

are large in the first two weeks of the intervention. Column [3] shows the effect on 'eat more:' there are no significant differencesal, though, the coefficients are all positive and of the same order of magnitude as those in column [2].

The picture looks slightly different in columns [4]-[6] when we consider the restricted sample, i.e. those who did not choose or consume a healthy item every day during the baseline and, thus, there was room for them to improve their nutritional habits. Roughly we see the same picture for each outcome variable of interest; the tournament scheme has a large significant effect that does not dissipate over the intervention period where as the piece-rate scheme has an initial positive effect in the first two weeks of the intervention and then drops off over there rest of the intervention period. By the last week of the intervention, the percentage of healthy choices increased by 17 percentage points, the percentage of portions tried increased by 18 percentage points and the percentage of portions eaten for more than half increased by 15 percentage points in the tournament scheme but there were no significant results for the piece-rate scheme. This means that the tournament scheme, on average, caused students to choose, try, and eat more than half of one additional healthy item per week.

These first results highlight the importance of testing different mechanisms and schemes while using incentives which a simple piece-rate scheme alone, as used in most other studies, could have missed.

Choice and consumption dynamics

Having established that there are differences in the effectiveness of the incentive schemes we now move onto explain why it might be the case the tournament scheme appears to work better in comparison to piece-rate scheme. While the previous analysis was based on week level regression we do have data at the day level for each student. In this section we will analyse the student-day level data and examine the dynamics of choice and consumption throughout the week. In particular we will look at if there are different dynamics based on the two types of treatments.

First when looking at choice, the students who were most responsive to the treatments were those who had not chosen a healthy item 100% of the time during the baseline period. Column [1] in panel A of table 4 shows the day level regressions for that

sample of students.[20] As in the weekly regressions, column [1] shows that competition had a large and significant effect on choice during treatment weeks; students assigned to the competition group were 16 percentage points more likely to choose a healthy item. There was a large imprecisely estimated effect due to piece-rate.[21] Columns [2]-[6] show the effect of the treatments for each day of the week. The effect of the tournament scheme started off very strong at the beginning of the week; on Mondays and Tuesdays students were 20 and 24 percentage points, respectively, more likely to choose a healthy item. As the week went on the effect dissipated, though; the point estimate decreased from 18 percentage points on Wednesday to 4 percentage points on Friday. Piece-rate had the opposite effect; students were more likely to choose their healthy item at the end of the week. The only significant increase in choice due to the piece-rate treatment took place on Friday when students were 25 percentage points more likely to choose a healthy item.

In the tournament scheme students did not know how many healthy items they would have to choose to get a prize at the end of the week; if they choose five healthy items, though, they were guaranteed a prize. Since students did not know who was in their group and some students did not choose a healthy item every day, a student could assign a subjective probability to winning given how many items she had chosen during the week.[22] Based on a student's subjective probability one could calculate the number of healthy items that a student would ideally want to consume each week to maximize her benefit from getting a prize subject to her disutility from having to choose a healthy item. Once a student has reached that number of healthy items she could switch back to her preferred unhealthy item. This type of pattern would explain why the effect of competition tapered off during the week.

In the piece-rate scheme the threshold to obtain the weekly prize was known and fixed. Given the exogenous pre-determined goal a student had to reach there was room for discouragement to take place; if a student had not eaten a healthy item on Monday or Tuesday then there was zero probability the student would get a prize that week. Besides

[20] There was no effect – either positive or negative – on the sample of students that had chosen a healthy item 100% of the time during the baseline week. The effect on all students is just a weighted average of these two groups.
[21] Redoing the weekly tables in the results section using day level data does not yield any qualitatively different results. The imprecision of the piece-rate estimate is primarily driven by the fact that year 5 students responded positively while year 2 students responded negatively; this will be discussed later when the heterogeneity of the effects are discussed.
[22] In fact there was an increasing probability of winning the prize based on the number of healthy items one chose. There was a small probability (under 5%) chance of winning if a student had chosen zero or one item, a 6.7% chance of winning if a student chose two items, a 21% chance of winning if a student chose three items, and a 39% chance of winning if a student chose 4 items.

having no external incentive from Wednesday onwards, a student might also feel discouraged and choose not to select a healthy option. Therefore, to examine this discouragement effect we break the sample into two groups in columns [7] and [8]. Column [7] contains students had 'missed' the prize as of Wednesday, i.e. they had not chosen a healthy item on Monday or Tuesday. Column [8] contains those students who had chosen at least one healthy item before Wednesday. The effect of piece-rate is large and significant for those who still have a chance of getting a prize, i.e. those in column [8]. However, for those that have missed the chance of getting a prize the effect of piece-rate is estimated to be negative, though, it is insignificant. This means that as the week goes on the incentive to choose a healthy item wears off for those that miss the goal in the piece-rate scheme. However, this is not the case in the competition treatment because there is always a positive probability of winning the prize no matter how many items the student has consumed during the week.[23]

These results speak to the intrinsic incentive differences between the two treatments. The external, known goal in the piece-rate can lead to a lack of incentive because of previous choice patterns. However, there is always a positive chance of winning in the competition treatment because the goal is unknown and endogenous to the system. In the habit formation literature with regards to healthy eating the goals have all been exogenous and known. Therefore, there is room to design rewards like the tournament scheme that can have a greater effect (than piece-rate) over the same period of time.

Panel B shows the effect of the treatment on consuming at least some of a chosen healthy item on any given day. Column [1] of panel B shows the same effect as the weekly results; competition has a large significant and positive effect and piece-rate has a positive effect, though, it is insignificant and less precisely estimated. Columns [2]-[6] show the daily effects and, as with choice, piece-rate only has an effect one day a week, in this case on Thursday. More importantly, columns [7] and [8] show the differential effect of piece rate for those who have made different choices previously in the week. There is no effect of the treatment for students that have missed the chance to get a prize by Wednesday, i.e. those that have not chosen a healthy item on Monday or Tuesday; shown in column [7]. There is a large and positive effect of piece-rate for those that still have a chance of getting the prize,

[23] Indeed the point estimate for competition is the same in columns [7] and [8] showing that the choice pattern before Wednesday does not change the effect that the competition treatment has from Wednesday onwards.

i.e. those that have chosen at least one healthy item on Monday or Tuesday; shown in column [8]. The dynamics around trying a healthy item are the same as choice for the piece-rate treatment.

The effect of the tournament scheme on consuming at least part of a healthy item is somewhat different than in choice. Columns [2]-[6] show a large positive effect of competition that does not dissipate over the week. Students were not incentivized to try the items therefore these results speak more to how a student behaves once a healthy item is on her plate. Since students were choosing, on average, less healthy items as the week went on, then this constant effect suggests that those who were choosing healthy items on Monday were less likely to consume them, while those choosing them on Wednesday were more likely to try them. On Friday, there was no effect of tournament on choice; therefore, the large effect seen in panel B has to mean that those who did choose a healthy item definitely tried it. Friday is typically "fish and chips" day at many schools in the UK; for a student to give up chips she probably has to really like the healthy option. The main effects on try, though, are those shown in columns [7] and [8]. Again, the point estimate for competition does not vary between columns [7] and [8]. This means that previous choice patterns in the week do not effect consumption choices later in the week systematically, unlike for the piece-rate treatment.

Heterogeneity in treatment effects

We now turn to the heterogeneity in treatment effects. We will focus on looking at the effect of the two incentive schemes by gender, year-group, and free-school meal status.

Tables 5A and 5B report the estimates broken down by gender. With boys, shown in Table 5A, we find significant short run effects of the piece rate only when we consider the whole sample. But if we look at the more restricted sample, we find large and sustained effects of the tournament on choosing and trying in particular and also on 'eat more.' As in the overall sample, boys initially respond to the piece rate scheme but the effects start to dissipate after the second week of the intervention. With girls, though, as shown in Table 5B we find a striking pattern: whether we consider the whole sample or not, we find that girls respond significantly to competition particularly in terms of trying a healthy item, but we find no significant effects of the piece rate, not even at the beginning of the intervention period.

Tables 6A and 6B show the effects for year two and year five students respectively. The two groups respond quite differently to the treatments. Year two students appear to respond adversely to the piece rate, while we find very large and significant effects for students in year five. In the whole sample, columns [1]-[3], we find that the percentage of portions chosen and the percentage of portions tried are estimated to have decreased to some extent for year two, although the effects are only significant for week 4. In the restricted sample, columns [4]-[6], there is also a significant decrease for the piece-rate in week two, the first week of the intervention, for trying a healthy time. Overall – in both the restricted and entire sample – there is a small positive effect of the tournament for students in year two, though. However, when looking at year five students the results are strikingly different: the piece-rate scheme increased by between 14 and 17 percentage points for choosing, 22 and 25 percentage points for trying, and 22 and 26 percentage points for eating more than half. Furthermore, with the year five students, the effect of the piece-rate does not decrease over the incentive period. When looking at the restricted sample in columns [4]-[6] the point estimates for piece-rate are even larger and significant over the entire intervention period. However the restricted sample does show that the tournament scheme also has a large significant effect that is similar in size and magnitude as that shown for the complete restricted sample in Table 3.

Next we examine the effect by two other sub-groups, those who are eligible for a free school meal (FSM) or not. The results for non-FSM students are reported in Table 7A and the results for FSM students are reported in 7B. Children eligible to receive free school meals at school usually come from lower socio-economic backgrounds. Children from lower-socioeconomic background have been shown to be much less responsive to interventions at aimed at tackling obesity and overweight (see for example Muller et al, 2005). For both of these groups it appears that the piece-rate scheme only had little or no significant effect on whether a student chose or consumed a healthy item during the intervention period. However, we find large and significant effects of the tournament scheme for both the FSM and non-FSM students on the probability of trying a healthy item. When looking at the restricted sample, we find that the likelihood that a student tried a healthy time increased by nearly 20 percentage points in the first week of the intervention both the FSM and non-FSM groups. Furthermore, the large significant effect of the tournament do not disappear over time for either the FSM or non-FSM group.

The examination of the heterogeneity of the treatment effects provides key insights to how average effects can mask potential downsides that are of particular importance when considering policy. With regards to the tournament, it is clear that the overall effect is large and significant: the tournament caused students to choose and consume more healthy times and those effects do not vary much by subgroup. However, looking at the average effect of the piece-rate, would mask extreme differences in the response by year groups. Table 3 suggests that the piece-rate had an initial effect on choice and consumption of healthy items but that the positive effect dissipated after two weeks. This general pattern is only present for boys (table 5A) and non-FSM students (table 7A); girls (table 5B) and FSM students (table 7B) did not have a significant response to piece-rate. However, year five students did respond positively and the effect did not dissipate over time and year two students responded negatively to the piece-rate. Given that piece-rate seems to be effected by discouragement, these heterogeneous effects could be due to differential responses to discouragement. Furthermore, if healthy eating habits developed at a young age are more likely to last, then it is uncertain that a policy maker would want to introduce an individual based scheme such as the piece-rate scheme used here since it appears to discourage younger students from choosing and eating healthy items.

4.4. Medium term effects

We now turn to the post-intervention weeks and evaluate whether the changes observed during the intervention persist in the week after the incentives are removed. The results are presented in Tables 8, 9 and 10. Given that tables 3-7 allow us to see how the incentives vary over the intervention period, we now look at the overall effect of the intervention period in one summary variable, called "wk 2-5." However, our main coefficient of interest is the interaction of the incentive scheme with week six, which is the week immediately following the intervention.

Table 8 shows the effects by gender. As expected there is no discernable effect of the treatment on the whole sample. Instead we see a strong positive effect on those most likely to have responded, the restricted sample in columns [4]-[6]. With regards to the tournament, there is an average increase of 18.5 percentage points in the probability of choosing a healthy item for boys, as shown in column [4]. In column [5] we see that there is a 15.4 percentage point increase in the probability of trying a healthy item. Column [6] shows no effect on eating

more than half. When the incentive is removed in week six, boys are more likely to continue eating part of the healthy items on their plate. However, there is no other effect. With regards to piece-rate, boys respond positively for all three outcomes of interest but there is not carry over into week six. When looking at girls, Table 8 panel B, we see that for the restricted sample, the significant results from the incentive period are for the tournament on consuming (try and eat more than half) of a healthy item. In both of those cases, we also find lasting positive effects after the incentives have been removed. These results suggest that when girls respond to a treatment it is more likely to have at least a medium-run effect than when a boy responds to a treatment.

Table 9 shows the effects by year group. The piece rate has a strong negative effect on choosing, trying, and eating more than half of a healthy item for students in year two. What is worse is that the negative effects persist even when the incentives are removed; all the point estimates for week six interacted with piece-rate in columns [1]-[6] are negative and significant. When looking at year five in panel B, we find very large and significant effects of the piece rate and most of them are sustained once the incentive is removed. We find a slightly different pattern for the tournament. The tournament, in general, effects both year groups positively, however the positive effect on choice and consumption only carries over into week six for the year two group – refer to column [2], [3], [5], and [6]. However, in no case do the positive effects of the incentive period carry over into week six for the year five group. This shows that when younger students respond to the incentive, even if it is negative, it is more likely to persist than when the older students respond to the incentives.

Finally, in Table 10, we split the sample by free school meal eligibility. Here again, we mainly find significant effects for the competitive scheme and these effects are particularly pronounced for FSM children. For non-FSM students we see that in the restricted sample there is a positive effect of the tournament, columns [4]-[6] in panel A. The positive behaviour only carries over into week six, when the incentives were removed, for consumption of healthy items. For FSM students, we see a positive effect of competition in columns [2]-[6]. In all of the consumption estimations, the positive effects carry over into week six, refer to columns [2], [3], [5], and [6] in panel B. There is no effect of piece-rate for either group.

As in the previous results we see that, in general, the tournament scheme works better than piece-rate. Indeed there are more lasting effects after the tournament incentives are removed. However, of particular interest, is that the effects on consumption – trying and

eating more than half – are more likely to persist than the effects on choice. Given that students are not more likely to choose a healthy item after the incentive is removed but are more likely to consume a healthy item, which implies that the intervention schemes have caused people who originally had healthy items as part of their meal to consume them rather than leave them on their plate. Given that people are now more likely to consume healthy items when they are given them, that could imply that a change in food preferences has occurred.

4.5. Habit formation

To evaluate whether the effects we find lead to permanent changes in habits, we contacted the schools again 6 months later and asked them to conduct an additional week of monitoring. Given that the medium run effects suggest that our intervention was more likely to have an effect on consumption, we will focus our analysis here on the consumption variables ("try" and "eat more than half"). To evaluate whether a habit has formed, we run a two-stage least squares regression, using the treatment dummies as instruments for the mean consumption during the intervention (that correspond to the mean of "try" or mean of "eat more than half" over the four weeks of intervention). Then, in the second stage, we regress the consumption variables ("try" or "eat more than half") corresponding to each of the post-intervention weeks (immediately after the intervention (week 6) and 6 months later) on the instrumented mean consumption during the intervention.

The results are presented in Table 11. We begin by presenting the estimates for all pupils. The first stage results show a positive and significant effect of the two treatment dummies on the proportion of healthy items tried during the intervention period for week 6 and on the sample where who tried less than 100% in the first week. The first stage estimates for 6 months later are typically weaker and are below 10. This is the case for all of our remaining first stage estimates except the year 5 group. To address this we have also estimated these results using limited information maximum likelihood (LIML) which is less biased when weak instruments are used. In most cases the results do not change[24]. We find a positive and significant effect of eating fruit and vegetables in the intervention period on consumption in week 6, and therefore evidence of habit formation in the immediate period

[24] There are two exceptions; when we run the estimation for the FSM children and boys in week 6 for the whole sample using LIML the estimates are of a similar magnitude but are less precisely estimated.

after the intervention, for all pupils and those who tried less than 100% in the first week. However, we do not find any evidence of habit formation when we examine consumption 6 months later. When examining various subgroups we find evidence of habit formation in week 6 for most groups. However, as with the entire sample we do not find any evidence of long run habit formation on trying.

Considering the outcome of eating more than half we find evidence of short run habit formation across our different groups, with LIML giving similar point estimates to 2SLS. For the longer term we find that the free school meals pupils in the less than 100% group in the first week developed a long run habit. Although, these results must be interpreted with caution due to the small sample size.

4.6. Learning: Food Knowledge

One question is whether the intervention triggered a response only through the incentives, or also through learning. It could be that the intervention taught children that fruit and vegetables are healthy and that they respond to that information rather than the incentives. We are able to test for this possibility by comparing the results on a knowledge test that was conducted just before and at the end of the intervention. The test shows pictures of 7 food items, including 3 or 4 fruit or vegetables and unhealthy items (such as sweets, chips, ice cream, crisps, fish fingers). Children were asked to write what the item is and whether the item was healthy or not (see Figure A2 for an example).

On average, we find that children described 92% items correctly and identified 83% of healthy items correctly before the intervention (see Table B1 for summary statistics). After the intervention, we find that results remained very similar (92% and 85% respectively). We estimate a simple linear model with the second test score results as dependent variable and include dummies for the experimental group of the school to which the child belongs (piece rate, competition or control). We also control for the first test score results. The results are presented in panel A of Table 12 for the sample of children who tried less than 100% in the first week (panel A of Table B3 presents the results for the whole sample which are very similar). We find very little difference in the second test score results across treatment and control groups. These results indicate that knowledge was very high before the intervention and was not affected by the intervention. Children are very well informed that fruit and vegetables are healthy and the responses to the intervention are not driven by learning.

4.7. Effects on cognitive outcomes

We now turn to the effects on cognitive outcomes. Cognitive ability is difficult to measure with a single test. Also, the intervention we consider is mild (it only targets one portion of fruit and vegetables a day). This is not a major shift in diet and, therefore, it is a tough test to evaluate the effects of nutrition on cognitive ability. We chose to use a partial measure of ability, capturing the ability to concentrate. Our conjecture is that concentration may be the dimension of ability that responds most quickly to changes in diet. These tests were given right after lunch on a Friday, both before and after the intervention.

The test is based on the well-known "spot-the-differences" puzzles (see Figure A3 for an example). Each puzzle consisted of two sets of 5 by 6 domino squares. There were five differences between each of the large squares that the pupils were required to find. They were given ten minutes to solve as many as they could up to a maximum of fourteen.

Table B2 presents summary statistics on the test score results. We report the fraction of completed puzzles (i.e. where 5 differences were correctly identified with a circle). On average pupils solve 40% of the puzzles correctly. There is no gender difference and no difference according to free school meal status, but Year 5 children perform better on average than Year 2 children, as we would expect. The score of the second test is also higher on average than the first test. The important question here is whether this improvement is more pronounced in the treatment schools than in the control schools.

We estimate a linear model with the fraction of correctly solved puzzles as dependent variable (see panel B of Table 12) and include treatment dummies, as well as the score of the first test. We find a positive effect (significant at the 10% level) of the piece rate on the restricted sample but we find no effects on average on the whole sample (panel B Table B3). On the other hand, if we look at the sub-groups, we find a negative effect on boys in the competition schools (again, not precisely estimated). Thus, these results do not provide convincing evidence that the intervention lead to an improvement in concentration. Of course, we should take these results with caution as we only consider a mild intervention in diet and a very partial measure of cognitive ability.

6. Conclusion

This paper provides field evidence on the effects of incentives on the formation of new dietary habits among children. We conducted a large scale field experiment in 31 primary schools in the UK testing for the effects of two different incentive schemes: a competition and a piece rate scheme. Both schemes are targeted at children's *choice* of fruit and vegetables at lunch. We implemented the rewarding schemes for a period of 4 weeks and monitored their choice and consumption over that period, as well as one week before, one week after and 6 months later.

We find two main results. First, competitive and piece rate incentives have very heterogeneous effects, particularly according to age. Younger children are more responsive to competition; older children are more responsive to piece rate incentives. Piece rate incentives work adversely on younger children (confirming some of the evidence in the literature on child nutrition). Overall, competitive incentives are more effective. This is because the piece-rate scheme contains a threshold which can create discouragement. This discouragement is not apparent in the competition scheme where the pupils, in principle, could receive an additional reward no matter how many items they had chosen. The competitive mechanism also seems to work particularly well to encourage tasting and trying fruit and vegetables among children from poorer socio-economic backgrounds. Second, we do find evidence that the intervention continues to affect behaviour after the incentives are removed. But we find that over the longer run (six months after the intervention) the changes are not sustained. One important exception relates to the children from poorer socio-economic backgrounds, for which we do find long term sustained effects. This is a notable finding given the lack of response of this group in many other school based health interventions. Finally, we do not find clear evidence of effects of the intervention on a measure of cognitive ability.

Figure 1: Average percentage of "healthy choice" (all sample)

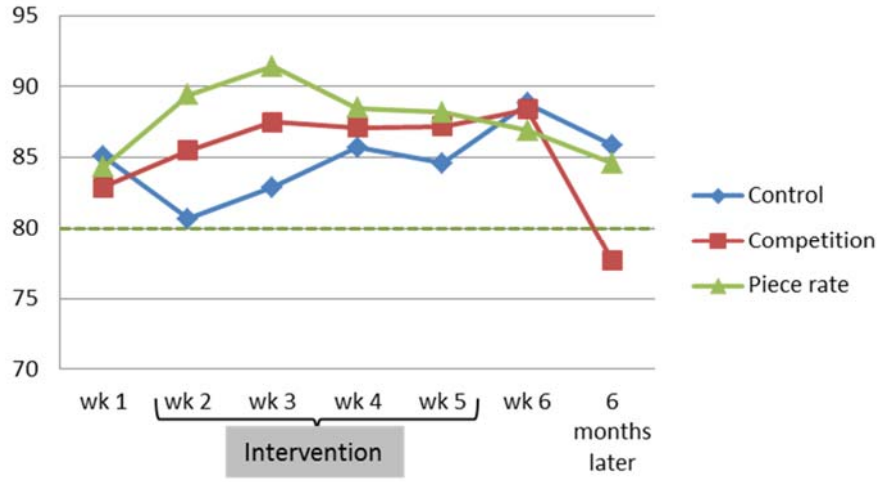


Figure 2: Average percentage of "try" (all sample)

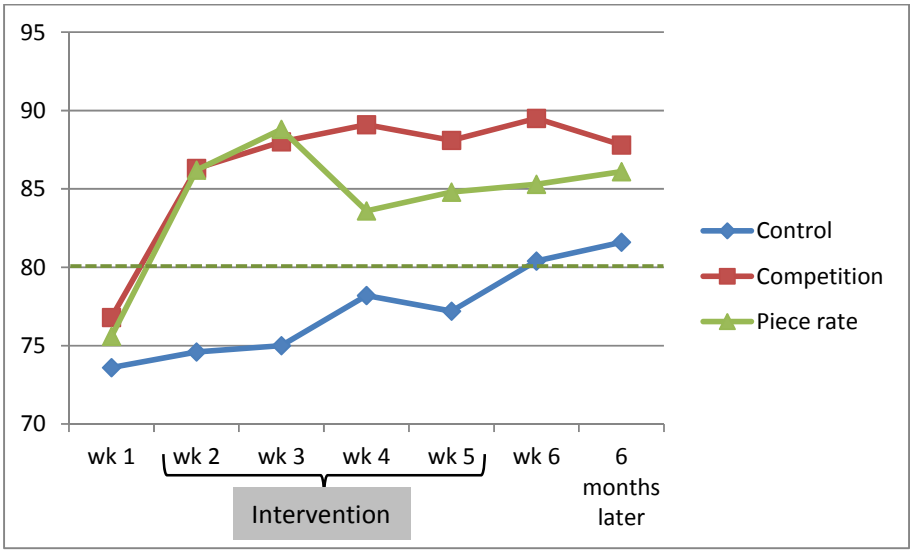


Figure 3: Average percentage of "eat more than half" (all sample)

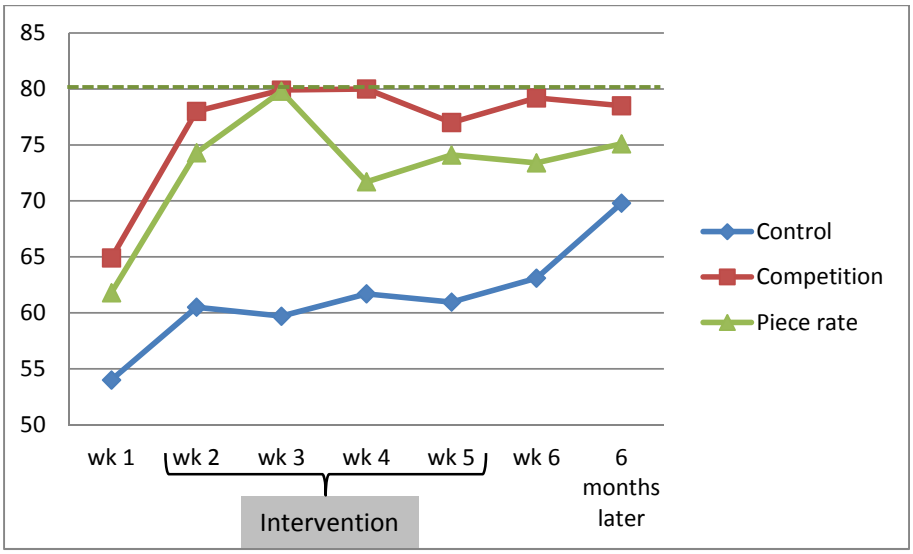


Table 1 – Summary statistics Control and Treatment groups

All	Control (C)	Piece-Rate (T1)	Competition (T2)	p-value C vs T1	p-value C vs T2
Choice	0.842	0.847	0.821	0.942	0.733
Ate something	0.738	0.78	0.767	0.507	0.657
Ate more than half	0.551	0.627	0.654	0.237	0.283
School Dinner	0.505	0.453	0.503	0.624	0.983
Packed Lunch	0.495	0.547	0.497	0.624	0.983
Female	0.52	0.447	0.552	0.202	0.521
1st Language English	0.979	0.982	0.932	0.872	0.228
Free School Meal	0.205	0.196	0.156	0.885	0.416
School meals per week	2.48	2.435	2.558	0.939	0.873
Always School Meal	0.44	0.433	0.444	0.955	0.973
Packed lunches per week	2.525	2.565	2.442	0.945	0.866
Always Packed Lunch	0.395	0.446	0.394	0.684	0.992
White British	0.903	0.9	0.785	0.963	0.221
Special dietary requirements	0.053	0.091	0.125	0.176	0.15
Specific health cond.	0.144	0.172	0.152	0.431	0.784
Normally eats breakfast	0.974	0.969	0.956	0.794	0.345

Notes: All variables are evaluated for the first week, before the start of the treatment. The first column shows the means for the pupils in the control school in the, the second column for schools in the piece-rate scheme and the third column in the competition schools. The fourth and fifth columns show the p-value difference in the means of each treatment compared to the control group. The p-value were calculated, to account for intra-school correlation, by regressing each baseline variable on one of the treatment indicators, and clustering the standard errors at the school level, the p-value is matches to the t-statistic on the treatment dummy.

Table 3 – Average Treatment Effects – All Sample and Less than 100%

	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	-0.022	0.032	0.110	-0.068	-0.032	0.014
	(0.058)	(0.065)	(0.098)	(0.074)	(0.050)	(0.072)
Piece rate	-0.009	0.020	0.078	-0.045	-0.061	0.016
	(0.075)	(0.072)	(0.072)	(0.059)	(0.073)	(0.060)
Comp x Week 2	0.059*	0.085**	0.080	0.196***	0.195***	0.153
	(0.031)	(0.035)	(0.081)	(0.060)	(0.037)	(0.103)
Comp x Week 3	0.070	0.095**	0.083	0.155**	0.182***	0.172*
	(0.047)	(0.045)	(0.079)	(0.077)	(0.048)	(0.090)
Comp x Week 4	0.025	0.073	0.083	0.109**	0.170***	0.190**
	(0.033)	(0.048)	(0.083)	(0.049)	(0.049)	(0.088)
Comp x Week 5	0.041	0.075*	0.063	0.168***	0.178***	0.149*
	(0.041)	(0.044)	(0.082)	(0.062)	(0.046)	(0.084)
Piece x Week 2	0.0710	0.0921	0.0720	0.158***	0.203**	0.143*
	(0.0473)	(0.0610)	(0.0679)	(0.0591)	(0.0905)	(0.0779)
Piece x Week 3	0.0797	0.102	0.0914	0.128	0.203**	0.166*
	(0.0576)	(0.0635)	(0.0739)	(0.0781)	(0.0921)	(0.0865)
Piece x Week 4	0.0170	0.0294	0.0324	0.0789	0.119	0.109
	(0.0698)	(0.0739)	(0.0849)	(0.105)	(0.105)	(0.0954)
Piece x Week 5	0.0202	0.0526	0.0611	0.0422	0.136	0.123
	(0.0605)	(0.0752)	(0.0829)	(0.0969)	(0.110)	(0.0923)
Constant	0.851***	0.736***	0.540***	0.546***	0.531***	0.399***
	(0.047)	(0.038)	(0.048)	(0.046)	(0.029)	(0.028)
N obs.	2,768	2,679	2,679	971	1,339	1,787
N pupils	626	609	609	203	304	410

Note: OLS regression results; including pupil random effects and with standard errors clustered at the school level. Additional controls: week dummies. Dependent variable is the pupil weekly mean.

*, ** and *** indicate significance at 10%, 5% and 1% level respectively.

Table 4: Dynamics of choice and consumption during treatment weeks.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: Choice								
Competition	-0.097	-0.155**	-0.184	-0.083	-0.106	0.088	0.045	-0.008
	[0.079]	[0.071]	[0.149]	[0.113]	[0.123]	[0.100]	[0.087]	[0.089]
Piece-Rate	-0.031	0.011	0.074	-0.017	-0.050	-0.174***	0.017	-0.144***
	[0.063]	[0.121]	[0.160]	[0.105]	[0.089]	[0.067]	[0.102]	[0.036]
Comp x Week 2-5	0.166***	0.199***	0.237*	0.179*	0.121	0.037	0.007	0.067
	[0.059]	[0.045]	[0.131]	[0.098]	[0.100]	[0.096]	[0.065]	[0.077]
Piece x Week 2-5	0.097	0.028	0.050	0.071	0.061	0.254**	-0.061	0.176***
	[0.079]	[0.068]	[0.131]	[0.102]	[0.125]	[0.111]	[0.183]	[0.051]
Constant	0.520***	0.499***	0.604***	0.618***	0.612***	0.440***	0.347***	0.607***
	[0.041]	[0.022]	[0.140]	[0.087]	[0.026]	[0.031]	[0.074]	[0.045]
Observations	4,743	910	977	951	975	930	841	1,970
Number of pupil_id	215	212	214	215	213	213	154	202
Panel B: Try								
Competition	-0.034	-0.010	-0.118*	0.018	-0.019	-0.042	0.016	0.002
	[0.047]	[0.098]	[0.071]	[0.081]	[0.101]	[0.100]	[0.061]	[0.059]
Piece-Rate	-0.003	0.115	-0.050	0.013	-0.164	0.042	-0.049	-0.041
	[0.060]	[0.116]	[0.098]	[0.068]	[0.105]	[0.142]	[0.113]	[0.046]
Comp x Week 2-5	0.187***	0.161*	0.268***	0.159*	0.125	0.235***	0.120**	0.126***
	[0.033]	[0.091]	[0.076]	[0.089]	[0.089]	[0.070]	[0.053]	[0.046]
Piece x Week 2-5	0.120	-0.022	0.160	0.109	0.250*	0.081	0.043	0.133**
	[0.079]	[0.101]	[0.108]	[0.076]	[0.147]	[0.131]	[0.190]	[0.062]
Constant	0.517***	0.453***	0.610***	0.560***	0.559***	0.532***	0.323***	0.620***
	[0.032]	[0.082]	[0.057]	[0.049]	[0.037]	[0.076]	[0.055]	[0.040]
Observations	6,296	1,203	1,303	1,290	1,299	1,201	871	2,866
Number of pupil_id	333	315	328	318	315	315	175	305
Sample Used	All	All	All	All	All	All	Missed	Not Missed
Days of Week Controls	Yes	No	No	No	No	No	Yes	Yes
Week Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 5A – Gender treatment effects (BOYS) – All and Less than 100%

	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	0.001	0.080	0.102	-0.117	-0.029	-0.009
	(0.078)	(0.068)	(0.093)	(0.107)	(0.058)	(0.072)
Piece rate	0.013	0.017	0.042	-0.096	-0.069	-0.048
	(0.093)	(0.080)	(0.075)	(0.108)	(0.086)	(0.064)
Comp x Week 2	0.034	0.037	0.071	0.224***	0.159***	0.144*
	(0.050)	(0.050)	(0.071)	(0.079)	(0.057)	(0.082)
Comp x Week 3	0.054	0.046	0.0533	0.181*	0.135**	0.130*
	(0.066)	(0.048)	(0.058)	(0.096)	(0.060)	(0.072)
Comp x Week 4	-0.002	0.020	0.0540	0.108*	0.154***	0.167*
	(0.045)	(0.056)	(0.083)	(0.064)	(0.055)	(0.085)
Comp x Week 5	0.028	0.029	0.052	0.219***	0.167**	0.114
	(0.055)	(0.060)	(0.080)	(0.083)	(0.073)	(0.088)
Piece x Week 2	0.070	0.149**	0.138*	0.209**	0.275***	0.226***
	(0.060)	(0.071)	(0.072)	(0.090)	(0.098)	(0.083)
Piece x Week 3	0.084	0.153**	0.131*	0.188*	0.289***	0.229***
	(0.070)	(0.069)	(0.069)	(0.101)	(0.088)	(0.073)
Piece x Week 4	-0.004	0.033	0.040	0.095	0.135	0.110
	(0.078)	(0.083)	(0.085)	(0.126)	(0.099)	(0.086)
Piece x Week 5	0.018	0.058	0.032	0.101	0.154	0.078
	(0.067)	(0.088)	(0.080)	(0.104)	(0.117)	(0.090)
Constant	0.818***	0.711***	0.544***	0.548***	0.532***	0.419***
	(0.063)	(0.050)	(0.047)	(0.069)	(0.044)	(0.029)
N obs.	1,302	1,236	1,236	496	660	850
N pupils	299	286	286	103	148	196

Note: OLS regression results; including pupil random effects and with standard errors clustered at the school level. Additional controls: week dummies. Dependent variable is the pupil weekly mean.

*, ** and *** indicate significance at 10%, 5% and 1% level respectively.

Table 5B – Gender treatment effects (GIRLS) – All and Less than 100%

	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	-0.034	-0.006	0.117	-0.017	-0.033	0.038
	(0.053)	(0.073)	(0.112)	(0.078)	(0.065)	(0.083)
Piece rate	-0.017	0.039	0.139	0.029	-0.054	0.109
	(0.063)	(0.079)	(0.089)	(0.065)	(0.112)	(0.077)
Comp x Week 2	0.078*	0.108***	0.071	0.172**	0.188***	0.135
	(0.043)	(0.041)	(0.100)	(0.079)	(0.049)	(0.133)
Comp x Week 3	0.078	0.120*	0.088	0.095	0.181**	0.178
	(0.052)	(0.064)	(0.103)	(0.093)	(0.071)	(0.121)
Comp x Week 4	0.036	0.103*	0.090	0.070	0.152***	0.187*
	(0.043)	(0.053)	(0.090)	(0.060)	(0.057)	(0.099)
Comp x Week 5	0.038	0.096*	0.056	0.071	0.150**	0.159
	(0.061)	(0.054)	(0.098)	(0.106)	(0.067)	(0.111)
Piece x Week 2	0.067	0.031	-0.003	0.072	0.126	0.052
	(0.045)	(0.065)	(0.076)	(0.084)	(0.125)	(0.093)
Piece x Week 3	0.071	0.050	0.036	0.019	0.111	0.085
	(0.056)	(0.080)	(0.095)	(0.086)	(0.157)	(0.129)
Piece x Week 4	0.024	0.013	0.006	0.009	0.099	0.094
	(0.065)	(0.077)	(0.100)	(0.105)	(0.146)	(0.124)
Piece x Week 5	0.0088	0.032	0.064	-0.088	0.105	0.141
	(0.064)	(0.082)	(0.102)	(0.118)	(0.144)	(0.123)
Constant	0.883***	0.760***	0.541***	0.542***	0.528***	0.384***
	(0.039)	(0.037)	(0.063)	(0.047)	(0.033)	(0.035)
N obs.	1,410	1,389	1,389	439	638	896
N pupils	315	311	311	92	147	205

Note: OLS regression results; including pupil random effects and with standard errors clustered at the school level. Additional controls: week dummies. Dependent variable is the pupil weekly mean.

*, ** and *** indicate significance at 10%, 5% and 1% level respectively.

Table 6A – Treatment effects by year group (YEAR 2): All and Less than 100%

	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	-0.090	0.007	0.129	-0.169*	-0.017	0.042
	(0.060)	(0.058)	(0.108)	(0.087)	(0.071)	(0.098)
Piece rate	0.014	0.065	0.144**	-0.078	0.043	0.093
	(0.056)	(0.057)	(0.0710)	(0.074)	(0.077)	(0.0638)
Comp x Week 2	0.089**	0.085**	0.018	0.189**	0.156***	0.076
	(0.045)	(0.040)	(0.117)	(0.080)	(0.051)	(0.167)
Comp x Week 3	0.067	0.065	0.046	0.129	0.091	0.130
	(0.057)	(0.050)	(0.124)	(0.093)	(0.069)	(0.159)
Comp x Week 4	0.014	0.041	0.064	0.034	0.064	0.156
	(0.046)	(0.046)	(0.124)	(0.084)	(0.059)	(0.145)
Comp x Week 5	0.076	0.101*	0.046	0.141	0.146**	0.103
	(0.048)	(0.052)	(0.123)	(0.088)	(0.066)	(0.147)
Piece x Week 2	-2.65e-06	-0.034	-0.087	-0.064	-0.070*	-0.062
	(0.033)	(0.024)	(0.056)	(0.096)	(0.042)	(0.064)
Piece x Week 3	-0.009	-0.016	-0.032	-0.120	-0.057	-0.012
	(0.036)	(0.042)	(0.073)	(0.119)	(0.058)	(0.077)
Piece x Week 4	-0.115**	-0.124**	-0.116	-0.358**	-0.208***	-0.111
	(0.055)	(0.059)	(0.093)	(0.151)	(0.079)	(0.093)
Piece x Week 5	-0.059	-0.078	-0.060	-0.271**	-0.180*	-0.067
	(0.048)	(0.061)	(0.098)	(0.130)	(0.095)	(0.101)
Constant	0.900***	0.771***	0.518***	0.598***	0.539***	0.381***
	(0.039)	(0.038)	(0.053)	(0.051)	(0.054)	(0.036)
N obs.	1,439	1,426	1,426	416	643	949
N pupils	340	337	337	90	154	228

Note: OLS regression results; including pupil random effects and with standard errors clustered at the school level. Additional controls: week dummies. Dependent variable is the pupil weekly mean.

*, ** and *** indicate significance at 10%, 5% and 1% level respectively.

Table 6B – Treatment effects by year group (YEAR 5): All and Less than 100%

	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	0.055	0.054	0.096	0.039	-0.063	-0.014
	(0.0672)	(0.103)	(0.119)	(0.0732)	(0.072)	(0.073)
Piece rate	-0.0276	-0.044	-0.001	-0.018	-0.158	-0.076
	(0.104)	(0.118)	(0.119)	(0.0760)	(0.107)	(0.100)
Comp x Week 2	0.017	0.080	0.131*	0.176**	0.220***	0.221***
	(0.0641)	(0.066)	(0.0690)	(0.069)	(0.073)	(0.0698)
Comp x Week 3	0.072	0.127*	0.113	0.150*	0.268***	0.207***
	(0.064)	(0.073)	(0.0751)	(0.088)	(0.058)	(0.0749)
Comp x Week 4	0.028	0.105	0.0959	0.147**	0.265***	0.218**
	(0.057)	(0.079)	(0.0890)	(0.071)	(0.069)	(0.096)
Comp x Week 5	-0.004	0.045	0.0743	0.173**	0.197***	0.189**
	(0.079)	(0.071)	(0.0799)	(0.077)	(0.066)	(0.078)
Piece x Week 2	0.139**	0.247***	0.255***	0.254***	0.442***	0.380***
	(0.066)	(0.095)	(0.097)	(0.058)	(0.103)	(0.102)
Piece x Week 3	0.169**	0.251**	0.244**	0.243**	0.432***	0.382***
	(0.083)	(0.108)	(0.117)	(0.097)	(0.120)	(0.120)
Piece x Week 4	0.144*	0.219**	0.222**	0.267***	0.403***	0.378***
	(0.087)	(0.088)	(0.099)	(0.050)	(0.080)	(0.0954)
Piece x Week 5	0.097	0.224***	0.220**	0.176**	0.414***	0.354***
	(0.077)	(0.085)	(0.090)	(0.078)	(0.071)	(0.088)
Constant	0.799***	0.699***	0.563***	0.512***	0.524***	0.419***
	(0.0502)	(0.0416)	(0.0600)	(0.0557)	(0.0329)	(0.0447)
N obs.	1,324	1,248	1,248	555	691	833
N pupils	285	271	271	113	149	181

Note: OLS regression results; including pupil random effects and with standard errors clustered at the school level. Additional controls: week dummies. Dependent variable is the pupil weekly mean.

*, ** and *** indicate significance at 10%, 5% and 1% level respectively.

Table 7A – Treatment effects by free school meal status (NO FSM): All and Less than 100%

	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	0.000	0.071	0.130	-0.040	-0.012	0.031
	(0.057)	(0.062)	(0.104)	(0.079)	(0.057)	(0.079)
Piece rate	-0.004	0.067	0.135*	-0.013	-0.020	0.049
	(0.079)	(0.062)	(0.0751)	(0.050)	(0.059)	(0.058)
Comp x Week 2	0.057*	0.063**	0.090	0.193***	0.189***	0.158
	(0.034)	(0.031)	(0.077)	(0.054)	(0.031)	(0.098)
Comp x Week 3	0.066	0.062	0.067	0.129*	0.143**	0.152
	(0.051)	(0.046)	(0.083)	(0.078)	(0.060)	(0.093)
Comp x Week 4	0.023	0.048	0.074	0.090*	0.143***	0.167*
	(0.034)	(0.047)	(0.086)	(0.055)	(0.053)	(0.093)
Comp x Week 5	0.035	0.051	0.066	0.157**	0.158***	0.136
	(0.042)	(0.042)	(0.079)	(0.066)	(0.048)	(0.084)
Piece x Week 2	0.075	0.082	0.058	0.138**	0.220***	0.145**
	(0.055)	(0.055)	(0.064)	(0.069)	(0.079)	(0.074)
Piece x Week 3	0.090	0.081	0.066	0.116	0.174**	0.151*
	(0.064)	(0.054)	(0.070)	(0.079)	(0.081)	(0.084)
Piece x Week 4	0.028	0.012	0.022	0.083	0.0881	0.103
	(0.077)	(0.068)	(0.083)	(0.095)	(0.098)	(0.095)
Piece x Week 5	0.026	0.044	0.044	0.038	0.113	0.108
	(0.066)	(0.069)	(0.079)	(0.085)	(0.109)	(0.089)
Constant	0.837***	0.710***	0.528***	0.533***	0.514***	0.391***
	(0.045)	(0.036)	(0.056)	(0.051)	(0.028)	(0.034)
N obs.	2,206	2,149	2,149	796	1,059	1,397
N pupils	498	487	487	168	242	322

Note: OLS regression results; including pupil random effects and with standard errors clustered at the school level. Additional controls: week dummies. Dependent variable is the pupil weekly mean.

*, ** and *** indicate significance at 10%, 5% and 1% level respectively.

Table 7B – Treatment effects by free school meal status (FSM): All and Less than 100%

	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	-0.112	-0.124	0.030	-0.204	-0.123	-0.033
	(0.097)	(0.106)	(0.133)	(0.151)	(0.102)	(0.099)
Piece rate	-0.005	-0.135	-0.096	-0.245*	-0.295*	-0.049
	(0.099)	(0.156)	(0.142)	(0.130)	(0.151)	(0.132)
Comp x Week 2	0.106	0.201**	0.056	0.277*	0.193*	0.117
	(0.066)	(0.086)	(0.113)	(0.165)	(0.109)	(0.160)
Comp x Week 3	0.101	0.231***	0.145*	0.313**	0.337***	0.235**
	(0.064)	(0.074)	(0.0842)	(0.146)	(0.125)	(0.118)
Comp x Week 4	0.055	0.193***	0.142	0.207	0.262***	0.279**
	(0.069)	(0.071)	(0.101)	(0.135)	(0.0669)	(0.113)
Comp x Week 5	0.065	0.184***	0.056	0.204	0.233***	0.190
	(0.061)	(0.068)	(0.107)	(0.131)	(0.0891)	(0.129)
Piece x Week 2	0.052	0.163	0.153	0.293*	0.252	0.182
	(0.051)	(0.136)	(0.120)	(0.175)	(0.201)	(0.117)
Piece x Week 3	0.020	0.199	0.182	0.119	0.432**	0.224
	(0.039)	(0.136)	(0.134)	(0.110)	(0.213)	(0.142)
Piece x Week 4	-0.051	0.083	0.066	-0.118	0.292	0.141
	(0.061)	(0.157)	(0.163)	(0.206)	(0.241)	(0.176)
Piece x Week 5	-0.046	0.058	0.085	-0.0979	0.236	0.136
	(0.069)	(0.154)	(0.149)	(0.346)	(0.234)	(0.161)
Constant	0.903***	0.829***	0.588***	0.595***	0.608***	0.426***
	(0.071)	(0.072)	(0.094)	(0.129)	(0.0688)	(0.0666)
N obs.	487	462	462	140	230	335
N pupils	113	108	108	28	52	77

Note: OLS regression results; including pupil random effects and with standard errors clustered at the school level. Additional controls: week dummies. Dependent variable is the pupil weekly mean.

*, ** and *** indicate significance at 10%, 5% and 1% level respectively.

Table 8 – Medium run effects (i.e. including week 6) by GENDER

BOYS	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	0.001	0.080	0.080	-0.117	-0.029	0.020
	(0.077)	(0.067)	(0.067)	(0.106)	(0.057)	(0.072)
Piece rate	0.013	0.017	0.017	-0.096	-0.069	-0.045
	(0.093)	(0.080)	(0.080)	(0.107)	(0.086)	(0.098)
Comp x Wk2-5	0.029	0.029	0.029	0.185**	0.154***	0.084
	(0.049)	(0.049)	(0.049)	(0.072)	(0.054)	(0.059)
Comp x Week 6	-0.040	0.003	0.003	0.047	0.135*	0.043
	(0.032)	(0.059)	(0.059)	(0.060)	(0.079)	(0.075)
Piece x Wk2-5	0.041	0.098	0.098	0.149*	0.213**	0.161*
	(0.063)	(0.073)	(0.073)	(0.086)	(0.091)	(0.091)
Piece x Week 6	-0.062	0.033	0.033	-0.054	0.118	0.073
	(0.062)	(0.078)	(0.078)	(0.115)	(0.102)	(0.099)
N obs.	1,565	1,490	1,490	597	794	1,025
N pupils	299	286	286	103	148	196

GIRLS	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	-0.034	-0.006	-0.006	-0.017	-0.033	-0.094
	(0.052)	(0.073)	(0.073)	(0.077)	(0.064)	(0.072)
Piece rate	-0.017	0.039	0.039	0.029	-0.054	0.004
	(0.063)	(0.079)	(0.079)	(0.064)	(0.111)	(0.103)
Comp x Wk2-5	0.060	0.109**	0.109**	0.106	0.170***	0.202***
	(0.045)	(0.047)	(0.047)	(0.076)	(0.048)	(0.053)
Comp x Week 6	0.033	0.077	0.077	0.056	0.108*	0.158***
	(0.055)	(0.048)	(0.048)	(0.068)	(0.062)	(0.047)
Piece x Wk2-5	0.043	0.032	0.032	0.004	0.110	0.075
	(0.053)	(0.072)	(0.072)	(0.087)	(0.135)	(0.103)
Piece x Week 6	-0.009	0.007	0.007	-0.095	0.070	0.043
	(0.067)	(0.090)	(0.090)	(0.098)	(0.176)	(0.127)
N obs.	1,698	1,675	1,675	531	769	1,079
N pupils	315	311	311	92	147	205

Table 9 – Medium run effects (i.e. including week 6) by YEAR GROUP

YEAR 2	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	-0.089	0.007	0.007	-0.169**	-0.017	-0.069
	(0.059)	(0.058)	(0.058)	(0.086)	(0.071)	(0.063)
Piece rate	0.014	0.064	0.064	-0.078	0.043	0.039
	(0.055)	(0.057)	(0.057)	(0.073)	(0.076)	(0.078)
Comp x Wk2-5	0.063	0.071*	0.071*	0.127	0.113**	0.138***
	(0.045)	(0.040)	(0.040)	(0.080)	(0.052)	(0.043)
Comp x Week 6	0.046	0.071*	0.071*	0.067	0.144**	0.146***
	(0.040)	(0.041)	(0.041)	(0.067)	(0.061)	(0.050)
Piece x Wk2-5	-0.046	-0.063*	-0.063*	-0.202***	-0.129***	-0.069**
	(0.034)	(0.035)	(0.035)	(0.073)	(0.045)	(0.035)
Piece x Week 6	-0.107**	-0.097**	-0.097**	-0.439***	-0.162**	-0.107**
	(0.043)	(0.042)	(0.042)	(0.075)	(0.078)	(0.047)
N obs.	1,743	1,727	1,727	505	779	1,151
N pupils	340	337	337	90	154	228

YEAR 5	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	0.055	0.054	0.054	0.039	-0.063	-0.031
	(0.067)	(0.102)	(0.102)	(0.073)	(0.072)	(0.107)
Piece rate	-0.028	-0.044	-0.044	-0.018	-0.158	-0.118
	(0.104)	(0.118)	(0.118)	(0.075)	(0.107)	(0.135)
Comp x Wk2-5	0.028	0.090	0.090	0.162**	0.238***	0.186**
	(0.060)	(0.067)	(0.067)	(0.064)	(0.054)	(0.082)
Comp x Week 6	-0.044	0.030	0.030	0.061	0.134	0.087
	(0.073)	(0.079)	(0.079)	(0.071)	(0.082)	(0.090)
Piece x Wk2-5	0.137*	0.235***	0.235***	0.236***	0.423***	0.356***
	(0.074)	(0.090)	(0.090)	(0.061)	(0.085)	(0.106)
Piece x Week 6	0.055	0.190*	0.190*	0.120*	0.336***	0.284**
	(0.084)	(0.108)	(0.108)	(0.071)	(0.109)	(0.124)
N obs.	1,582	1,498	1,498	667	828	997
N pupils	285	271	271	113	149	181

Table 10 – Medium run effects (i.e. including week 6) by FREE SCHOOL MEAL STATUS

NO FSM	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	0.000	0.071	0.071	-0.039	-0.012	0.003
	(0.057)	(0.062)	(0.062)	(0.079)	(0.056)	(0.064)
Piece rate	-0.003	0.067	0.067	-0.013	-0.020	0.003
	(0.078)	(0.062)	(0.062)	(0.060)	(0.059)	(0.075)
Comp x Wk2-5	0.048	0.056	0.056	0.148***	0.161***	0.122***
	(0.035)	(0.036)	(0.036)	(0.054)	(0.036)	(0.045)
Comp x Week 6	0.005	0.025	0.025	0.068	0.112**	0.073*
	(0.038)	(0.043)	(0.043)	(0.042)	(0.056)	(0.044)
Piece x Wk2-5	0.054	0.055	0.055	0.094	0.149*	0.119
	(0.061)	(0.056)	(0.056)	(0.067)	(0.082)	(0.077)
Piece x Week 6	-0.009	0.024	0.024	-0.029	0.085	0.069
	(0.067)	(0.065)	(0.065)	(0.089)	(0.104)	(0.089)
N obs.	2,661	2,597	2,597	962	1,277	1,687
N pupils	498	487	487	168	242	322

FSM	ALL			LESS THAN 100% IN WEEK 1		
	Chooses	Tries	Eats more than half	Chooses	Tries	Eats more than half
Competition	-0.112	-0.124	-0.124	-0.204	-0.123	-0.230*
	(0.096)	(0.105)	(0.105)	(0.147)	(0.100)	(0.119)
Piece rate	-0.005	-0.135	-0.135	-0.245*	-0.295**	-0.137
	(0.099)	(0.155)	(0.155)	(0.126)	(0.149)	(0.178)
Comp x Wk2-5	0.081	0.202***	0.202***	0.250*	0.255***	0.310***
	(0.061)	(0.070)	(0.070)	(0.132)	(0.089)	(0.087)
Comp x Week 6	0.018	0.174**	0.174**	0.086	0.226**	0.295***
	(0.091)	(0.084)	(0.084)	(0.159)	(0.106)	(0.105)
Piece x Wk2-5	-0.005	0.126	0.126	0.049	0.304	0.158
	(0.046)	(0.138)	(0.138)	(0.199)	(0.211)	(0.160)
Piece x Week 6	-0.140	0.024	0.024	-0.314	0.192	0.061
	(0.107)	(0.167)	(0.167)	(0.371)	(0.247)	(0.186)
N obs.	581	552	552	168	276	401
N pupils	113	108	108	28	52	77

Table 11: habit Formation (IV Results)

	Try				Eats more than half			
	Week 6		6 months later		Week 6		6 months later	
	All	Less than 100%	All	Less than 100%	All	Less than 100%	All	Less than 100%
All sample	0.646**	0.637**	0.085	-0.056	0.976***	0.887***	0.1	0.095
	(0.262)	(0.268)	(0.525)	(0.546)	(0.219)	(0.221)	(0.460)	(0.398)
1st stage F-stat	10.22	12.08	7.892	5.623	5.11	5.6	4.986	6.196
Boys	0.494*	0.610***	-0.092	-0.065	0.810***	0.594***	0.484	0.559
	[0.281]	[0.225]	[0.795]	[0.576]	[0.192]	[0.207]	[0.671]	[0.416]
1st stage F-stat	3.845	4.987	2.183	3.464	5.188	6.164	1.87	4.093
Girls	0.689**	0.619*	0.028	-0.125	0.990***	1.037***	-0.228	-0.248
	[0.295]	[0.353]	[0.401]	[0.602]	[0.288]	[0.304]	[0.477]	[0.402]
1st stage F-stat	8.239	6.436	7.387	3.849	3.522	3.642	5.881	4.569
Year 2	1.354***	1.358***	1.369	1.572*	1.427***	1.452***	0.583	1.947*
	[0.295]	[0.249]	[1.213]	[0.926]	[0.345]	[0.406]	[0.689]	[1.140]
1st stage F-stat	2.988	4.531	0.588	0.222	1.139	1.045	0.742	0.813
Year 5	0.723***	0.705***	-0.044	-0.13	0.732***	0.696***	0.144	-0.044
	[0.197]	[0.149]	[0.555]	[0.478]	[0.216]	[0.206]	[0.438]	[0.367]
1st stage F-stat	26.19	34.48	19.09	27.19	14.71	17.82	8.748	15.01
No FSM	0.654**	0.609**	-0.232	-0.212	0.923***	0.822***	-0.135	-0.21
	[0.289]	[0.306]	[0.579]	[0.655]	[0.261]	[0.265]	[0.509]	[0.446]
1st stage F-stat	6.46	8.813	4.621	3.33	5.377	5.193	3.942	3.798
FSM	0.960**	0.748	0.959	0.324	1.352***	1.179***	0.471	0.868***
	[0.474]	[0.486]	[0.604]	[0.479]	[0.457]	[0.341]	[0.539]	[0.250]
1st stage F-stat	3.421	3.919	4.714	5.019	1.121	3.055	3.643	9.227

Table 12: Food knowledge & Spot the difference tests - difference across treatment and control groups - OLS estimates (Tried less than 100% in the first week sample)

	All	Boys	Girls	Year 2	Year 5	FSM	No FSM
Panel A: Food Knowledge test 2 score							
Competition	-0.032	-0.002	-0.057	-0.082	0.011	-0.095	-0.017
	[0.029]	[0.050]	[0.040]	[0.053]	[0.034]	[0.078]	[0.034]
Piece rate	0.019	0.062	-0.039	-0.033	0.057	-0.118	0.045
	[0.033]	[0.049]	[0.074]	[0.078]	[0.051]	[0.244]	[0.038]
Test 1	-0.017	-0.077	0.065	-0.048	0.013	0.284	-0.1
	[0.083]	[0.107]	[0.134]	[0.126]	[0.119]	[0.196]	[0.090]
Constant	0.910***	0.936***	0.864***	0.890***	0.852***	0.709***	0.960***
	[0.067]	[0.093]	[0.120]	[0.112]	[0.092]	[0.159]	[0.082]
Observations	142	66	75	64	77	21	120
R-squared	0.072	0.094	0.082	0.046	0.03	0.259	0.074
Panel B: Spot the difference test 2 score							
Competition	-0.048	-0.085*	-0.017	0.009	-0.061	0.022	-0.062
	[0.035]	[0.047]	[0.054]	[0.048]	[0.050]	[0.086]	[0.039]
Piece rate	0.076*	0.088	0.065	0.088	0.093	0.182	0.066
	[0.043]	[0.056]	[0.070]	[0.065]	[0.058]	[0.116]	[0.047]
Test 1	0.733***	0.782***	0.691***	0.901***	0.618***	0.751***	0.718***
	[0.060]	[0.080]	[0.092]	[0.087]	[0.092]	[0.142]	[0.068]
Constant	0.302***	0.296***	0.344***	0.138***	0.374***	0.229	0.306***
	[0.053]	[0.072]	[0.068]	[0.048]	[0.071]	[0.160]	[0.058]
Observations	202	103	99	103	99	32	170
R-squared	0.599	0.646	0.565	0.534	0.38	0.709	0.572

APPENDIX A – EXPERIMENTAL MATERIALS

1 – Stickers and rewards

2 – Example of food knowledge test

3 – Example of spot the differences test

1. **STICKERS**

CONTENTS OF EACH REWARD BOX

16 highlighters

10 animal pencils

11 animal yoyos

12 animal erasers

18 bookmarks

6 pencils on the move

12 wooden boomarks

12 pencil sharpeners

6 sea pencils

2. Food knowledge test - example

Please write for each item "what it is" and tick **"yes**" if you think it is healthy and **"no"** if you think it is <u>not</u> healthy. [max 10 minutes]

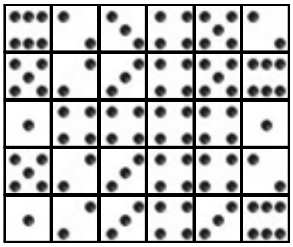
	What is it?	Is it healthy?
		Yes ☐ No ☐
		Yes ☐ No ☐
		Yes ☐ No ☐
		Yes ☐ No ☐
		Yes ☐ No ☐
		Yes ☐ No ☐
		Yes ☐ No ☐

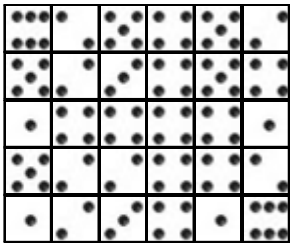
3. Example of spot the differences test

Name: _____

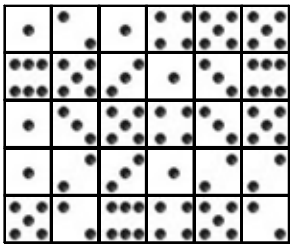
Complete as many puzzles as you can.

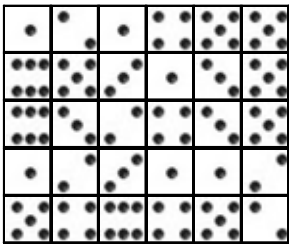
1) Spot the 5 differences and circle them



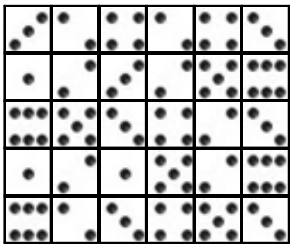


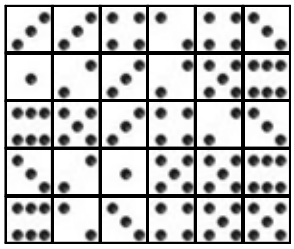
2) Spot the 5 differences and circle them





3) Spot the 5 differences and circle them





Appendix B: Additional Tables (not for publication)

Table B1: Summary statistics – Food knowledge test – Correct answers

	Test 1			Test 2		
All	N	Mean	SD	N	Mean	SD
"What is it"	460	0.922	0.14	414	0.92	0.146
"Is it healthy"	462	0.827	0.177	409	0.854	0.157
Boys						
"What is it"	219	0.911	0.155	189	0.915	0.154
"Is it healthy"	221	0.805	0.19	186	0.838	0.165
Girls						
"What is it"	235	0.933	0.124	217	0.93	0.135
"Is it healthy"	234	0.85	0.158	215	0.869	0.149
Year 2						
"What is it"	244	0.9	0.152	233	0.898	0.159
"Is it healthy"	245	0.802	0.193	227	0.834	0.17
Year 5						
"What is it"	216	0.947	0.12	181	0.95	0.12
"Is it healthy"	217	0.855	0.154	182	0.878	0.136
FSM						
"What is it"	73	0.902	0.131	69	0.907	0.139
"Is it healthy"	75	0.765	0.197	67	0.826	0.181
No FSM						
"What is it"	377	0.926	0.142	336	0.926	0.145
"Is it healthy"	376	0.841	0.168	333	0.859	0.152

Table B2 – Summary statistics – "Spot the Differences" test – Correct answers

	Test 1			Test 2		
	N	Mean	SD	N	Mean	SD
All	545	0.393	0.326	494	0.520	0.367
Boys	262	0.386	0.324	230	0.494	0.375
Girls	274	0.404	0.327	255	0.546	0.357
Year 2	294	0.241	0.276	273	0.364	0.333
Year 5	251	0.570	0.288	221	0.712	0.313
FSM	94	0.378	0.350	81	0.437	0.375
No FSM	440	0.394	0.319	402	0.537	0.363

Table B3: Food knowledge & Spot the difference tests - difference across treatment and control groups - OLS estimates (All Pupils)

	All	Boys	Girls	Year 2	Year 5	FSM	No FSM
Panel A: Food Knowledge test 2 score							
Competition	-0.008	0.002	-0.013	-0.039	0.036	-0.068	0.008
	[0.023]	[0.032]	[0.026]	[0.029]	[0.028]	[0.047]	[0.022]
Piece rate	-0.001	0.018	-0.019	-0.015	0.027	-0.043	0.017
	[0.022]	[0.038]	[0.042]	[0.041]	[0.041]	[0.104]	[0.029]
Test 1	0.077	0.056	0.072	0.05	0.108	0.215	0.026
	[0.055]	[0.078]	[0.082]	[0.074]	[0.093]	[0.132]	[0.060]
Constant	0.815***	0.823***	0.829***	0.809***	0.752***	0.702***	0.843***
	[0.043]	[0.069]	[0.077]	[0.062]	[0.079]	[0.105]	[0.056]
Observations	292	136	155	160	132	45	247
R-squared	0.032	0.063	0.028	0.024	0.031	0.16	0.017
Panel B: Spot the difference test 2 score							
Competition	-0.007	-0.004	-0.008	0.01	0.013	0.046	-0.019
	[0.028]	[0.042]	[0.038]	[0.038]	[0.045]	[0.076]	[0.030]
Piece rate	0.022	0.051	-0.003	0.021	0.049	0.092	0.007
	[0.033]	[0.045]	[0.049]	[0.043]	[0.052]	[0.073]	[0.037]
Test 1	0.740***	0.805***	0.693***	0.823***	0.641***	0.704***	0.749***
	[0.040]	[0.057]	[0.055]	[0.053]	[0.067]	[0.089]	[0.045]
Constant	0.297***	0.267***	0.338***	0.170***	0.352***	0.208**	0.304***
	[0.040]	[0.056]	[0.053]	[0.038]	[0.053]	[0.100]	[0.044]
Observations	423	202	221	234	189	68	355
R-squared	0.546	0.572	0.525	0.479	0.353	0.547	0.544