You're Not You When You're Hungry: Measuring The Impact of a Supplemental Nutrition Program on Childhood Test Scores

Daniel Mangrum*

October 20, 2019

Abstract

I estimate the impact of a supplemental nutrition intervention on math and language arts test scores at low income elementary schools in the Mississippi Delta. The intervention provided meals to students in order to replicate school breakfast and lunch over the course of the weekend. Using a difference-in-differences design, I estimate the effect of the intervention on the mean and the distribution of test scores. I find that treated students performed better on both language arts and math standardized tests. The average gains stem from a reduction in the share of students achieving at the lowest threshold and shifts toward higher thresholds. I also use administrative daily attendance data to estimate how the intervention affected attendance by day of the week. Attendance on Fridays improved likely due to the transfer incentive, however I find improvements in attendance on Mondays and Tuesdays which is evidence of improved nutrition over the course of the weekend.

Keywords: nutrition, education, test scores, food insecurity.
JEL: I15, I25, I26, I28, H51, H52.

*Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235; *daniel.mangrum@vanderbilt.edu*.

1 Introduction

Food insecurity in the United States remains a topic of public concern despite federal, state, and local programs designed to alleviate hunger for low income households. In 2018, almost 14% of households with children in the United States experienced at least one instance of food insecurity during the year. Incidence of food insecurity varies across region with the lowest rates in the Midwest and the highest in the South. Mississippi has had the highest rate of food insecurity since 2010 with one in eight households experiencing bouts of nutritional scarcity. Recent research finds that lack of nutrition can cause children to lose focus and perform worse in school, however interventions that relieve food insecurity are shown to improve test scores and reduce behavioral issues.(Frisvold, 2015; Schwartz and Rothbart, 2017; Figlio and Winicki, 2005; Maluccio et al., 2009).

In this paper, I explore the impact of a supplemental nutrition program targeting low socioeconomic status children in the Mississippi Delta region on standardized test scores and attendance patterns. The intervention was designed to replicate the Free and Reduced Lunch Program (FRLP) students receive during the week to last over the course of the weekend for children in grades three through five at two schools in the Mississippi Delta region. A survey provided to teachers and parents suggests the intervention increased attendance on Fridays and resulted in fewer behavioral issues. This paper more rigorously studies the effect of the intervention by employing panel data on grade-by-school test scores to estimate the effect of the supplemental nutrition program on mean test scores, the percentage of students achieving at different thresholds, and daily attendance. I use a difference-in-differences design to compare the schools selected for treatment to a set of schools who were unaffected. I find that the one-time intervention improved students' test scores, particularly for language arts. The improvements in mean scores are largely driven by shifts away from students achieving the lowest achievement threshold and toward achievement of a proficient standard.

I also use a triple-difference design where I incorporate administrative daily attendance records and incorporate kindergarten through second grade students who were not included in the intervention as another dimension for comparison. I find that attendance was higher for students selected for treatment and the improvements were concentrated on Fridays, Mondays, and Tuesdays. The improvement in Friday attendance is likely due to the transfer effect of receiving the bundles of food. However, improvements in Monday and Tuesday attendance suggest that students had better nutrition over the course of the weekend.

The remainder of the paper is organized as follows: Section 2 reviews the literature for various nutrition interventions across the world to serve as benchmark treatment effects. Section 3 discusses the institutional details of this intervention. Section 4 describes the data used in this study. Section 5 discusses the empirical strategy employed to estimate the casual impact of the intervention. Section 6 summarizes the results of the analysis and Section 7 concludes the paper.

2 Literature

The link between nutrition and education has been explored extensively in both developed and developing countries. Many early studies were largely observational or relied on targeting children exhibiting signs of malnutrition. A study of 3,055 third-grade students in Vietnam examined the relationship between anthropometric status and educational achievement and found low test scores in mathematics and Vietnamese were correlated with both low height for age and low weight for age after controlling for age, sex, and school (Hall et al., 2001). A study in Chile surveyed a random sample of children graduating elementary school and high school and the results suggested academic achievement is positively correlated with consumption of dairy, meat, and eggs while consumption of fruits and vegetables is negatively correlated with low academic achievement. Food habits explained nearly 24% of variation in achievement for elementary school students (Ivanovic et al., 1992). A study in the United States used data from the Third National Health and Nutritional Examination Survey to test correlations between food insufficiency and academic achievement and other behavioral outcomes for respondents answering positively to instances of food insufficiency. After controlling for various covariates, the study finds students between 6 and 11 years old were more likely to report not getting along with other students, more likely to repeat a grade, have poorer arithmetic scores, and more likely to have seen a psychologist. Separately studying teenage respondents revealed similar results for behavior responses but no significant results for arithmetic scores suggesting food insecurity is more damaging to younger children than teenagers (Alaimo, Olson, and Frongillo, 2001).

In the case of studies using observation data, omitted variables correlated with diet and academic performance can bias the estimated effect of nutrition on achievement. As a result, more recent studies use more rigorous identification strategies to estimate the casual effect of nutrition on educational achievement. In 1995, the Minnesota state legislature approved a grant providing free breakfast to a set of six treatment schools to test the effectiveness of extending a similar program statewide. In addition to the six treatment schools, three control schools were chosen to provide a natural set of counter-factual students. During the three years of treatment, each treated school reported an increase in math and reading achievement while the control schools' scores were relatively flat. In addition, teachers in treatment schools reported fewer students complained of headaches and stomachaches, students were more energetic, and had an easier time concentrating. Teachers also reported fewer disciplinary issues in the treated schools with morning disciplinary referrals declining between 15 and 50% for all treated schools (Wahlstrom and Begalle, 1999). The use of a set of control schools and the use of administrative data allow for a more plausible argument for causality.

A randomized control trial (RCT) in Jamaica selected seventh grade students in the lower-third of academic performance and randomized the students into a control group and two treatment groups. In one treatment group, students were given a school lunch and in the other treatment group, students were given a syrup drink. The study found that the students who received the school lunch performed better than the control and alternate treatment on an arithmetic test and had better attendance records. The results were also robust to controlling for attendance (Powell, Grantham-McGregor, and Elston, 1983). Another study in Jamaica randomized a breakfast treatment to rural students in grades 2 through 5. Treated students received a school breakfast while control students received one-quarter of an orange and an equal amount of attention. While treated students were shown to have improved in height and weight, significant arithmetic results were only apparent in the youngest of treated children (Powell et al., 1998). An RCT conducted in South Africa randomized 108 students into a treatment and control group and provided breakfast every school day to the treatment group for six weeks. The school breakfast was found to have a positive effect on cognitive performance for the treatment group (Richter, Rose, and Griesel, 1997).

In contrast to randomized control trials, the use of longitudinal data to study diet and academic achievement has become popular due to the ability to perform within-unit comparisons over time. While observational analysis is prone to omitted variable bias, panel data on children can eliminate time invariant unobserved heterogeneity that could otherwise bias observational studies. One particular study performed a randomized control trial and followed Guatemalan children during early childhood and through adulthood. Between 1969 and 1977, four villages in were randomly assigned a high protein drink and a low protein drink meant to be given to children between birth and 36 months of age. The children that were randomly chosen for treatment into the more nutritious drink were tracked and interviewed in 2002. The study found positive effects of the intervention for the treated children. Treated women were found to have completed 1.2 more grades while both treated men and women had increases in both reading comprehension and non-verbal cognitive ability of one-quarter of a standard deviation (Maluccio et al., 2009).

In the case of the Guatemalan study, statistically and economically significant results can be found long after the end of treatment. This suggests that investments made to enhance nutrition in children at critical stages can have lasting effects on educational outcomes. On the other hand, an analysis of standardized test scores in Virginia show that interventions lasting as short as a week can have significant and immediate effects. This study identified schools that were under the threat of accountability sanctions if mean test scores were not improved. Researchers discovered that the school administrators systematically altered school lunch menus in an attempt to increase caloric counts during testing and finds that the schools who increased the caloric content of lunches the most saw the highest test score gains with an increase of 100 calories corresponding to increases of 7, 4, and 7 percentage points for mathematics, English, and social studies, respectively (Figlio and Winicki, 2005).

A more recent wave of studies estimate the effect of the National School Lunch Program and the rollout of School Breakfast programs across the U.S. Frisvold (2015) conducts a rigorous analysis of the School Breakfast Program using two identification strategies and multiple datasets and finds improvements in mathematics of 0.09 standard deviations and improvements in reading of 0.05 standard deviations for schools that adopt the School Breakfast Program. Similarly, Schwartz and Rothbart (2017) studies the impact of universal free lunch in New York City middle schools and the resulting impact on achievement. The study finds that an additional school lunch every two weeks improves math scores by around 0.08 standard deviations and improves language arts test scores by around 0.07 standard deviations. On the contrary, one recent study finds evidence of lower test scores when students are furthest away from the benefit receipt date. The study finds that when the students' family received SNAP benefits between 27 and 30 days prior to the test, math scores decline on average between 0.024 and 0.046 standard deviations (Cotti, Gordanier, and Ozturk, 2018). This suggests that families who receive federal food benefits may exhibit food insecurity at the end of the benefits cycle and that food insecurity may adversely impact academic performance.

While the literature covers a multitude of interventions spanning nutrition supplements, food stamps, school lunches, and school breakfasts, there are no studies (to my knowledge) covering interventions providing supplemental nutrition over the weekend for students reliant on free and reduced school lunches during the week. Since the introduction of the School Lunch Program and Free or Reduced Lunch Program, families have become increasingly reliant on school lunches to provide adequate nutrition to children in low socioeconomic settings. In addition to school lunches, schools with a large proportion of free or reduced lunch eligible students tend to also provide breakfast to students. While these students receive a majority of their caloric intake from school during the school-week, the students must rely on their own household for nutrition over the weekend. The literature suggests that improving nutrition over the course of the weekend should positively impact student health, attendance, academic achievement, and behavior.

3 Background

During the 2011-2012 school year, a local non-profit conducted an intervention in two elementary schools in the Mississippi Delta. Third- and fourth-grade students at Brooks Elementary School in Bolivar County and third-, fourth-, and fifth-grade students at Stampley Elementary School in Coahoma County were selected for inclusion into the treatment. Students at these schools were overwhelmingly eligible for the Free and Reduced Lunch Program (FRLP) with 99% eligibility at Brooks Elementary and 94% eligibility at Stampley Elementary. According to questionnaires administered by the non-profit, food insecurity was prevalent at both schools with more than half of parents at the two schools reporting some degree of food insecurity. In addition to the intervention at Brooks Elementary and Stampley Elementary, the non-profit conducted another intervention at two more schools during the 2015-2016 academic year, however public test score data for the most recent intervention is not available at this time.

Table 1: Contents of Weekend Supplemental Nutrition

2 bowls of cereal
2 small containers of fat-free milk
2 pieces fresh fruit
3 canned meats
1 cup applesauce
1 cup mixed fruit in syrup

In order to be included in the intervention, students were required to return a permission form signed by a parent or guardian. In total, 174 students were included in the treatment (73 at Stampley and 101 at Brooks).[1] Each Friday between September 30, 2011 and May 18, 2012 recyclable bags filled with food were distributed to all students involved in the treatment. Each bag contained food intended to last for weekend consumption denoted in Table 1. In order for the student to receive the food, the student must have attended school on Friday. Food was not distributed over weeks in which students were on break including: Fall Break, Thanksgiving Break, Winter Break, and Spring Break.

4 Data

I use data from the Mississippi Department of Education that include performance metrics on the Mississippi Curriculum Test, 2nd Edition (MCT2). The MCT2 is the standardized test administered every spring to public school students in the state of Mississippi during the sample period. MCT2 tests students in Language Arts and Mathematics beginning in the third grade. The data for this analysis spans the academic years 2007-2008 through the 2011-2012 school years. MCT2 scores are reported for each public school in the state of Mississippi for each grade level. In addition to average scores, the data also include information about the distribution of test scores. This data include the percentage of students in each school-grade cell achieving between three thresholds creating four bins of test scores: Minimal, Basic, Proficient, and Advanced. These metrics allow for analysis on both the average and the distributional change in test scores as a result of the intervention.

I also obtained attendance records for all Mississippi public schools for kindergarten through

[1] In 2011-2012, there were 56 third graders and 51 fourth graders who took the standardized tests at Brooks Elementary. At Stampley, there were 28 third graders, 34 fourth graders, and 38 fifth graders who took the tests.

fifth grade students for the sample years. These records contain the number of students absent from each school-grade cell on a daily basis throughout each school year. To create an outcome measure that is consistent across various school sizes, I compute the number of absences by day-of-week per enrolled student at the school-grade cell. Enrollment data comes from the Common Code of Data (CCD).

Lastly, I include control variables for changes in other programs that might also affect food insecurity. I merge data on Free and Reduced Lunch Program (FRLP) eligibility for each school-grade-year cell from the Common Core of Data and I include county level Supplemental Nutrition Assistance Program (enrollment) for each school. Table 2 reports summary statistics for outcome measures and demographics between the 2007-2008 and 2010-2011 school years.[2] Column 1 includes the full sample of Mississippi public schools that contain at least one of grades three, four, or five. Column 2 contains only the two treated schools, J.W. Stampley Elementary and Brooks Elementary. Comparing these columns reveals that the schools selected for the intervention score worse than the average Mississippi elementary school in both Language Arts and Mathematics test scores. The distributions for both tests follow a similar pattern with the treated schools having a larger percentage of students achieving Minimal and Basic standards and fewer students achieving Proficient and Advanced. Students at the schools selected for treatment also missed more days of school per student with those increases largely occurring on Mondays and Fridays. Lastly, the treated schools are populated by a larger share of Black students, have smaller average grade sizes, and have higher FRLP eligibility.

5 Empirical Strategy

5.1 Identification

Previous qualitative surveys administered to teachers and parents of the treated students suggest that students were better behaved, more attentive in class, and more likely to attend school on Friday during the year of the intervention.[3] However, these impacts cannot be interpreted as casual effects of the intervention without considering how other students performed during the

[2] Schools with missing information on the percent of students receiving Free and Reduced School lunches are omitted from the analysis.

[3] From summary report from the administering non-profit organization. Available upon request.

Table 2: Descriptive Statistics: Test Scores and Demographic Variables

Variable	(1) All MS Schools	(2) Treated Schools
LA Score	-0.048	-0.317
% Minimal	15.4	20.1
% Basic	36.2	45.5
% Proficient	37.7	27.9
% Advanced	10.7	6.4
Test Takers	87.6	42.7
Math Score	-0.046	-0.245
% Minimal	16.2	16.3
% Basic	28.5	37.9
% Proficient	45.1	41.6
% Advanced	10.2	4.3
Test Takers	87.6	42.7
Absences per student	6.00	6.35
Monday	1.31	1.37
Tuesday	1.17	1.21
Wednesday	1.11	1.19
Thursday	1.08	1.13
Friday	1.33	1.44
% White	41.4	0.0
% Black	54.5	98.4
% Hispanic	1.9	0.3
% Asian	0.5	0.0
% Free Lunch	77.5	98.3
Schools	532	2

The table above reports means for outcome variables (test scores) and demographic variables for the full sample of MS schools separately from the two treated schools. Outcome variables include normalized mean test scores in Langauage Arts and Mathematics on the MCT2 state test along with the number of test takers for each school-grade level. Demographic variables include the proportion of each school-grade that is white, black, Hispanic, or Asian as well as the percent of each school-grade cell that is eligible for Free and Reduced School Lunch. Demographic and FRLP data come from IPEDS while test score data come from Mississippi Department of Education.

2011-2012 school year. It is possible students that were not exposed to the intervention also performed better in school and had better attendance in the 2011-2012 school year for reasons unrelated to the intervention. Further, it could be the case that parents and teachers were more likely to answer positively on these surveys as a result of knowledge of the treatment in hopes of continued aid.

To help mitigate the concerns, I employ a difference-in-differences identification strategy to estimate the effect of the supplemental nutrition program on test scores and attendance for the students in the schools selected for the intervention. This design compares changes in outcomes of treated school-grade units to changes in outcomes for otherwise similar untreated school-grade units to control for any factors that might influence students in the same manner in the same years. If the necessary assumption, discussed below, is satisfied the estimated effect is the Intent to Treat (ITT) estimate of the intervention.[4]

The baseline specification for the test score outcomes follows a basic difference-in-differences framework:

$$y_{gst} = \gamma treat_{st} + \delta_g + \delta_s + \delta_t + \delta_s \times t + \varepsilon_{gst}, \tag{1}$$

where y_{gst} is a test score outcome measure for grade g at school s during academic year t. $treat_{st}$ is a binary variable which equals one for grades three through five at the two treated schools in the 2011-2012 academic year where γ is the estimated ITT. Also included are fixed effects for schools (δ_s), grades (δ_g), and academic years (δ_t). I also include a school specific linear time trend to capture differences in test score trends for each school. Standard errors are clustered at the school level to allow for correlation between grade levels within a school.

For the attendance outcomes, I can include one additional level of differences in the analysis. While standardized tests are only administered for students grade three and above, the attendance data contains records for all students in kindergarten through fifth grade. I use this additional data to compare the absence rate *within* treatment schools across the treated grades (3-5) and the untreated grades (K-2). The DDD specification is similar to Equation (1):

$$y_{gst} = \gamma treat_{gst} + \delta_g + \delta_s + \delta_t + \delta_s \times t + \varepsilon_{gsct}, \tag{2}$$

[4] Since not all students returned permission slips to enroll in the intervention, the effect estimated is not the Average Treatment Effect on the Treated.

except the DDD coefficient, $treat_{gst}$, also varies by grade level since the sample also includes absence data for the untreated grades K-2.

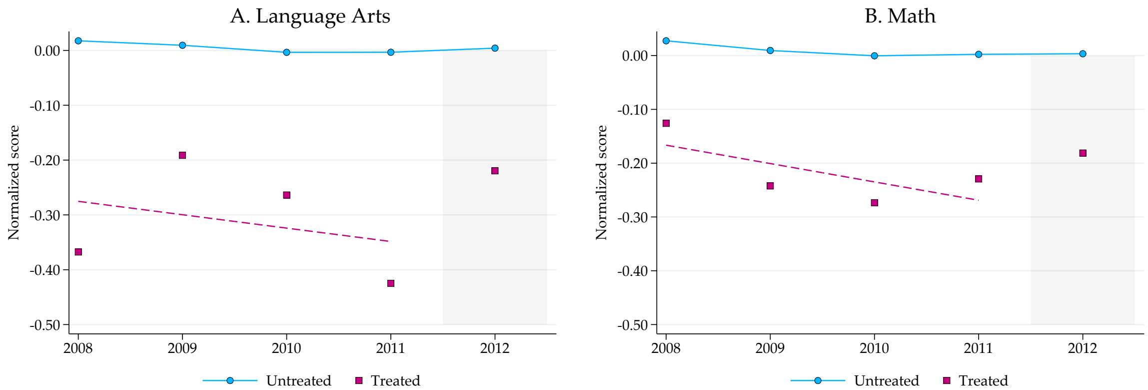
5.2 Selection of Comparison Group

As mentioned above, the estimated coefficient γ can only be interpreted as the ITT if the treated units would have evolved similarly to the control units in the absence of treatment. If the untreated units used as controls in the sample are systematically different from the treated units, the observed outcomes in the year of treatment cannot be reasonably used as a counter-factual outcome for the treated units. Thus, it must be the case that *changes* in test scores in the academic years prior to 2011-2012 are similar across the treated and untreated units. As described in Section 4, the schools selected for treatment had, on average, lower test scores, higher participation in FRLP, and had a larger share of Black students than the overall sample of schools in Mississippi. Figure 1 shows the trend in Language Arts and Mathematics test scores for the schools selected for treatment versus the sample of all other Mississippi Schools. Since the test scores for the full sample are normalized, the full sample of schools will be (near) mean zero for all years.[5] However, for the treated schools, the scores report where in the distribution of the sample the treated schools fall. Language Arts and Math scores for the treated schools are around 0.3 and 0.2 standard deviations lower than the mean, respectively. Additionally, both scores are slightly decreasing in the years leading up to the intervention. This trend could be the result of falling raw scores in the schools selected for treatment, gains in other schools, or both. Regardless, the trend in scores between the schools selected for treatment and the full sample of schools suggests that the full sample of schools might make for a suitable control group for the treated schools despite the level differences.

I also select a subsample of schools from the sample of all Mississippi public schools that is more similar to the two treated schools on the observables in which the treated schools most differ from the full sample: FRLP eligibility, size of grade level, and proportion of the student body who is Black. In addition, I also consider distance from the two treated schools as-the-crow-flies. In many cases, matching (or inverse probability weighting) methods use a nonlinear model such as a probit or a logit model to find comparison units that are similar to the units selected for treatment. However, in this setting, a nonlinear model is not feasible since there are few schools selected for

[5]Some schools are removed from the sample for missing data.

Figure 1: Mean Test Score Trends for Full Sample versus Treated Schools



Each figure above plots the mean test scores for the full sample of MS schools against the mean test scores for the treated schools in each year of the data. In addition, the red dashed line shows the trend in test scores for the treated schools in the three years before treatment. The shaded area, the 2011-2012 school year, denotes the year of treatment.

treatment and because the treated schools are largely boundary values for the selection criterion. Instead, I use Mahalanobis multivariate distance (MD) to rank the schools on similarity metrics which allocates more weight to schools that are more similar to the two treatment schools (Rubin, 1980). This algorithm takes selection variables and corresponding values on which to the match observations and calculates the distance between each observation in the sample from the mean values of the treated schools. The distance for each matching variable is weighted by the inverse of the variance-covariance matrix of the selection variables. The result is a weighted distance measure, measured in standard deviations, ranking how similar units are to the variable values on which the matching is constructed.

Table 3 lists the top 15 Mississippi schools with the highest Mahalanobis distance scores including the two treated school, Brooks Elementary and J.W. Stampley Elementary, in bold. The schools measured as most similar to the treated schools are all largely similar in the measured demographics by construction. In fact, there are many schools which are have smaller distance measures than the schools chosen for treatment since these schools better match the mean values of the variables than each individual school does alone. To create the matched sample of comparison schools, I select the subsample of comparison schools in top quartile in Mahalanobis distance so that the 25% of the sample most similar to the treatment schools are chosen. Table 3 also details the bottom 15 schools selected in this subsample in terms of similarity to the treatment schools.

Table 3: Comparison Group by Mahalanobis Multivariate Distance

Rank	School	District	% FRLP	% Black	Test Takers	Avg. Dist	Mahalanobis Distance
1	Shelby School	North Bolivar	98.6	98.4	48.2	12.8	0.065
2	I T Montgomery Elementary	Mound Bayou	99.0	99.3	44.3	17.2	0.075
3	Myrtle Hall IV Elementary	Clarksdale	98.5	99.8	47.7	8.1	0.166
4	Lyon Elementary	Coahoma County	98.9	98.0	41.8	11.7	0.174
5	**J W Stamply Elementary**	Clarksdale	97.9	98.8	37.2	0.0	0.195
6	**Brooks Elementary**	North Bolivar	98.9	97.9	52.3	0.0	0.195
7	Booker T Washington	Clarksdale	96.8	99.0	45.3	7.7	0.224
8	Nailor Elementary	Cleveland	98.7	97.0	33.3	26.6	0.248
9	Hunter Middle	Drew	96.4	96.3	41.3	23.7	0.252
10	A W James Elementary	Drew	95.8	90.7	51.2	23.5	0.261
11	Geo H Oliver Elementary	Clarksdale	96.8	99.5	58.8	8.5	0.288
12	Jonestown Elementary	Coahoma County	98.9	99.5	45.5	18.3	0.299
13	McEvans School	Shaw	96.8	97.0	47.1	33.8	0.335
14	Friars Point Elementary	Coahoma County	97.9	98.0	30.9	17.5	0.335
15	West Bolivar Elementary	West Bolivar	95.9	92.8	65.4	27.9	0.346
121	Galloway Elementary	Jackson Public	96.1	99.0	41.8	126.6	1.844
122	Lake Elementary	Jackson Public	96.7	98.3	75.1	126.4	1.846
123	Philadelphia Elementary	Philadelphia	84.3	73.0	89.4	129.0	1.847
124	Sykes Elementary	Jackson Public	89.0	86.3	72.7	131.2	1.847
125	Hopkins Elementary	Jackson Public	96.0	98.3	81.9	124.8	1.854
126	Johnson Elementary	Jackson Public	94.7	98.8	56.6	125.6	1.858
127	Poindexter Elementary	Jackson Public	96.9	99.3	31.4	127.8	1.862
128	Lester Elementary	Jackson Public	94.4	96.5	48.2	129.6	1.874
129	George Elementary	Jackson Public	96.2	97.8	26.6	129.0	1.878
130	Noxapater Attendance Center	Louisville	80.6	50.3	30.0	120.6	1.880
131	Smith Elementary	Jackson Public	94.6	99.8	64.3	125.0	1.881
132	H W Byers Elementary	Marshall County	90.7	51.3	63.4	89.3	1.890
133	Key Elementary	Jackson Public	96.2	98.5	66.7	130.0	1.892
134	Green Elementary	Jackson Public	93.3	99.5	60.4	123.8	1.895
135	Houlka Attendance Center	Chickasaw County	78.7	39.5	43.1	94.9	1.907

The table above ranks the top 15 and bottom 15 schools selected in the subsample as determined by the Mahalanobis multivariate distance using the mean values of the two treated schools for %FRLP, % Black, number of Test Takers and average distance from Brooks and Stampley Elementary. Data on FRLP eligibility and % Black come from IPEDS while the number of test takers comes from the Mississippi Department of Education. Avg. Dist. is the average distance between the school and Brooks Elementary and Stampley Elementary and are computed using the Stata program geodist which calculates distance as-the-crow-flies. Treated schools in bold.

Table 4 repeats Table 2 adding the summary statistics for the schools in the comparison group determined by the constructed Mahalanobis multivariate distance measure. Selection on the three demographic characteristics and the average distance creates a comparison group that is much more similar to the treated units than the overall sample. The 133 schools in the comparison group only differ in FRLP eligibility by 4.2 percentage points and differ in the proportion of Black students by only 10.3 percentage points compared to the overall sample difference of 20.8 and 43.9 percentage points, respectively. The number of test takers for each grade differs by 14.6 students compared to the overall sample difference of 44.9. The more suitable control group generates average test scores that are also far more similar than the overall sample: average test scores in Language Arts are within 0.035 standard deviations and test scores in Math are within 0.021 standard deviations.

Figure 2 repeats Figure 1 with the comparison group selected using the Mahalanobis similarity score. This subsample better matches the *level* of test scores for both Language Arts and Math than the full sample but a worse job matching the *trend* in scores. In this figure, the potential treatment impact is more apparent with 2012 Language Arts and Math scores for the treated units outpacing the same year scores for the comparison group. This trend break suggests that the intervention had a positive impact on the treated schools in both subject areas. If the 2012 scores for the comparison group are used as a counter-factual for the treated units, this would suggest improvements in test scores larger than 0.1 standard deviations. In the next section, I present results from the estimation of Equation (1) for the full sample and the selected comparison group in order to estimate the causal impact of the supplemental nutrition intervention on test scores. In addition, I also estimate Equation (1) for the full sample and the subsample using the Mahalanobis distance measure to weigh the observations. This specification allows for the schools most similar to the treated schools to take on more weight in the analysis than the schools that are less similar.

5.3 Inference

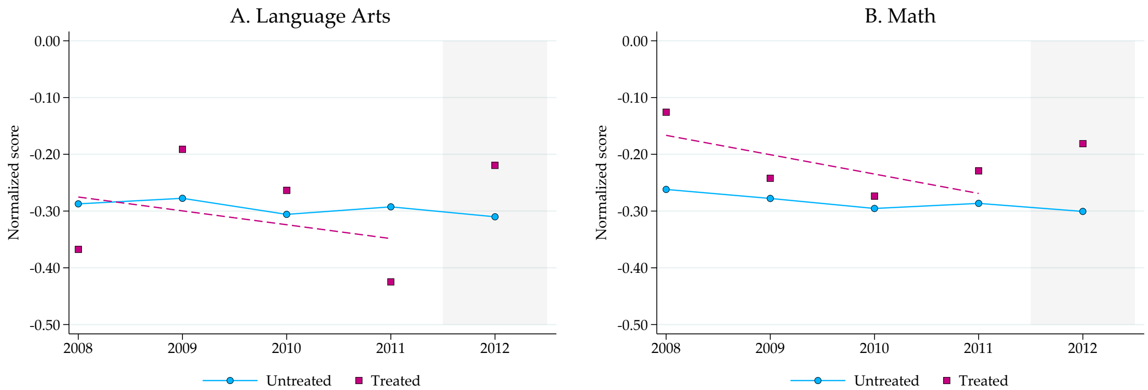
In the baseline specification, I cluster standard errors at the school level to allow for correlation in grades within the same school. However, recent research suggests that cluster robust standard errors perform remarkably poorly when there are few treated clusters (MacKinnon and Webb,

Table 4: Descriptive Statistics: Test Scores and Demographic Variables

Variable	(1) All MS Schools	(2) MD Matched Schools	(3) Treated Schools
LA Score	-0.048	-0.282	-0.317
% Minimal	15.4	20.7	20.1
% Basic	36.2	41.7	45.5
% Proficient	37.7	31.2	27.9
% Advanced	10.7	6.5	6.4
Test Takers	87.6	57.3	42.7
Math Score	-0.046	-0.266	-0.245
% Minimal	16.2	21.5	16.3
% Basic	28.5	32.6	37.9
% Proficient	45.1	40.1	41.6
% Advanced	10.2	5.7	4.3
Test Takers	87.6	57.3	42.7
Absences per student	6.00	5.61	6.35
Monday	1.31	1.25	1.37
Tuesday	1.17	1.09	1.21
Wednesday	1.11	1.03	1.19
Thursday	1.08	1.01	1.13
Friday	1.33	1.24	1.44
% White	41.4	9.2	0.0
% Black	54.5	88.1	98.4
% Hispanic	1.9	1.2	0.3
% Asian	0.5	0.1	0.0
% Free Lunch	77.5	94.1	98.3
Schools	532	133	2

The table above reports means for outcome variables (test scores) and demographic variables for the full sample of MS schools, the upper quartile of MS schools most similar to the treated schools by Mahalanobis multivariate distance, and the two treated schools. Outcome variables include normalized mean test scores in Langauage Arts and Mathematics on the MCT2 state test along with the number of test takers for each school-grade level. Demographic variables include the proportion of each school-grade that is white, black, Hispanic, or Asian as well as the percent of each school-grade cell that is eligible for Free and Reduced School Lunch. Demographic and FRLP data come from IPEDS while test score data come from Mississippi Department of Education.

Figure 2: Mean Test Score Trends for Matched Schools versus Treated Schools



Each figure above plots the mean test scores for the subsample of MS schools most similar to the treated group as described in 5 against the mean test scores for the treated schools in each year of the data. In addition, the red dashed line shows the trend in test scores for the treated schools in the three years before treatment. The shaded area, the 2011-2012 school year, denotes the year of treatment.

forthcoming). As a result of the few number of treated units, the finite sample standard errors tend to be underestimated causing over-rejection of the null hypotheses. In this setting, there are only two treated schools and since schools are the unit of clustering, it is likely that the typical cluster robust standard errors are too small. To correct for this, I follow the RI-β algorithm prescribed by MacKinnon and Webb (forthcoming) by conducting a randomization inference exercise. I perform 3,500 placebo replications in which, instead of the observed treated schools, I suppose that two other random schools were chosen for treatment and I estimate Equation (1) under this supposition. This exercise compares the difference-in-difference coefficient estimated in Equation (1) with the distribution of placebo estimates. For each placebo replication, I draw two schools at random from the set of all schools in Mississippi with FRLP greater than 75%. This restriction of potential treatment schools is chosen so that only schools that might have been reasonably chosen for the intervention are included. I collect all estimated placebo $\hat{\gamma}$ and generate two empirical p-values for each outcome variable. First, I use the two-sided test where I estimate the share of $\hat{\gamma}$ that are larger in absolute value than the $ga\hat{m}ma$ estimated using the actual treated schools:

$$\bar{p} = \frac{1}{3500} \sum_{n=1}^{3500} 1 \cdot \left\{ |\hat{\gamma}_n| \geq |\hat{\gamma}_{true}| \right\} \tag{3}$$

In addition, I construct a one-sided test that tests where there is an improvement in the outcome.

For mean test scores and percent achieving Proficient and Advanced, the empirical p-value is

$$\bar{p} = \frac{1}{3500}\sum_{n=1}^{3500} 1 \cdot \left\{\hat{\gamma}_n \geq \hat{\gamma}_{true}\right\}, \tag{4}$$

and for percent achieving Minimal and Basic and for absences, the empirical p-value is

$$\bar{p} = \frac{1}{3500}\sum_{n=1}^{3500} 1 \cdot \left\{\hat{\gamma}_n \leq \hat{\gamma}_{true}\right\}. \tag{5}$$

Since MacKinnon and Webb (forthcoming) note that this algorithm has a tendency to under-reject, the one-sided test will provide a lower threshold for rejecting the null hypothesis of no improvement in outcomes. For each outcome, I present the clustered standard errors which are likely under-estimated along with the one- and two-sided tests using the empirical p-values.

6 Results

6.1 Standardized Test Scores

Table 5 reports the results from the estimation of Equation (1). Columns 1 through 4 report the results for Language Arts scores while columns 5 through 8 report the results for Math scores. The first two columns for each subject score include the full sample while the last two columns include only the subsample of the most similar quartile of schools based on the multivariate distance measure described above. Odd numbered columns report unweighted regressions while the even numbered columns report regressions that are weighted by the Mahalanobis distance measure. For each outcome, I report p-values using clustered standard errors in parenthesis along with one- and two-sided empirical p-values using RI-β from MacKinnon and Webb (forthcoming) in brackets. Columns 1 through 4 suggest improvements in Language Arts test scores regardless of specification. The estimates using the full sample range from 0.24 to 0.28 from a pre-treatment mean of -0.317. Hence, the results suggest the intervention closed 75% to 88% of the gap in test scores between the treated school and the state average. The effect on math scores is slightly smaller at between 0.16 to 0.18 standard deviation improvement from a mean of -0.245 representing a closing of the test score gap of 65% to 73%. In total, the point estimates suggest significant test score improvements as a results of the nutrition intervention for the school-grade cells selected for treatment.

Table 5: Difference-in-Differences Estimates: Language Arts and Math Scores

	Language Arts Score				Math Score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.240	0.241	0.284	0.276	0.166	0.161	0.177	0.174
CRVE p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RI-β 2-sided p	[0.184]	[0.203]	[0.157]	[0.158]	[0.378]	[0.388]	[0.235]	[0.237]
RI-β 1-sided p	[0.097]	[0.109]	[0.083]	[0.083]	[0.207]	[0.208]	[0.139]	[0.139]
Outcome Mean	-0.317	-0.317	-0.317	-0.317	-0.245	-0.245	-0.245	-0.245
Sample	Full	Full	Matched	Matched	Full	Full	Matched	Matched
Weights	No	Yes	No	Yes	No	Yes	No	Yes
Schools	523	523	132	132	523	523	132	132
Grades	3	3	3	3	3	3	3	3

Each column contains the estimated difference-in-differences coefficient where the outcome variable is either the standardized language arts or mathematics test score as denoted in the column title for a school-grade cell. Odd numbered columns are unweighted and even numbered columns use the normalized Mahalanobis distance measure as a regression weight. Columns One, Two, Five, and Six use the full sample of schools. Columns Three, Four, Seven, and Eight use the matched subsample of schools using the Mahalanobis distance measure. The Outcome Mean reports the mean value for the given outcome across the treated schools in the years prior to treatment. P-values using standard errors clustered at the school level are presented in parenthesis. One- and two-sided empirical p-values using RI-β are presented in brackets.

Across all eight specification, the p-values corresponding to each coefficient suggest the null hypothesis can be rejected at the 1% level. This is likely evidence that indeed the standard errors are under-estimated due to the low number of treated clusters in the sample. As such, these p-values should not be trusted and the RI-β empirical p-values are likely to be better estimates for inference. Using the 2-sided empirical p-values, we cannot reject the null hypothesis that the language arts treatment effect is not equal to zero. The p-values from the one-sided test lie just outside of the rejection region for a test of improving test scores. Additionally, the p-values for the mathematics scores are almost twice those of the languages arts scores and, as such, the null hypothesis also cannot be rejected. In total, the evidence suggests large improvements in test scores for the treated units but the small number of treated clusters produces errors that make this conclusion difficult.

6.2 Distributional Treatment Effects

The discussion in the previous section presents evidence that average standardized test scores for the schools selected for treatment improved in the year of the intervention. The magnitude of these estimates are quite large and suggest that the treated students performed almost as well as the average Mississippi student despite having average test scores three-tenths of a standard deviation in previous years. However, due to the few number of treated units, inference on the average is difficult and the null hypothesis cannot be rejected by traditional.

Table 6: Difference-in-Differences Estimates: Distributional Effects for Language Arts and Mathematics Test Scores

	Language Arts Score				Math Score			
A. Percent Minimal	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-8.488	-8.441	-8.798	-8.720	-6.615	-6.482	-6.473	-6.440
CRVE p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.004)	(0.010)	(0.011)
RI-β 2-sided p	[0.112]	[0.138]	[0.108]	[0.108]	[0.256]	[0.287]	[0.193]	[0.195]
RI-β 1-sided p	[0.066]	[0.075]	[0.051]	[0.051]	[0.133]	[0.142]	[0.096]	[0.097]
Outcome Mean	20.147	20.147	20.147	20.147	16.284	16.284	16.284	16.284
B. Percent Basic								
Treatment	-11.661	-11.347	-11.237	-10.892	3.767	3.914	3.297	3.442
CRVE p-value	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
RI-β 2-sided p	[0.075]	[0.085]	[0.072]	[0.078]	[0.475]	[0.488]	[0.273]	[0.268]
RI-β 1-sided p	[0.031]	[0.032]	[0.033]	[0.030]	[0.792]	[0.782]	[0.896]	[0.895]
Outcome Mean	45.511	45.511	45.511	45.511	37.858	37.858	37.858	37.858
C. Percent Proficient								
Treatment	19.222	18.848	17.804	17.624	-2.817	-3.059	-2.981	-3.087
CRVE p-value	(0.000)	(0.001)	(0.002)	(0.002)	(0.270)	(0.237)	(0.274)	(0.262)
RI-β 2-sided p	[0.013]	[0.016]	[0.029]	[0.029]	[0.641]	[0.660]	[0.337]	[0.337]
RI-β 1-sided p	[0.003]	[0.003]	[0.004]	[0.004]	[0.692]	[0.681]	[0.861]	[0.861]
Outcome Mean	27.884	27.884	27.884	27.884	41.595	41.595	41.595	41.595
D. Percent Advanced								
Treatment	0.925	0.935	2.220	1.977	5.699	5.662	6.189	6.117
CRVE p-value	(0.405)	(0.395)	(0.097)	(0.125)	(0.000)	(0.000)	(0.000)	(0.000)
RI-β 2-sided p	[0.803]	[0.808]	[0.347]	[0.353]	[0.193]	[0.196]	[0.107]	[0.107]
RI-β 1-sided p	[0.424]	[0.419]	[0.186]	[0.186]	[0.102]	[0.101]	[0.052]	[0.053]
Outcome Mean	6.447	6.447	6.447	6.447	4.268	4.268	4.268	4.268
Sample	Full	Full	Matched	Matched	Full	Full	Matched	Matched
Weights	No	Yes	No	Yes	No	Yes	No	Yes
Schools	523	523	132	132	523	523	132	132
Grades	3	3	3	3	3	3	3	3

Each column contains the estimated difference-in-differences coefficient. Each panel contains a different outcome variable where the outcome is the percentage of students in the school-grade cell achieving at a certain achievement threshold. The thresholds increase in achievement level from Minimal (lowest), Basic, Proficient, and Advanced (highest). Odd numbered columns are unweighted and even numbered columns use the normalized Mahalanobis distance measure as a regression weight. Columns One, Two, Five, and Six use the full sample of schools. Columns Three, Four, Seven, and Eight use the matched subsample of schools using the Mahalanobis distance measure. The Outcome Mean reports the mean value for the given outcome across the treated schools in the years prior to treatment. P-values using standard errors clustered at the school level are presented in parenthesis. One- and two-sided empirical p-values using RI-β are presented in brackets.

On the other hand, the Mississippi Department of Education provides information on the distribution of students in each school-grade cell. Table 6 presents the difference-in-difference estimates for the percentage of students in each performance bin from worst (Minimal) in Panel A to best (Advanced) in Panel D for both language arts and mathematics. For both language arts and mathematics, there are declines in the percentage of students achieving at the lower threshold. The effect is larger again for language arts and the point estimate suggests a 42% improvement.[6] The reduction for mathematics is similar in proportional terms but smaller in magnitude. Again, the empirical p-values for the two-sided test are about twice as large for mathematics than for language arts. We are able to reject the null hypothesis of no improvement for language arts at the 10% level but the empirical p-value for mathematics is larger than 0.100 at 0.136.

The results for the second bin of achievement are mixed across language arts and mathematics. For language arts, the point estimates suggest declines around 11 percentage points while math achievement in this bin is around 3.5 percentage points larger. The point estimate for the percentage of students in the Basic bin is quite imprecise, however the reduction in students achieving Basic for language arts is significant at the 10% level for a two-sided test and at the 5% level for a one-sided test.

Since students move from one bin to another by construction, the decline in both Minimal and Basic for language arts suggest students are shifting into higher achievement. This result is reflected in the point estimates in Panel C for the percentage of students achieving Proficient standards. The point estimate is quite large at between 17 and 19 percentage points and the estimate is significant at the 5% level for the two sided test and the 1% level for the one-sided test using the empirical p-values. This suggests that treated students are shifting from lower achievement bins to higher achievement bins for language arts scores. On the contrary, the point estimates suggest a reduction in students achieving Proficient in math. This could be due to students leaving the Proficient bin and moving into either Basic or Advanced. However, this point estimate is very imprecise: even the under-estimated clustered standard errors are unable to reject the null hypothesis for these estimates.

Lastly, the effect of the intervention on the percentage of students achieving Advanced is presented in Panel D. For language arts, the estimates are economically small and are imprecisely

[6] 8.5 percentage point reduction off of a base of 20.1 percentage points

estimated. However, the effect on mathematics is quite large relative to the baseline average for the treated schools in previous years. This large improvement in mathematics might explain the negative point estimates for the Proficient bin in Panel C: it may be the case that students previously achieving at Minimal are able to move to Basic and some students who previously achieved at Proficient are able to move into Advanced. Further, the point estimate is significant at the 10% level for the one-sided empirical p-value.

In total, analysis on the distribution of test scores suggest reductions in students achieving at lower thresholds with shifts towards higher achievement bins. For language arts, these improvements largely stem from students exiting the Minimal and Basic bins and entering the Proficient bins. For mathematics, there is some evidence students leave the Minimal bin and other students enter the Advanced bin but without micro-level data, it is not possible to track which students move across which bins.

6.3 Attendance

In addition to test score outcomes, I also explore how daily attendance is affected by the supplemental nutrition intervention. As mentioned above, attendance data is available for students in kindergarten through fifth grade. As a result, this data allow for a third dimension of comparison in which untreated students (K-2) in treated schools can be compared to treated students (3-5) in treated schools. The results from this section can help to explain the mechanisms by which the nutrition intervention improves test scores. If the supplemental nutrition program indeed improves nutrition over the weekend, students should feel healthier on Mondays and should be less likely to miss school on Mondays. On the other hand, the intervention could also act as an in-kind transfer in which students must attend school on Fridays to receive the transfer. If this is the case, attendance on Fridays should improve.

Table 7 reports the results of this strategy where each panel separately reports the estimated impact of the intervention on the number of absences per enrolled student for each day of the week in the year of the intervention. Panel A reports these estimates for Mondays. On average, students in the treated schools missed around 1.53 Mondays per year making Monday the most frequently missed school day of the week. The estimates suggest a reduction in Mondays missed

Table 7: Triple Difference Estimates: Absences per Enrolled Student by Day-of-Week

A. Monday	(1)	(2)	(3)	(4)
Treatment	-0.214	-0.208	-0.177	-0.175
CRVE p-value	(0.033)	(0.037)	(0.073)	(0.076)
RI-β 2-sided p	[0.134]	[0.158]	[0.111]	[0.110]
RI-β 1-sided p	[0.052]	[0.063]	[0.040]	[0.040]
Outcome Mean	1.533	1.533	1.533	1.533
B. Tuesday				
Treatment	-0.202	-0.206	-0.207	-0.202
CRVE p-value	(0.004)	(0.003)	(0.004)	(0.005)
RI-β 2-sided p	[0.253]	[0.273]	[0.162]	[0.172]
RI-β 1-sided p	[0.118]	[0.134]	[0.071]	[0.068]
Outcome Mean	1.318	1.318	1.318	1.318
C. Wednesday				
Treatment	-0.042	-0.044	-0.046	-0.042
CRVE p-value	(0.008)	(0.006)	(0.037)	(0.057)
RI-β 2-sided p	[0.839]	[0.827]	[0.358]	[0.347]
RI-β 1-sided p	[0.516]	[0.503]	[0.751]	[0.749]
Outcome Mean	1.333	1.333	1.333	1.333
D. Thursday				
Treatment	0.000	-0.003	0.000	-0.001
CRVE p-value	(0.999)	(0.964)	(0.994)	(0.991)
RI-β 2-sided p	[0.808]	[0.814]	[0.333]	[0.338]
RI-β 1-sided p	[0.353]	[0.352]	[0.122]	[0.123]
Outcome Mean	1.187	1.187	1.187	1.187
E. Friday				
Treatment	-0.351	-0.342	-0.311	-0.304
CRVE p-value	(0.000)	(0.000)	(0.000)	(0.000)
RI-β 2-sided p	[0.039]	[0.051]	[0.036]	[0.035]
RI-β 1-sided p	[0.017]	[0.023]	[0.028]	[0.028]
Outcome Mean	1.504	1.504	1.504	1.504
Sample	Full	Full	Matched	Matched
Weights	No	Yes	No	Yes
Schools	523	523	132	132
Grades	6	6	6	6

Each column contains the estimated difference-in-differences coefficient where the outcome variable is the number of absences on each day of the week per enrolled student. The third difference in this specification uses the untreated kindergarten through second grade students. Odd numbered columns are unweighted and even numbered columns use the normalized Mahalanobis distance measure as a regression weight. Columns One and Two use the full sample of schools. Columns Three and Four use the matched subsample of schools using the Mahalanobis distance measure. The Outcome Mean reports the mean value for the given outcome across the treated schools in the years prior to treatment. P-values using standard errors clustered at the school level are presented in parenthesis. One- and two-sided empirical p-values using RI-β are presented in brackets.

in the year of treatment by 0.18 to 0.21 days per enrolled student representing a 11-14% decline in Monday absenteeism. These estimates are also precisely estimated with the one-side p-value significant at the 10% level. The results in Panel B are similar to those in Panel A with declines in attendance around 15-16% These estimates are somewhat less precise with the one-side empirical p-value near the cutoff of the rejection region at 0.109.

On the other hand, the results in Panels C and D suggest no change in attendance for Wednesday or Thursdays. Both estimates are economically small and not precisely estimated across all specifications.

Lastly, the effect on Friday attendance seems to be the largest across the week. The reduction in absenteeism on Fridays is between 20-23% and is significant at the 5% level for both one- and two-sided tests. These results lend evidence to the mechanisms presented above. The largest effect of the intervention on attendance is due to the transfer effect on Fridays. However, Monday is the most frequently missed day of school for the treated units and the nutritional intervention reduces absenteeism on Mondays and Tuesdays by around 15%.

7 Conclusion

In this paper, I estimate how improved nutrition affects elementary school students living in the most food insecure region in the United States. Although the total expenditure per student is relatively small, the estimated improvements in test scores and attendance are quite large. I find that students attending the school selected for treatment had larger, albeit imprecise, average test scores in both language arts and mathematics. The improvements in average scores is higher for language arts than for mathematics, but the intervention was able to close the gap in test score by around 70% from the state average.

I also decompose the effect of the intervention on the distribution of test scores by using the percentage of students achieving at various levels as outcomes. For language arts, I find large reductions in the percentage of students achieving at the lowest two bins and large increases in the percentage of students achieving at the Proficient bin. I find no effect of the intervention on moving students in the Advanced bin for language arts.

The results for mathematics are smaller but perhaps more divergent. While I find small reductions in those achieving at the lowest bin, I am unable to trace movements into the middle two bins. The estimates are smaller and less precise and do not represent traceable shifts from the lower tail. On the other hand, I do find large improvements relative to the baseline in the percentage of students achieving Advanced in mathematics that might be driven by a reduction in the students achieving at Proficient. These improvements in the highest achievement threshold are the most precisely estimated effects for mathematics scores.

Lastly, I also find improvements in attendance for students at the schools selected for treatment, but only for those in the treated grades. I use daily administrative attendance records to track changes in attendance by day of the week. The results provide evidence for the mechanisms by which test scores are improved. First, I find the largest effect of the intervention on attendance on Fridays with reductions in absenteeism by 20%. This evidence is consistent with the transfer program mechanism by which students must be present at school on Friday to receive the nutrition bundle. However, I also find significant improvements in attendance on both Monday and Tuesday. For the set of selected schools, Monday was the most commonly missed day of school prior to the intervention and attendance increased by over 10% on Mondays and almost 15% on Tuesdays. This is evidence that indeed the students selected for treatment had better nutritional intake over the course of the weekend subsequently improving general physical condition at the beginning of the week.

The results from this paper add to the literature of the role nutrition plays in human capital accumulation. The effect sizes from this one-shot intervention are on the higher end of the distribution from the education treatment effects literature. Further, these gains were achieved at a rather low cost. Kraft (2018) categorizes an intervention as "Large Effect Size/Low Cost" if the effect size is greater than 0.2 standard deviations and the cost is less that \$500 per pupil. Using the estimated improvement in language arts test scores, this intervention meets the benchmark for a large effect size. Further, the cost of the intervention must be below \$16.67 per week per student to be categorized as low cost. At an estimated \$3.63 per food bundle[7] plus administrative costs, it is very likely this intervention remains beneath the \$500 per student cost over the course of the 30-week school year. Lastly, this intervention is likely quite easy to scale. Since schools already

[7] This estimate is arrived at by assuming the following unit costs: cereal \$0.33, milk \$0.25, fruit \$0.25, canned meat \$0.50, applesauce \$0.23, and mixed fruit \$0.50.

have sufficient infrastructure to provide breakfast and lunch during school days, these items can be bought in bulk with the existing meal orders and distributed to students through the existing school infrastructure. As a result, this intervention is very likely to achieve the definition on Easy to Scale from Kraft (2018). However, the setting of this intervention is important when considering the external validity of these treatment effects. Since the Mississippi Delta region is the most food insecure region in the United States, it might be unreasonable to expect similar effect sizes if the intervention were to be repeated in another setting. Students that face a lower degree of food insecurity are unlikely to respond in a similar manner as the students in this setting. However, increasing the scale of this intervention across the Mississippi Delta is likely to achieve similar results for elementary school students in the area.

References

Alaimo, Katherine, Christine M Olson, and Edward A Frongillo. 2001. "Food insufficiency and American school-aged children's cognitive, academic, and psychosocial development." *Pediatrics* 108 (1):44–53.

Cotti, Chad, John Gordanier, and Orgul Ozturk. 2018. "When does it count? The timing of food stamp receipt and educational performance." *Economics of Education Review* 66:40–50.

Figlio, David N and Joshua Winicki. 2005. "Food for thought: the effects of school accountability plans on school nutrition." *Journal of Public Economics* 89 (2):381–394.

Frisvold, David E. 2015. "Nutrition and cognitive achievement: An evaluation of the School Breakfast Program." *Journal of public economics* 124:91–104.

Hall, Andrew, LN Khanh, TH Son, NQ Dung, RG Lansdown, DT Dar, NT Hanh, Helen Moestue, Ha Huy Khoi, and DA Bundy. 2001. "An association between chronic undernutrition and educational test scores in Vietnamese children." *European Journal of Clinical Nutrition* 55 (9):801–804.

Ivanovic, Daniza, Magaly Vásquez, Marcela Aguayo, Digna Ballester, Maximiliano Marambio, and Isabel Zacarías. 1992. "Nutrition and education. III. Educational achievement and food habits of Chilean elementary and high school graduates." *Archivos latinoamericanos de nutricion* 42 (1):9–14.

Kraft, Matthew A. 2018. "Interpreting effect sizes of education interventions." Tech. rep., Brown University Working Paper. Downloaded Tuesday, April 16, 2019, from . . . .

MacKinnon, James G and Matthew D Webb. forthcoming. "Randomization inference for difference-in-differences with few treated clusters." *Journal of Econometrics* .

Maluccio, John A, John Hoddinott, Jere R Behrman, Reynaldo Martorell, Agnes R Quisumbing, and Aryeh D Stein. 2009. "The impact of improving nutrition during early childhood on education among Guatemalan adults." *The Economic Journal* 119 (537):734–763.

Powell, Christine, Sally Grantham-McGregor, and M Elston. 1983. "An evaluation of giving the Jamaican government school meal to a class of children." *Human Nutrition. Clinical Nutrition* 37 (5):381–388.

Powell, Christine A, Susan P Walker, Susan M Chang, and Sally M Grantham-McGregor. 1998. "Nutrition and education: a randomized trial of the effects of breakfast in rural primary school children." *The American journal of clinical nutrition* 68 (4):873–879.

Richter, Linda M, Cynthia Rose, and R Dev Griesel. 1997. "Cognitive and behavioural effects of a school breakfast." *South African Medical Journal* 87 (1 Suppl):93–100.

Rubin, Donald B. 1980. "Bias reduction using Mahalanobis-metric matching." *Biometrics* :293–298.

Schwartz, Amy Ellen and Michah W Rothbart. 2017. "Let them eat lunch: The impact of universal free meals on student performance." .

Wahlstrom, Kyla L and Mary S Begalle. 1999. "More Than Test Scores: Results of the Universal School Breakfast Pilot in Minnesota." *Topics in Clinical Nutrition* 15 (1):17–29.