NBER WORKING PAPER SERIES

ADEQUATE (OR ADIPOSE?) YEARLY PROGRESS: ASSESSING THE EFFECT OF "NO CHILD LEFT BEHIND" ON CHILDREN'S OBESITY

Patricia M. Anderson
Kristin F. Butcher
Diane Whitmore Schanzenbach

Working Paper 16873
http://www.nber.org/papers/w16873

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2011

We thank Jannine Riggs, Denise Airola, and Jim Boardman of the Arkansas Department of Education for helpful discussions about the Arkansas education data and accountability rules. We received generous financial support from RWJ grant number #57922, Rockefeller Center at Dartmouth College Reiss Family Faculty Research Grant, and a Wellesley College Faculty Award. Brian Dunne, A.J. Felkey, Brenna Jenny and Alan Kwan provided excellent research assistance. We thank Eric Edmonds, Jonathan Guryan, seminar participants at UC-Davis, Boston College, Louisiana State University, the University of Toulouse, and conference participants at the American Economic Association 2010 annual meetings, and the Rockefeller Center Health Policy Workshop for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Adequate (or Adipose?) Yearly Progress: Assessing the Effect of "No Child Left Behind" on Children's Obesity
Patricia M. Anderson, Kristin F. Butcher, and Diane Whitmore Schanzenbach
NBER Working Paper No. 16873
March 2011
JEL No. I18,I21

ABSTRACT

This paper investigates how accountability pressures under No Child Left Behind (NCLB) may affect children’s rate of overweight. Schools facing increased pressures to produce academic outcomes may reallocate their efforts in ways that have unintended consequences for children’s health. For example, schools may cut back on recess and physical education in favor of increasing time on tested subjects. To examine the impact of school accountability programs, we create a unique panel data set of schools in Arkansas that allows us to test the impact of NCLB rules on students’ weight outcomes. Our main approach is to consider schools to be facing increased pressures if they are on the margin of passing – that is, if any subgroup at the school has a passing rate that is close to the AYP passing threshold, where we define close as being 5 percentage points above or below the threshold. We find evidence of small effects of accountability pressures on the percent of students at a school that are overweight. A follow-up survey of school principals points to reductions in physical activity and worsening of the food environment as potential mechanisms.

Patricia M. Anderson
Department of Economics
Dartmouth College
6106 Rockefeller
Hanover, NH 03755-3514
and NBER
patricia.m.anderson@dartmouth.edu

Kristin F. Butcher
Department of Economics
Wellesley College
106 Central Street
Wellesley, MA 02481
and NBER
kbutcher@wellesley.edu

Diane Whitmore Schanzenbach
School of Education and Social Policy
Northwestern University
Annenberg Hall, Room 205
2120 Campus Drive
Evanston, IL 60208
and NBER
dws@northwestern.edu

I. Introduction

Childhood obesity has increased dramatically over the past three decades, from about 4 percent during the mid-1970s to 16 percent today. There have been many changes in children's lives during the period when children's obesity has been increasing (see Anderson and Butcher 2006a). In particular, there have been changes at home and at school that may contribute to increased obesity. Understanding how the school environment may contribute to obesity is critical as the school environment may be relatively more within the control of public policy makers than the family environment. In recent years, especially, pressures on schools have changed dramatically. No Child Left Behind (NCLB) added explicit "accountability" incentives for schools to improve test score outcomes without necessarily providing more resources with which to produce these outcomes.

This paper investigates how accountability pressures under NCLB may affect children's obesity. Children's health is not among the outcomes for which schools are held accountable – standardized test achievement is the primary area monitored, with secondary emphasis on attendance and graduation rates. Schools facing increased pressure to produce academic outcomes may reallocate their efforts in ways that have unintended consequences for children's health. The new financial pressures due to accountability rules may, for example, induce school administrators to try to raise new funds through outside food and beverage contracts, or time pressures may cause them to cut back on recess and physical education in favor of increased time teaching tested subjects.[1]

To examine the impact of school accountability programs, we create a unique data set that allows us to test the impact of NCLB rules on students' weight outcomes. This dataset combines school-level rates of "obesity" and "overweight" for school children in Arkansas

[1] Additionally, schools may use food as rewards to motivate students, and there is the possibility that testing pressures increase cortisol secretions in children due to the increased stress, which may lead to weight gain.

with data from the Arkansas Department of Education on standardized test proficiency rates in English and math for all schools, by grade and subgroup.[2] The standardized test pass rates are those used for determining whether a school is making Adequate Yearly Progress (AYP) under NCLB. The main methodological difficulty in trying to assess the causal effect of accountability pressures on children's health outcomes is that schools that have children who perform poorly on tests may also have children who have worse health outcomes. Our empirical strategy, described in detail below, addresses this issue by taking advantage of the fact that the determination of which schools do and do not make Adequate Yearly Progress is very complicated, and even schools with generally high performing students (who are likely to be healthier) face accountability pressures that may induce changes at the school that adversely affect students' health status.

The main empirical results examine *whether* schools facing accountability pressure have more overweight and obese students. These results, however, cannot examine the mechanisms through which changes that are made to be NCLB compliant may affect children's obesity. In order to directly examine those mechanisms, we designed and fielded a survey of school principals in Arkansas. An analysis of Arkansas principals' reports of changes they have made at their schools in response to No Child Left Behind supports the notion that these changes could plausibly affect children's caloric intake or expenditure.

Below, we describe the Arkansas data and accountability program in more detail, and discuss our approach to modeling the role that accountability pressures may play. We then present results from empirical models on obesity rates, as well as some results of a survey of school principals, before concluding.

---

[2] Obesity is defined as having a body mass index (BMI) greater than the 95th percentile of a distribution of age- and sex-specific BMIs from a baseline population from the 1970s. Overweight is defined analogously, with BMI greater than the 85th percentile. The official Arkansas documentation follows CDC convention and labels these thresholds differently as "overweight" and "at risk of overweight," respectively. We will use the more common terms "obese" and "overweight" instead.

II. Policy and Research Background

In 2002, the Federal No Child Left Behind (NCLB) legislation was passed, requiring states to define and implement stringent accountability standards and prescribing increasing penalties for schools that fail to meet their state's standard. Under NCLB in Arkansas, schools are held accountable for the fraction of children in a school who earn a passing score on the state standardized tests in math and literacy.[3] For an elementary school to be deemed passing[4], approximately 30 percent of students in the school had to pass the test in 2002. The percent passing goal ratchets up by about 7 percentage points each year until it reaches 100 percent in 2014. The starting points are slightly lower for higher grades, and the annual increase in the goals are slightly higher in order to reach the Federally mandated goal of 100 percent proficiency by 2014.[5] In addition to the overall percentage passing in the school, each student subgroup – as defined by race, socio-economic status and other educational categories – must meet the same percent passing rate. A school that fails to meet the passing percentage in a certain year can still be awarded passing status through the "safe harbor provision" if the fraction of students failing in the school declines by 10 percent or more between one year and the next. This is a simplified overview of the program. There are more details that are important for the data work but add little to the intuition of the program – such as minimum subgroup size rules and the ability of schools to use a 3-year average percent passing instead of their current pass rate – that are described in more detail in Appendix A.

---

[3] The passing threshold on the Arkansas state test is lower than the threshold on the Nation's Report Card (NAEP) test. In particular, 62 (61) percent of students passed the 4th grade state test in literacy (math), while 28 (26) percent of 4th graders passed the NAEP test. This 34-35 percentage point difference across tests in pass rates is in line with the U.S. average of 32-37 points (Education Week, 2006). The information about Arkansas's NCLB program is taken from http://arkansased.org/nclb/nclb.html (accessed March 25, 2010).

[4] We refer to schools as "failing" schools or "passing" schools. However, the official nomenclature is that schools that we refer to as "failing" schools are in "School Improvement Status."

[5] Annual AYP percent passing goals by grade and subject are listed in Appendix Table 1.

If a school fails to meet the AYP goals for 2 consecutive years, they must implement corrective actions that increase in severity over time if the school continues to fail from allowing students to transfer to a different non-failing school in the district in year 1, to being required to offer supplemental instruction to students in year 2, to more extreme measures such as school restructuring in year 5. In the first few years of NCLB, approximately 25 percent of Arkansas schools were out of compliance with AYP (Blankenship and Barnett, 2006), while 46 percent were failing in 2009 (Arkansas Department of Education, 2009).

A substantial amount of research has documented that schools respond to accountability pressures. Overall, test scores improve – sometimes quite substantially – after accountability is enacted (Carnoy and Loeb 2002, Jacob 2005, Figlio and Rouse 2006, Dee and Jacob 2009, Wong et al. 2010). These gains are made in part by schools' strategic responses that alter *whom* they teach and test – for example strategically assigning kids to special education or English Language Learner status (Cullen and Reback 2006), strategically suspending low performing students (Figlio 2006) or shifting teaching resources toward students on the cusp of passing (Reback 2008, Neal and Schanzenbach 2009). Bokhari and Schneider (2009) find that stricter accountability laws result in more children being prescribed psychostimulates for ADHD (and being diagnosed with the disorder). Schools also strategically alter *what* they teach, shifting effort toward literacy and math and away from un-tested subjects (Jacob 2005, Rouse et al. 2007).

It is intuitive that accountability rules surrounding standardized test performance may affect test scores. In order for accountability pressures to affect children's obesity, it must affect either calories expended or ingested, and there is less prior research that directly addresses this question. The small number of papers that address these issues, however, point to changes in the food or physical activity environment that could affect children's calorie balance. For example, Figlio and Winicki (2005) find that schools that face accountability sanctions increase the number of calories offered in their school lunches during the testing period. Anderson and Butcher (2006b) find that schools in states with accountability measures are more likely to give students access to junk food.[6] Additionally, they find that schools that are more likely to give students access to junk have students with higher BMI.

As for school changes that affect the calorie expenditure side of the equation, the Center on Education Policy (2007) finds 20 percent of school districts have decreased recess time since NCLB was enacted, with an average decrease of 50 minutes per week. Yin (2009) uses cross-state differences in the implementation of accountability laws (pre-NCLB) to explore the effects of accountability on obesity and finds that high school students in states with accountability laws show a significant increase in body mass index and obesity rates. Her analysis also includes an examination of the potential mechanisms by assessing how physical education participation varies with the introduction of accountability laws. She finds evidence that female adolescents' participation in PE classes declines with the advent of accountability. It is worth noting, however, that there is not much compelling evidence that increased PE time decreases obesity rates among high school students. Using state changes in state PE requirements, Cawley et al. (2007) find that while increased time in PE may increase the reported level of physical activity among high school students, there is no evidence that it reduces students' body weight or likelihood of being overweight or obese.[7] Note that our empirical analysis includes elementary school students as well as high school students, thus the mechanisms for changes in physical activity also include changes in recess time.

---

[6] This paper uses a two sample two-stage least squares estimation strategy, and whether or not the school is in a state that has an accountability rule is one of the factors used in the first stage which predicts the fraction of schools in a county that give students access to junk food.

[7] It is surprisingly difficult to find research using credible sources of exogenous variation in physical activity that shows a large and statistically significant link to obesity. Not only is it difficult to find credible sources of exogenous variation in reported *time* in physical activity, it is usually impossible to measure *effort or intensity* of that physical activity. And, as anyone who has tried to track their own calories consumed and expended knows, it is distressingly easy to offset calories burned in physical activity with changes in consumption.

Beyond our school-level exploration of the effect of accountability on obesity rates, we provide more insight into the possible mechanisms with a survey of Arkansas school principals, described below.

III. Background and Data

In order to measure the impact of NCLB pressure on children's obesity status, we construct a unique dataset that merges school-level information on test scores, obesity, and other demographic characteristics from different sources. Using these different sources, we construct our final data set which has 2852 observations for Arkansas schools from 2004 to 2007. The data contain information on 1190 different schools in Arkansas; 447 of the schools include grades 7 and above, the rest include only lower grades.

A. Arkansas Assessment of Childhood and Adolescent Obesity[8]

In 2003 the state of Arkansas passed a sweeping act intended to help combat childhood and adolescent obesity. Although obesity has been increasing nationwide, obesity levels were particularly high in Arkansas. In 2003, about 21 percent of school aged children in Arkansas were obese or overweight, while this figure was about 18 percent for the nation as a whole.[9] A multifaceted coalition came together to address the challenge of childhood obesity, and passed Act 1220 of the 2003 Arkansas General Assembly.[10] A central component of this initiative was the reporting of health risk information to parents (ACHI 2004).

---

[8] This section draws heavily from the yearly reports on the Arkansas Assessment of Childhood and Adolescent Obesity released by the Arkansas Center for Health Improvement. Reports are available online at: www.achi.net

[9] Comparison of Table 1 in ACHI (2004) to NHANES 2003-2004 calculations (http://www.cdc.gov/nccdphp/dnpa/obesity/childhood/prevalence.htm).

[10] The coalition included parents, school nurses, teachers, and administrators, private foundations (including Robert Wood Johnson), physicians, hospitals, universities, Governor Mike Huckabee, the Arkansas Departments of Education and Health, among many others.

The Arkansas Center for Health Improvement (ACHI) spearheaded the effort to collect height and weight information for each school child in the state of Arkansas. This effort included ensuring that each school had the equipment and trained personnel necessary to accurately weigh and measure each child.[11] After children were weighed and measured, a letter then went home to each parent describing the child's BMI, where this fit in the BMI distribution (whether the child was obese, overweight, healthy weight, or underweight), and the type of health risks that might be associated with the child's BMI. Parents of children with an unhealthy weight were urged to consult a physician. An implicit assumption of this effort was that if better information was in the hands of parents, they could make – or help their children make – better informed, more healthful, choices that would improve their weight outcomes. Additionally, an annual public report is produced (available on the ACHI website) with the percent of students who are underweight, normal weight, overweight, and obese at each public school in Arkansas. Thus, due to the Arkansas Assessment of Childhood and Adolescent Obesity, we have panel data on school-level rates of overweight and obesity from 2004 to 2007.

B. School Academic Performance Reports

One of the requirements of NCLB is to make available school-level information on the passing rate, both overall and for student subgroups. School report cards for Arkansas were provided by the Department of Education. These school report cards provide information on the percent of students scored as proficient on the literacy test and the percent scored as proficient on the math test. As described above, a school's AYP designation is determined by the average passing rate of its students overall. In addition, the passing rate of all designated

[11] Training included taking each measure a number of times to ensure accuracy.

sub-groups that have a large enough enrollment in the school must also meet the goal.[12] Student sub-groups are defined by race (for whites, African Americans, Hispanics, etc.), and as low socio-economic status, English language learners, and students with disabilities. If any one of the student subgroups fails to attain AYP, then the entire school is designated as failing to meet AYP.[13]

While these basic rules are straightforward enough, in practice a school can be deemed to meet or fail to meet AYP for several other reasons. For example, even if a school (or subgroup) has a lower fraction of students meeting AYP than the passing standard requires, it still might make AYP through the "Safe Harbor" provision, which allows a school to be deemed as passing if the percentage of failing students (within subject and subgroup) declines by ten percent relative to the prior year. On the other hand, a school will be deemed as failing despite its passing rate if too low a fraction of its students participate in the test, or if attendance or graduation rates are below the target threshold.[14]

There are many details involved in determining AYP status and this means it is very difficult using publicly available data to predict which schools will be deemed as making AYP and which schools will be placed on "School Improvement Status".[15] Since only aggregate data on grade level by subject by subgroup are available to us, we cannot perfectly predict AYP status. However, it seems very likely that if it is difficult for researchers to determine which schools will make AYP and which will not, it is likely also difficult for school administrators to make such a determination as the school year progresses. If school administrators change the way resources are deployed at the school in order to help insure that the school makes AYP

[12] The definition of "large enough" varies by state. In Arkansas, there must be 40 students in the subgroup for it to count toward accountability, or if there are more than 800 students in the school the subgroup must be at least 5 percent of the total enrollment.

[13] Again, although we refer to these as "failing" schools, the official nomenclature is that these schools are in "School Improvement Status."

[14] More information on the Arkansas accountability plan is available at http://arkansased.org/nclb/pdf/accountability_wkbk_021208b.pdf.

[15] It is, in fact, impossible to do this with school level data on aggregate student performance. The official determination is made using micro level data on student performance and detailed information on student characteristics.

status, these changes are unlikely to be starkly discontinuous around the threshold of making and not making AYP. Administrators may have a sense of whether they are likely to be far below or far above the requirements for making AYP, but there is likely to be a substantial fraction of schools where the administrators believe they are on the margin of making or failing to make Adequate Yearly Progress. It is these schools where administrators should have the most incentive to make changes to improve test scores. The data we have access to, percent proficient by grade-level by subject by subgroup, will allow us to identify schools where administrators are likely to be worried about marginally making or missing AYP.

C. Other Demographic Characteristics

Subgroup proficiency rates are reported on the school report cards for groups with a sample size of at least 10, but only subgroups that have sample size of 40 (or 5 percent of total enrollment, whichever is larger) count toward the official accountability rating. As a result, we need to obtain measures of subgroup sample sizes. Although perfect data on this are not publicly available, we are able to estimate population sizes from the Common Core of Data (CCD) for the years 2002-2006.[16] The CCD data report annual school-by-grade enrollment overall and for several subgroups of interest (whites, African Americans and Hispanics). In addition, we were able to proxy for the number of low-income students in each grade by multiplying the school-level fraction of students on free or reduced-priced lunch by the grade-specific enrollment. We used the information on enrollment from the CCD to omit test scores from accountability calculations if they were based on too small of a population and therefore did not count toward accountability status under Arkansas rules. We also use the CCD data to

[16] Because the BMI and test score data extend to 2007 but at the time we began our project the CCD data were not yet available for that year, we assign 2007 CCD data to be the same as the 2006 data.

create school-level demographics on percent nonwhite and percent poor (i.e. receiving free or reduced-price lunch).

IV. Methodology

The main challenge in isolating the effect of NCLB on children's weight outcomes is that low-income children are more likely to score poorly on standardized tests, and are also more likely to be overweight. As a result, a simple cross-sectional relationship between NCLB status and children's weight will be a biased estimate of the impact of accountability pressure on obesity. We take two basic approaches to investigating the possibility that accountability pressures may be contributing to childhood overweight. The first approach assumes that accountability-induced behavioral changes are likely to be greatest among schools that are close to meeting their targets. That is, schools easily meeting the current proficiency levels are unlikely to feel the need to change their behaviors much in the face of accountability, while schools very far from meeting the standards may feel pressure, but will be less likely to think that a small change such as a reduction in recess time or using a portion of the time allocated for gym class to work on mathematics concepts will be useful in addressing their deficiencies. However, we expect that schools with test scores just above and just below the target in year t-1 are the most likely to make the types of changes that could result in more overweight and obese students by year t. By comparing these "marginal" schools to those far away from the thresholds, we can determine if accountability is having an unintended impact on children's health. Our second approach assumes that if a school has seen larger proficiency gains over a year, then that school probably undertook some types of behavioral changes to achieve those gains. As a result, we compare schools that make large gains in a given year to schools that make more modest gains, and measure whether there is a significant difference in obesity rates or changes in obesity rates between them.

With our first empirical approach we address the main methodological challenge by taking advantage of the fact a school being categorized as "marginal" in our framework is not straightforwardly related to the socio-economic status of students in that school. Under NCLB rules, the entire school fails to make AYP if any single subgroup fails to meet the proficiency goal; as a result the school's status is defined by its worst-performing subgroup.[17] For each school-grade-year, for each test, for each subgroup with a group size large enough to count toward the official accountability rating, we standardize the proficiency rates around that year's AYP threshold. For example, for the 4th grade math test, the initial threshold is 40 percent. If a subgroup had a 45 percent proficiency rate, their standardized rate for 4th grade math in the initial year is 5. Similarly if a subgroup had a proficiency rate of 30, their standardized rate for 4th grade math in the initial year is -10. Thus, positive standardized rates represent meeting AYP, while negative ones represent failure to meet AYP. Since AYP is determined at the school level, not grade level, we then aggregate the data to the school-year level. We define school-level proficiency score as the minimum standardized rate across all grades and subgroups in the school. We also maintain the minimum-across-grades math and literacy proficiency rates for the overall school population, as they reflect more generally on the school's overall academic performance. Additionally, based on the CCD, we calculate the percentage of the school's students who are nonwhite, and the percent poor to control for observable demographics.

Our first approach, then, is to consider schools to be "marginal" if they have a minimum subgroup passing rate that is close to the AYP threshold, and to hold constant other factors that are related to student socio-economic status, test scores, and health outcomes. We define "close to the AYP threshold" as being 5 percentage points above or below the

[17] However, as seen below, we investigate the robustness of our results to this choice.

threshold.[18] While schools may have some idea that they are going to be close to making or missing the AYP threshold and change behaviors contemporaneously, we will nonetheless estimate current rates of overweight or obesity based on the previous year's test results to ensure that the school has had time to react to being close to the AYP threshold. We estimate the following model:

$$(1)\ \text{Overwgt}_{st} = \alpha + \beta_1 \text{marginal}_{st-1} + \sum_{i=1}^{4} \gamma_i \text{pctnw}_{st}^{i} + \sum_{i=1}^{4} \delta_i \text{pctpoor}_{st}^{i} + \sum_{i=1}^{4} \varphi_i \text{mathprof}_{st}^{i} + \sum_{i=1}^{4} \theta_i \text{litprof}_{st}^{i} + \upsilon_{st}$$

where $Overwgt$ is the percentage of students in school s at time t who are overweight or obese, $mathprof$ and $litprof$ are the overall current-year proficiency rate in math and literacy, respectively, $pctnw$ is the fraction of the student enrollment that is non-white, and $pctpoor$ is the fraction of student enrollment on free or reduced-price lunch. The variable $marginal$ is an indicator for whether the worst-performing subgroup last year was within 5 points of the passing threshold, and proxies for whether the school might implement short-term policies that might harm student obesity rates, such as curtailing recess. A nice feature of this approach is that we can flexibly control for the overall achievement in a school and identify schools at risk of missing AYP based on their worst-performing subgroup. For example, two schools with similar overall achievement rates may be very similar, but one school is at risk under AYP because of a struggling subgroup, while the other school does not face that risk because they have too few students to have that subgroup count toward the school's accountability status.[19] In some specifications we augment the specification with the lagged dependent variable – that is, the prior year's rate of overweight. Additionally, we can compare the "marginal" schools separately to schools that are farther above and below the threshold. To do this, we add a separate control for whether the minimum subgroup in a

[18] We have also experimented with alternative divisions with qualitatively similar results.
[19] One might be interested in measuring whether there is a particularly strong impact on the rate of overweight of students in the actual subgroups that were marginal. Unfortunately, we do not have access to sufficiently disaggregated data to do such an estimate.

school always scores more than 5 points above the cutoff; in these specifications, the coefficient on “marginal” and the coefficient on “>5 points above” are both relative to the schools for which the worst performing subgroup is more than 5 points below the threshold. This allows us to compare schools in the marginal group separately to schools in each “tail” of performance (for the worst performing subgroup). To further probe these relationships, we also estimate these equations based on whether *any* subgroup (and not just the minimum subgroup) is on the margin of passing in the prior year, and whether any subgroup has *ever* been on the margin of passing.

Our second approach focuses on the relationship between overall school performance and obesity rates, not subgroup performance. The school’s change in math proficiency rate is added to the school’s change in literacy proficiency rate to define the total proficiency gain over the prior year. Gains in proficiency scores might occur for for several reasons, including: 1)mean reversion, if the prior years scores were aberrantly low; 2) changes in the type of student who took the test; or 3) changes in the way the school allocates its resources that are intended to improve test scores. The idea with this specification is that if schools that managed test score gains did it by changing the way it resources are allocated, then those changes may have the potential to affect children’s weight as well as test performance. As with the models focusing on the marginal schools, the main threat to validity in this approach is that factors that are correlated with test performance, like the composition of the student body, may also be correlated with obesity rates. In these specifications, we also control flexibly for current demographics and academic performance. Additionally, it is important to keep in mind that any mechanical change in the student body that would tend to improve test scores would also be expected to reduce obesity rates, since higher socio-economic status students, on average, have higher test scores and lower BMI. Thus, any unmeasured socio-economic characteristics in these regressions would be likely to bias the coefficient toward finding a negative relationship between gains in test performance and obesity rates.

V. Results

A. The Role of NCLB on Students' Overweight Status

We start by examining some basic descriptive statistics using the final data set. Looking at the top rows of Table 1, in the first column we see that almost a fifth of school-by-year cells in Arkansas are classified as being "marginal" (i.e. the school's minimum scoring grade-test-subgroup cell falls within 5 points of the AYP threshold), with the remaining school-by-year cells being almost evenly split between being more clearly in failing territory (scoring more than 5 points below the AYP threshold) and passing territory (scoring more than 5 points above the AYP threshold). Moving down the column, we see that overall in the average school-by-year observation 38.3 percent of students are overweight or obese, 53.6 percent receive free or reduced-price lunch and 26.9 percent are nonwhite. Finally, on average schools met the target English proficiency rate by 13.5 percentage points and the target math proficiency rate by 12.6 points, but over the previous year the total proficiency rate actually dropped by almost one percentage point because the target passing rate increased in line with the schedule presented in Appendix Table 1. The second and third columns present characteristics for the average school which is marginal in terms of meeting the AYP goal and the average school which is not. Keeping in mind that "non-marginal" schools are comprised of schools with minimum scoring groups that are both farther below and above the threshold for AYP, one sees that there is a very slightly higher rate of overweight and obesity at the marginal schools, and that this does not appear to be driven primarily by demographic differences. Marginal schools have a very slightly lower rate of free and reduced-price lunch recipients and a much lower nonwhite percentage. In terms of overall school performance, both marginal and non-marginal schools are fairly similar, with the overall test performance of marginal schools actually being a bit better in both English and Math. Note however, the marginal schools had higher proficiency rate growth over the past year, up almost 3 percentage points, while non-marginal schools lost almost 2 percentage points.

Table 2 presents the results of the basic model described in equation (1) above, with schools categorized as "marginal" by their lowest-performing subgroup and percent overweight (including obese) as the dependent variable. Standard errors are clustered by school. Starting with column (1), we see that marginal schools in Arkansas have a rate of overweight that is just under 1 percentage point higher than the non-marginal schools. When we augment the controls with the lagged value of overweight in column (2), schools that were marginal in the previous year see an increase in the rate of overweight that is almost 0.5 percentage points larger than those that were non-marginal.[20] Columns (3) and (4) continue to control for the lagged rate over overweight, and compare the marginal group of schools separately against the top and bottom of the distribution.[21] Marginal schools have 0.59 percentage points more overweight students compared to schools where the worst performing subgroup is more than 5 points above the AYP threshold, and this difference is statistically signifcant at conventional levels. Note, however, that the point estimate in column (4) is also positive and indicates that marginal schools have 0.361 percentage points more overweight students than schools where the worst performing subgroup is more than 5 points below the AYP threshold. While this coefficient is not statistically different from zero, it also is not statistically different from the coefficient estimated in column (3). If being

[20] If we simply replace the flexible function of overall test performance with the lagged value of overweight, the results are very similar to those presented (e.g. the coefficient (standard error) for column (2) is 0.519 (0.200)), but since the test score variables are jointly significant, we retain them in the model presented.
[21] We recognize that column (4) can be derived from column (3), but estimate both models for ease of determining statistical significance.

a "marginal" school were straightforwardly correlated with socio-economic status, then we would expect that marginal schools would have children with worse health outcomes than schools in the "more than 5 points above AYP" group and better health outcomes than schools in the "more than 5 points below AYP." Instead, schools in the "marginal" category have students with worse health outcomes than schools where the worst performing subgroup is farther away from the threshold, either above or below. This strongly suggests that "marginal"schools are undertaking actions to try to make adequate yearly progress, and that these actions may have unintended consequences.

Our alternate approach is based on the gains in proficiency levels across years. Schools with larger year-to-year gains have likely done something to improve scores, and in doing so may have adopted policies that result in increased overweight. The specification in columns (5) and (6) uses the sum of the math and reading proficiency gains experienced by the school overall (not any particular subgroup). In column (5), we see a significantly positive estimate indicating that a school making a bigger proficiency gain under NCLB over the past year has a slightly higher rate of overweight. Because Arkansas set up its schedule for AYP to involve equal increments each year until reaching 100 percent proficiency, a typical school that just met its goal in one year would be expected to add about another 7 or 8 percentage points to each of the literacy and math proficiency rates (for a total gain of about 15) in order to make AYP again the next year. Based on the results in column (5), a school making this size gain would be expected to have a rate of overweight that is about 0.43 percentage points higher than a school not making any gains, which is similar to the effect for marginal schools found in column (2). However, when we control for last year's rate of overweight in column (6), the estimated effect of the total gain becomes very small and insignificant. Note, however, that if a school is regularly making large gains, changes resulting in weight gain may have been put in place in the past, with no additional changes in policies or weight occurring in the current year. Controlling for the lagged value of percent overweight will absorb the effects of any past changes in how the school conducts business. In such a case, we would expect the coefficient in column (6) to be zero.

Table 3 repeats the specifications in Table 2, with the obesity rate as the dependent variable and schools classified as marginal by their lowest performing subgroup. The results are quite similar, but smaller in magnitude. Overall, students in schools where the minimum scoring subgroup was within 5 points of the proficiency threshold in the prior year are 0.6 percentage points more likely to be obese (see column 1). When the lagged obesity rate is controlled (column 2), the coefficient falls by about half but remains statistically significant. When we compare marginal schools to the top and bottom of the distribution separately in columns (3) and (4), marginal schools are only statistically significantly more likely to have obese students when compared to the higher performing schools, however, both point estimates are positive. Again, the point estimates indicates that "marginal" schools have more obese students than schools that are either farther below or farther above the NCLB threshold. Turning to our alternate specification, schools with larger annual test score gains see higher obesity rates, but the results are not robust to controlling for lagged obesity rates (columns 5 and 6). In the remaining tables and figures, we limit the analysis to the school's rate of overweight, though results for obesity are qualitatively similar.[22]

Table 4 presents some robustness checks. Columns (1) – (3) change the independent variable of interest to whether any subgroup – not just the minimum-scoring subgroup – falls within 5 points of the passing threshold.[23] The idea here is that even if a school has no real chance to meet AYP for its worst subgroup, if there is a marginal group at the school, policy changes may take place to get that group over the AYP threshold. The results are fairly similar

[22] Results for Table 4 and Figure 1 with obesity rate as the dependent variable are available upon request.

[23] This definition results in a much larger number of school-year-cells being defined as marginal – about 47 percent.

to those in Table 2, but slightly smaller in magnitude once the prior rate of overweight is controlled. Columns (4) – (6) define the independent variable of interest as whether any subgroup has been marginal in any prior year.[24] The intuition behind this final approach is that schools may have implemented policies to improve performance after a year in which they were marginal, and have continued to implement those policies in future years when they were no longer on the margin. When this specification is employed, the results are still similar but with slightly larger point estimates than those found in Table 2.

Overall, then, Tables 2 through 4 suggest that accountability pressures may have unintended side effects on children's health. In particular, the coefficients are robust to controlling for the prior year's rate of overweight, indicating that schools that are marginal on test scores see a discrete increase in overweight. We further build the case that this likely reflects a causal relationship by next turning to an event study analysis to investigate the time pattern of increases in the rate of overweight. We then report results of a survey we conducted that asked Arkansas school principals to about their policies and practices, and how those have changed after the implementation of NCLB.

B. Event Study

The timing of increases in the rate of overweight can be estimated more directly using an event study analysis. Specifically, we fit the following equation:

$$(2)\quad \text{Overwgt}_{st} = \alpha + \sum_{-2}^{5} \beta_i 1(\tau_{st} = i) + \sum_{i=1}^{4} \gamma_i \text{pctnw}_{st}^{i} + \sum_{i=1}^{4} \delta_i \text{pctpoor}_{st}^{i} + \sum_{i=1}^{4} \varphi_i \text{mathprof}_{st}^{i} + \sum_{i=1}^{4} \theta_i \text{litprof}_{st}^{i} + \upsilon_{st}$$

where τ_{st} denotes the event-year, defined so that τ=0 in the first year that a school's minimum subgroup scores within 5 points of the passing threshold (i.e. the first year a school is "marginal.") τ=1 denotes the first year after a school is declared to be marginal, and so on. In

[24] This definition results in just over 72 percent of the school-year-cells being defined as marginal, a large increase that is not surprising.

the years in which $\tau<0$, the school was not yet marginal. The coefficients are measured relative to the omitted coefficient $\tau\leq-3$. Schools that never have marginal status are still in the dataset, but they only help identify the relationship between covariates and overweight since the vector of τ's are all zero for such schools. We include all available observations, even though we do not observe the full range of event-time for all schools.

Figure 1 plots the event-year coefficients from estimating equation (2). Although the coefficients are imprecisely estimated, the figure shows relatively flat trends in the years prior to first being marginal under NCLB, then a moderate jump in the following year. There is some additional upward movement in subsequent years which may reflect the intensifying of efforts to improve test scores that have negative spillover effects on students' weight. Overall this provides further evidence that the timing lines up well with the hypothesis that accountability pressures under NCLB may be causing the increases in the rate of overweight.

C. The Effect of NCLB on School Policies

At the start of this paper, we presented several theories on why accountability pressures might have an effect on students' obesity: there may be changes in the school that unintentionally affect either students' expenditures of energy (like a shift away from recess toward test preparation) or in-take of energy (like sales of junk food to raise extra money for test preparation) . In order to evaluate the validity of these theories, we fielded a survey to the principals of the Arkansas schools used in our study. Principals were invited via email to participate in an online survey, were offered a $10 incentive for participation in the form of a gift certificate to an online bookstore, and were reminded via email three times after the original request. Approximately 5 percent of the sample either did not have a valid email or had previously opted out of receiving survey invitations. Of the remainder, approximately 20 percent responded to the survey, resulting in a dataset of 191 schools; however, not every principal answered every question. We matched the survey responses with our data set on test scores, demographics, and obesity rates to examine how schools that responded to the survey compare to other schools in Arkansas. It is clear that the survey data set is not a random sample of Arkansas schools: respondents are principals at relatively better-off schools as measured by demographic characteristics, test scores, and the likelihood that they are passing under NCLB (see Appendix Table 2 for descriptive statistics). Nonetheless, the responses are informative about the types of changes that schools implemented in response to NCLB.

Table 5 presents results from the survey. First, we see that 86 percent of schools report increasing spending since NCLB went into effect. Ten percent report raising funds through sales of foods or beverages, and 17 percent report increased funding from vending machines or pouring rights contracts. At the same time, 38 percent report using food as a reward for students, either in the form of parties in school or coupons that can be redeemed for food items outside of school. Thus, we see some evidence that school changes in response to NCLB may be increasing calories consumed.

There is also evidence that changes in response to NCLB may be reducing time spent in physical activity. Schools report having increased instructional time on tested subjects (English/language arts and math) by almost an hour and a half per week. To make room for this additional instruction, 38 percent of schools report that time spent on non-academics has been reduced. Schools in the sample currently spend an average of almost 190 minutes (or about 3.2 hours) per week in PE classes and about 127 minutes (or about 2.1 hours) at recess. Taking both activities combined for all schools, students spend an average of 304 minutes (or about 5.1 hours) per week in PE and recess. This is substantially higher than the 150 minutes per week of "physical activity" required by the state.[25] Since NCLB, only 3 percent report reducing time spent in PE, while 12 percent report reducing time spent at recess.[26] While only a small fraction of schools report decreasing the amount of time spent in physical activities, over 80 percent have not had increases since NCLB. In addition, a substantial fraction of schools report offering before or after school sessions for instruction, or increasing the length of the school day. To the extent that these sessions may be displacing active play or other physical activity, this may also contribute to weight gain among students. Finally, the overwhelming majority of schools report engaging in some explicit test preparation. Many (over 25 percent) report having test prep sessions after school, and 5 percent report conducting test preparation during recess, lunch or PE class.

Thus, there is evidence that NCLB may have reduced the opportunities for physical activity in Arkansas schools, although the changes do not seem overwhelming. That said, for the 22 schools for which we can calculate the drop in recess time, it averages 51 minutes per week.[27] A simulation of a weekly reduction of that magnitude over the course of a 36 week school year implies that the rate of overweight of a population would increase by about 1.6 percentage points.[28] Interestingly, an increase of about 0.5 percentage points, as is found in Tables 2 and 3, is simulated to be achieved with less than a 20 minutes reduction in physical activity per week.

While this snapshot of the state of Arkansas school policies is quite interesting, we would ideally like to know whether the policies differ by our definitions of a "marginal"

---

[25] For K-8 students, at least 60 minutes of weekly physical activity must be in PE class. The remainder can include recess and after- or before-school activities. See http://arkansased.org/programs/csh/physed_faqs.html for more details.
[26] Note that not all schools answered both of these questions. In particular, only elementary schools tend to have recess, so while 186 schools reported time spent in PE, only 129 reported time spent at recess.
[27] The decrease in PE time among those reporting a decrease is 33 minutes per week.
[28] The simulation uses children age 6-16 from the NHANES who have an overweight rate of 36.4 percent. Using age-gender-weight-specific formulas for base metabolic rate (BMR), and assuming the children are in caloric balance before the change in weekly activity, it is possible to simulate the caloric imbalance and subsequent weight gain.

school. To this end, we are hampered by our small sample sizes, as well as differential response rates across groups and some issues of timing. The remaining columns of Table 5 break out the responses by the school's most recent (2007) test performance relative to the passing threshold. Although policy changes that might contribute to obesity have been implemented in schools across the board, there is some suggestive evidence that the food and exercise environments have changed more at lower-performing schools.

Perhaps the most reasonable approach would be to regress a policy such as reduction in PE or recess on the marginal definition based on any group being close to the AYP threshold at any time in the past (along with the basic demographic and school controls).[29] The drawbacks to this approach are twofold. First, for many of the policy variables, the samples are really quite small, potentially limiting our ability to find significance for coefficients that correspond to large effects. Second, a given school may have had a group scoring within 5 points of the AYP goal after 2007, meaning they really should be coded as being marginal in terms of the 2009 survey data. This measurement error would lead to attenuation bias in our estimates.

With these caveats in mind, it is worth noting that the coefficient on being marginal under NCLB on combined minutes of recess and PE is -29, but the estimate is not significantly different from zero. This point estimate is, however, quite in line with the type of change that was simulated to produce the estimated change in the rate of overweight.[30] No other models have estimates that are significant at conventional levels.

[29] Note that controlling for the lagged rate of overweight does not really make sense in this context, since the inclusion of the lagged dependent variable in Tables 2 and 3 was meant to control for unobserved determinants of current overweight.

[30] A drop in activity of 30 minutes per week is simulated to increase the rate of overweight by 0.86 percentage points, which is close to the estimate in column (1) of Table 2. Note though that an activity drop of 76 minutes per week would imply just over a 2 percentage point increase in overweight, which is in line with column (4) of Table 3.

V. Conclusions and Further Avenues for Research

Through the No Child Left Behind Act, schools face increasing pressure to improve performance on standardized tests. Past research has clearly documented that school behaviors are affected by accountability pressures (e.g. Cullen and Reback 2006, Figlio 2006, Jacob 2005, Neal and Schanzenbach 2009, Rouse et al. 2007). Since schools are graded based primarily on standardized test scores, but not on other student outcomes such as children's health, schools facing accountability pressure may well make decisions designed to increase test scores that have unintended negative consequences for children's weight.

This paper adds to the small amount of evidence on the effect of accountability on the food and physical activity environment (e.g. Figlio and Winicki 2005, Yin 2009, Anderson and Butcher 2006b). By focusing on schools in Arkansas that are close to meeting AYP standards, we find schools that were on the margin of passing have about a 0.5 percentage point higher rate of overweight in the following year. This effect is based on models that also control for the school's lagged rate of overweight. Models without the lagged dependent variable find larger effects. Because it is clear that schools with lower test scores also have students with worse socioeconomic outcomes it is important that we are not just comparing poor performing schools to better performing schools. Rather, being a "marginal" school has a positive point estimate for rates of obesity and overweight whether the marginal schools are compared to those scoring farther above or farther below the AYP threshold.

These results present prima facie evidence that the NCLB accountability rules may have unintended adverse consequences for student health. As a result, parents and school administrators, and policy-makers should keep in mind the potential for impacts on children's health as they consider how to reallocate school resources in pursuit of test score gains.

References

Anderson, Patricia M. and Kristin F. Butcher, "Trends in and Potential Causes of Childhood Obesity," *The Future of Children: Child Overweight and Obesity*, vol. 16, no. 1, Spring 2006a, Brookings Institution Press.

Anderson, Patricia M. and Kristin F. Butcher, "Reading, Writing, and Refreshments: Do School Finances Contribute to Childhood Obesity?" *Journal of Human Resources,* vol. 41, no. 3, Summer 2006b, pp.467-494.

Arkansas Center for Health Improvement (ACHI), "The Arkansas Assessment of Childhood and Adolescent Obesity," 2004.

Arkansas Center for Health Improvement, "The 2005 Arkansas Assessment of Childhood and Adolescent Obesity: Online State Report," September 2005.

Arkansas Center for Health Improvement, "The Arkansas Assessment of Childhood and Adolescent Obesity – Tracking Progress: Online State Report, Year 3 (Fall 2005-Spring 2006)," 2006.

Arkansas Center for Health Improvement, "Assessment of Childhood and Adolescent Obesity in Arkansas: Year Four (Fall 2006 – Spring 2007)," 2007.

Arkansas Department of Education (2009). 2008-2009 School Improvement List, Press Release. http://arkansased.org/communications/pdf/ayp_release_100909.pdf

Blankenship, Virginia H. and Joshua H. Barnet, "AYP in Arkansas: Who's on Track?" *Arkansas Education Report* 3(2), Summer 2006.

Bokhari, Farasat A. S. and Helen Schnedier, "School Accountability Laws and the Consumption of Psychostimulants," *Journal of Health Economics,* forthcoming.

Carnoy, Martin and Susanna Loeb (2002), "Does External Accountability Affect Student Outcomes? A Cross-State Analysis," Educational Evaluation and Policy Analysis 24(4): 305 – 331.

Center on Education Policy. *Choices, Changes, and Challenges: Curriculum and Instruction in the NCLB Era*. Washington DC: 2007.

Cullen, Julie Berry and Randall Reback (2006). "Tinkering Toward Accolades: School Gaming under Performance Accountability System." *Advances in Applied Microeconomics* 14: 1 – 34.

Dee, Thomas and Brian Jacob (2009) "The Impact of No Child Left Behind on Student Achievement." *NBER Working Paper No. 15531.*

Education Week (2006) *Quality Counts 2006* (available online at http://www.edweek.org/ew/toc/2006/01/05/index.html)

Figlio, David (2006) “Testing, Crime and Punishment.” *Journal of Public Economics*, 90(4-5): 837 – 851.

Figlio, David and Cecilia Elena Rouse (2006). “Do Accountability and Voucher Threats Improve Low-Performing Schools?” *Journal of Public Economics*, 90(1-2): 239 – 255.

Figlio, David and Joshua Winicki, “Food For Thought: The Effects of School Accountability Plans on School Nutrition,” *Journal of Public Economics*, Volume 89, Issues 2-3, February 2005, Pages 381-394.

Going, Scott B., Timothy G. Lohman, and Harold B. Falls, “Body Composition Assessment” in Fitnessgram/Activitygram Reference Guide (Gregory J. Welk. and Marilu D. Meredith eds.), Dallas, TX: The Cooper Institute.

Jacob, Brian (2005) “Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in Chicago Public Schools.” *Journal of Public Economics* 89(5-6): 761 – 796.

Neal, Derek and Diane Schanzenbach (2009) “Left Behind by Design: Proficiency Counts and Test-Based Accountability.” *Review of Economics and Statistics*, 92(2): 263 ‐ 283.

Reback, Randall, “Teaching to the Rating: School Accountability and Distribution of Student Achievement,” *Journal of Public Economics*, 92 (2008): 1394 – 1415.

Rouse, Cecelia Elena, Jane Hannaway, Dan Godhaber and David Figlio (2007) “Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure.” *NBER Working Paper No. 13681*.

Schanzenbach, Diane Whitmore “Do School Lunches Contribute to Childhood Obesity?,” *Journal of Human Resources*, vol. 44, no. 3, Summer 2009, pp.684-709.

Yin, Lu, “Are School Accountability Systems Contributing to Adolescent Obesity?” mimeograph, 2009.

Appendix A: Determining Adequate Yearly Progress in Arkansas

During the time period studied, high-stakes math and literacy tests were given in grades 4, 6 and 8, and in algebra, geometry and 11th grade literacy. The publicly available school report card data report the percentage of students with scores on each test in various ranges (below basic, basic, proficient and advanced), separately for math and literacy. For each school, this information is reported not only for the "combined population," but also for a variety of subgroups defined by race/ethnicity (African-American, Caucasian, Hispanic) and special classification status (economically disadvantaged, limited English proficiency, students with disabilities).[31]

Average proficiency rates are reported in the public-use data for all group-by-test cells with a sample size of at least 10. However, the cell does not count toward proficiency unless there are at least 40 students in the cell (for schools with fewer than 800 students) or the cell represents at least 5 percent of enrollment (for schools with 800 or more students). Since sample sizes are not included in the report card data, we merge on information from the Common Core of Data (CCD). The CCD reports for each grade and each year the total number of students, as well as the number of African-American, Caucasian and Hispanic students in the grade. We use these data to code whether a subgroup is large enough to count toward AYP. The CCD only reports the total number of economically disadvantaged students in the school (as measured by the fraction on free or reduced-price lunch). To estimate the grade-specific population, then, we multiply the school-level fraction that is disadvantaged by the grade-specific enrollment. No information is available from either source for the number of students with disabilities or limited English proficiency. After trying several approaches to proxy for whether the sample sizes in these two groups would have been large enough to

[31] The report cards also list separate proficiency rates by gender, and for migrants, but these subgroups do not count toward AYP.

count toward AYP, we dropped both of these subgroups from the analysis. Note that in the end this improved our predictive power when we modeled the official AYP status as coded by the state as a function of the program rules and the public-use data.

If fewer than 95 percent of students are tested in math and reading, the school automatically fails AYP. The report card lists whether the school meets the 95 percent threshold, and if not it reports the school-wide percentage of students tested. The denominator of this calculation is all students enrolled in the grade or course at the time of testing.

If the proficiency rate in the cell is greater than or equal to that year's passing target (termed "Annual Measurable Objective" or AMO and displayed in Appendix Table 1), and at least 95 percent of the eligible students are tested, then that group-by-test meets AYP. The school as a whole meets AYP if both its combined population and all subgroups that count toward proficiency meet the AYP requirements. If any subgroup fails to meet that year's AMO (and does not meet AYP through the safe harbor provision described below), then the school has failed to meet AYP. As a result, the binding constraint is the proficiency rate of the worst-performing subgroup that is large enough to count toward the rating. Our empirical approach reflects this by defining each school by the performance of its subgroup with the lowest passing rate.

A school can have its AYP status determined based either on the current year's performance or the average over the prior 3 years. The metric chosen can vary across years, but must be consistent across all subgroups within the same year. Our empirical work defines being marginal under AYP based only on the current year's results. Robustness checks that use the maximum of the current year and the 3-year average are substantially the same.

Schools can also be classified as meeting AYP requirements even if the passing rate is below the year's AMO under the safe harbor provision, as long as at least 95 percent of students are tested in the year. Under safe harbor, if the fraction of students scoring below proficient in each subgroup declines by at least 10 percent from the prior year, then the school meets AYP requirements for that year. In addition, the school meets AYP under safe harbor if its observed increase in proficiency rates falls within a 95 percent confidence interval of the safe harbor goal. Over this time period, the minimum increase necessary to meet AYP under safe harbor was always greater than 5 points, which is the cutoff for being a "marginal" school in the analysis.

In order to meet AYP, schools must also satisfy the appropriate "secondary indicator." For high schools, graduation rate must be at least 70 percent, and for non-high schools attendance rates must be at least 91.13 percent. Among schools that are passing due to test rates, relatively few fail due to the secondary indicator. We do not use this information in our analysis.

Figure 1: Estimated Impact of “Marginal” Status under NCLB on Overweight Rate, by Year

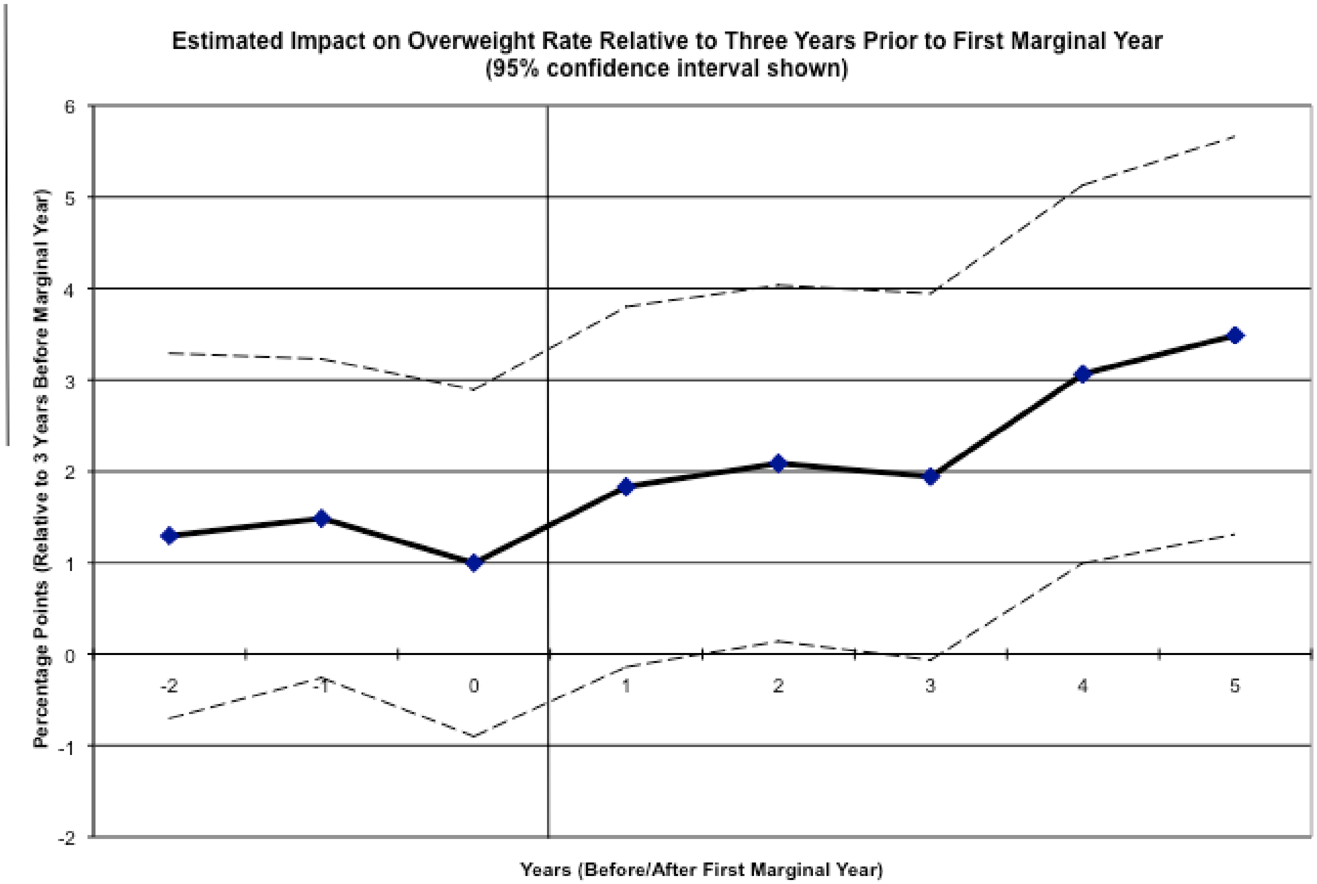

Table 1: Summary Statistics

	Full Sample	Marginal Schools (w/in 5 pts of AYP)	Non-marginal Schools
Marginal	0.183	1	0
	(0.387)	--	--
Below AYP	0.405	0	0.496
	(0.491)	--	(.500)
Above AYP	0.412	0	.504
	(0.492)	--	(.500)
Overweight Rate	38.3	39.0	38.1
	(5.8)	(5.2)	(6.0)
Obesity Rate	21.1	21.6	20.9
	(4.7)	(4.4)	(4.8)
Percent Free/Reduced-	53.6	52.5	53.8
Price Lunch	(18.9)	(16.4)	(19.4)
Percent Nonwhite	26.9	20.8	28.2
	(28.3)	(24.7)	(28.9)
English Proficiency	13.5	15.0	13.2
Rate	(16.8)	(15.3)	(17.1)
Math Proficiency	12.6	13.1	12.5
Rate	(18.8)	(15.2)	(19.5)
Annual Gain	-0.992	3.0	-1.88
	(27.3)	(25.3)	(27.6)
Observations	2852	523	2329

Notes: Means (std. deviations) are shown. "Marginal" is defined as the lowest scoring subgroup in the school being within 5 points of the AYP goal in the prior year, while "Below AYP" and "Above AYP" are defined as that subgroup being more than 5 points below or above the AYP goal in the prior year, respectively. Overweight (obesity) rate is based on students being above the 85^{th} (95^{th}) percentile of the age and gender-specific BMI distribution set by the Centers for Disease Control. "Annual Gain" is defined as the overall school combined increase in English and math proficiency rates over the prior year. Observations are at the school-year level.

Table 2:
Effect of Accountability Pressures on the Overweight Status of Students

	(1)	(2)	(3)	(4)	(5)	(6)
Marginal minimum subgroup	0.934*** (0.27)	0.485** (0.21)	0.591** (0.23)	0.361 (0.23)		
Minimum subgroup >5 points below AYP			0.230 (0.21)			
Minimum subgroup >5 points above AYP				-0.230 (0.21)		
Annual Gain					0.029*** (0.0036)	0.00009 (0.0032)
Lag Overweight Rate		0.708*** (0.017)	0.705*** (0.017)	0.705*** (0.017)		0.709*** (0.017)
Constant	34.68 (150)	-585.2*** (194)	-598.8*** (195)	-598.5*** (195)	15.87 (151)	-564.5*** (194)
Observations	2852	2090	2090	2090	2869	2078
R-squared	0.28	0.65	0.65	0.65	0.29	0.65

*** p<0.01, ** p<0.05, * p<0.1

Notes: Standard errors (in parentheses) are corrected for within school correlation. The dependent variable is the school-wide rate of overweight, where overweight is defined as being above the 85^{th} percentile of the age and gender-specific BMI distribution set by the Centers for Disease Control. All models include a time trend and four powers of the overall English proficiency rate relative to AYP, four powers of the overall math proficiency rate relative to AYP, four powers of the percent of students who are nonwhite, and four powers of the percent of students who are poor. "Marginal" is defined as the lowest scoring subgroup in the school being within 5 points of the AYP goal in the prior year, while "Below AYP" and "Above AYP" are defined as that subgroup being more than 5 points below or above the AYP goal in the prior year, respectively. "Annual Gain" is defined as the overall school combined increase in English and math proficiency rates over the prior year.

Table 3:
Effect of Accountability Pressures on the Obesity Status of Students

	(1)	(2)	(3)	(4)	(5)	(6)
Marginal minimum subgroup	0.637***	0.281*	0.441**	0.094		
	(0.198)	(0.160)	(0.180)	(0.180)		
Minimum subgroup >5 points below AYP			0.347**			
			(0.166)			
Minimum subgroup >5 points above AYP				-0.347**		
				(0.166)		
Annual Gain					0.026***	0.003
					(0.003)	(0.002)
Lag Obesity Rate		0.709***	0.703***	0.703***		0.709***
		(0.016)	(0.016)	(0.016)		(0.016)
Constant	119.2	-461.8***	-482.6***	-482.3***	87.83	-472.9***
	(119)	(146)	(146)	(146)	(121)	(145)
Observations	2,849	2,087	2,087	2,087	2,866	2,075
R-squared	0.325	0.668	0.669	0.669	0.342	0.671

*** p<0.01, ** p<0.05, * p<0.1

Notes: Standard errors (in parentheses) are corrected for within school correlation. The dependent variable is the school-wide rate of obesity, where obesity is defined as being above the 95^{th} percentile of the age and gender-specific BMI distribution set by the Centers for Disease Control. All models include a time trend and four powers of the overall English proficiency rate relative to AYP, four powers of the overall math proficiency rate relative to AYP, four powers of the percent of students who are nonwhite, and four powers of the percent of students who are poor. “Marginal” is defined as the lowest scoring subgroup in the school being within 5 points of the AYP goal in the prior year, while “Below AYP” and “Above AYP” are defined as that subgroup being more than 5 points below or above the AYP goal in the prior year, respectively. “Annual Gain” is defined as the overall school combined increase in English and math proficiency rates over the prior year.

Table 4:
Robustness Checks for the
Effect of Accountability Pressures on the Overweight Status of Students

	(1)	(2)	(3)	(4)	(5)	(6)
	Any subgroup within 5 points last year			Any subgroup within 5 points in any previous year		
Marginal	1.463***	0.382**	0.428**	2.605***	0.571***	0.703***
	(0.22)	(0.17)	(0.20)	(0.32)	(0.19)	(0.20)
Below AYP			0.164			0.628
			(0.28)			(0.44)
Lag Overweight Rate		0.705***	0.704***		0.699***	0.697***
		(0.02)	(0.02)		(0.02)	(0.02)
Constant	-1.429	-594.9***	-599.4***	459.1***	-460.9**	-445.2**
	(151)	(195)	(195)	(156)	(198)	(198)
Observations	2852	2090	2090	2852	2090	2090
R-squared	0.29	0.65	0.65	0.31	0.65	0.65

*** p<0.01, ** p<0.05, * p<0.1

Notes: Standard errors (in parentheses) are corrected for within school correlation. The dependent variable is the school-wide rate of overweight, where overweight is defined as being above the 85^{th} percentile of the age and gender-specific BMI distribution set by the Centers for Disease Control. All models include a time trend and four powers of the overall English proficiency rate relative to AYP, four powers of the overall math proficiency rate relative to AYP, four powers of the percent of students who are nonwhite, and four powers of the percent of students who are poor. In columns (1) – (3) "Marginal" is defined as any subgroup in the school being within 5 points of the AYP goal in the prior year, while "Below AYP" is defined as any subgroup being more than 5 points below the AYP goal (provided no subgroup was within 5 points) in the prior year. In columns (4) – (6) "Marginal" is defined as any subgroup in the school being within 5 points of the AYP goal in any of the prior years, while "Below AYP" is defined as any subgroup being more than 5 points below the AYP goal (provided no subgroup was within 5 points) in any of the prior years.

Table 5:
Results from a Survey of School Principals

	(1)	(2)	(3)	(4)
		by NCLB category		
	Overall	Missed by more than 5 points	within 5 points of cutoff	Made by more than 5 points
% total responses	1.000	0.217	0.111	0.672
Spending has increased =1	0.864	0.897	0.895	0.847
If increased spending, from what source? (Check all that apply)				
Food fund raisers	0.105	0.114	0.059	0.110
Vending machines/pouring rights	0.171	0.343	0.118	0.120
Any extrinsic rewards for student success on test? 1=yes	0.619	0.732	0.667	0.575
If extrinsic rewards, what types? (Check all that apply)				
Food	0.376	0.400	0.429	0.356
Weekly minutes spent on subjects:				
ELA	589.4	575.4	451.2	616.9
Math	400.4	375.2	369.0	413.8
Lunch	191.2	192.2	197.6	189.8
PE	188.9	208.6	237.9	174.4
Recess	126.5	70.1	79.3	151.9
PE + Recess	303.7	265.1	317.1	314.0
Change in min/week since NCLB:				
ELA	51.7	62.8	55.6	47.0
Math	35.0	39.5	52.5	30.1
Lunch	-0.9	-0.3	-4.2	-0.5
PE	13.5	11.9	11.7	14.4
Recess	-6.5	-5.5	3.0	-8.3
PE + Recess	5.8	6.3	12.1	4.6
Where did extra instructional time come from? (Check all that apply)				
Before/after school sessions (optional or required)	0.450	0.585	0.429	0.409
Reduced time on non-academic activities	0.381	0.390	0.333	0.386
Increased length of school day	0.111	0.171	0.095	0.094
Increased length of school year	0.026	0.024	0.048	0.024
School does some test prep =1	0.887	0.950	0.905	0.864

When does test prep occur? (Check all that apply)				
After school	0.255	0.474	0.158	0.194
During other non-academic classes	0.121	0.105	0.263	0.102
During recess, lunch or PE	0.055	0.053	0.053	0.056
Before school	0.055	0.079	0.053	0.046

Notes: Not all questions were answered by each of the 191 schools.

Appendix Table 1: Proficiency Rate Goals in Arkansas, by Grade Span and Year

	Grades K-5		Grades 6-8		Grades 9-12	
	Literacy	Math	Literacy	Math	Literacy	Math
2002	31.80	28.20	18.10	15.30	19.50	10.40
2003	37.48	34.18	24.93	22.36	26.21	17.87
2004	43.16	40.16	31.76	29.42	32.92	25.34
2005	48.84	46.14	38.59	36.48	39.53	32.81
2006 amendment changed schedule						
2006	42.40	40.00	35.20	29.10	35.50	29.20
2007	49.60	47.50	43.30	38.06	43.56	38.05
2008	56.80	55.00	51.40	47.02	51.62	46.90
2009	64.00	62.50	59.50	55.98	59.68	55.75
2010	71.20	70.00	67.60	64.55	67.75	64.60
2011	78.40	77.50	75.70	73.41	75.81	73.45
2012	85.60	85.00	83.80	82.28	83.88	82.30
2013	92.80	92.50	91.90	91.14	91.94	91.15
2014	**100**	**100**	**100**	**100**	**100**	**100**

Appendix Table 2: Descriptive Statistics of Survey Sample vs. Overall Sample

	Overall	**Have survey**	**No survey**
Demographics			
Percent non-white	0.264	0.181	0.280
Percent Free/RP lunch	0.558	0.504	0.570
Majority Non-white=1	0.198	0.087	0.221
Majority poor = 1	0.602	0.516	0.620
School eligible for Title I funds	0.749	0.721	0.755
Receiving Schoolwide title I	0.833	0.782	0.843
Total enrollment	409.0	451.2	400.4
Geography			
City=1	0.191	0.206	0.188
Town=1	0.192	0.152	0.201
Suburb=1	0.079	0.109	0.073
Rural=1	0.537	0.533	0.538
% Overweight or At-Risk	0.382	0.368	0.385
Test performance			
Overall ELA standard score	10.1	14.4	9.1
Overall Math standard score	14.3	20.8	12.9
Min sub-group ELA score	3.6	6.8	2.9
Min sub-group math score	8.1	13.8	6.9
Min group is close (w/in 5 points)	0.184	0.142	0.193
Fail: Min group more than 5 points away	0.391	0.277	0.416
Pass: Min group more than 5 points away	0.425	0.581	0.391