NATIONAL
CENTER *for* ANALYSIS *of* LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Teaching Assistants and Nonteaching Staff: Do They Improve Student Outcomes?

Steven W. Hemelt
Helen F. Ladd

WORKING PAPER 169 • August 2017

Teaching Assistants and Nonteaching Staff: Do They Improve Student Outcomes?

Steven W. Hemelt
University of North Carolina, Chapel Hill

Helen F. Ladd
Duke University

Contents

Acknowledgements

The authors gratefully acknowledge financial support from the National Center for Analysis of Longitudinal Data in Education Research (CALDER) funded through grant #R305C120008 to American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education. The views expressed are those of the authors and should not be attributed to the American Institutes for Research, its trustees, or any of the funders or supporting organizations mentioned herein. Any errors are attributable to the authors.

The authors want to thank research assistance from Adrienne Jones, Winnie Biwott, and Ying Shi who provided excellent research assistance for this project.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Teaching Assistants and Nonteaching Staff: Do They Improve Student Outcomes?
Steven W. Hemelt, Helen F. Ladd
CALDER Working Paper No. 169
October 2016

Abstract

This paper examines the influence of teacher assistants and other personnel on student outcomes in elementary schools during a period of recession-induced cutbacks in teachers and teacher assistants. Using panel data from North Carolina, we exploit the state's unique system of financing its local public schools to identify the causal effects of teacher assistants and other staff on student test scores in math and reading and other outcomes. We find strong and consistent evidence of positive contributions of teacher assistants, an understudied staffing category, with larger effects on outcomes for minority students than for white students.

Keywords: Teaching assistants, school finance, class size

JEL Categories: I2 (Education and Research Institutions); I 22 (Educational finance); J45 (Public sector labor markets)

I. Introduction

This paper takes a new look at a perennial question in education: Do resources matter and, if so, which ones? Our perspective differs from that of most other studies in that we focus directly on staffing levels rather than on funding levels or teacher quality. We do so in the context of elementary schools in North Carolina. The peculiarities of that state's approach to funding schools makes it possible for us to estimate plausibly causal estimates of the effects of teachers and teacher assistants and other staffing categories on student test scores and behaviors. The most compelling results emerge for the category of teacher assistants.

The study builds directly on two literatures related to school finance and policy. One is the extensive literature on whether money matters. In the many early studies summarized by Hanushek (1986, 1997), researchers explore whether spending on teachers in the form of higher salaries for years of experience, master's degrees or National Board Certification, or on larger numbers of teachers in the form of smaller class sizes pays off in the form of higher student outcomes, typically as measured by test scores. Based on Hanushek's conclusion of no clear and consistent effects of spending, the standard mantra for many policymakers has been that "money doesn't matter." That view has justified decisions either not to increase education spending or to implement accountability efforts designed to make schools use whatever resources they have, especially teachers, more productively. More recent work that makes use of funding changes induced by state school finance court cases challenges the Hanushek conclusion and indicates a more positive role for additional funding (Jackson et al., 2016; Hyman, 2017). Moreover, other recent studies confirm positive effects of school funding, especially in the lower grades and often for pupils in disadvantaged communities (e.g., Guryan 2001; Chadhary 2009; & Papke 2005).

A second smaller, but growing, body of literature focuses attention on expenditures for specific programs within schools, such as health centers or social workers, and often looks at a broader set of outcome measures, including graduation rates, teen pregnancy, and grade retention (Carrell & Carrell 2006; Reback 2010a; & Lovenheim et al., 2014).

Motivating the present study is the significant reduction in state-funded staffing levels for teachers and teacher assistants in North Carolina that took place during the Great Recession and that have continued to the present. The variation in staffing levels over the period 2001-2012 in those two categories is sufficiently large, especially for teacher assistants, for us to estimate causal effects on student outcomes at the elementary level. We conclude that cutbacks in teacher assistants have had clear adverse consequences for students.

The paper proceeds as follows. Section II describes the North Carolina funding context and Section III summarizes the relevant literature. We describe our data and methodology in Section IV, present our results in Section V, and report falsification and robustness checks in Section VI. The paper ends with a concluding discussion of our findings in Section VII.

II. North Carolina Context and School Funding

North Carolina has 9.8 million people distributed among several thriving metropolitan areas centered around cites such as Charlotte, Raleigh, Greensboro and Winston-Salem in the middle of the state, as well as many poorer rural areas to the east and to the mountainous west. The state is relatively poor – it ranks 38th among the states in terms of average annual income and 39th in terms of its overall poverty rate. Its per pupil education funding puts it at 45th in the nation and its average teacher salaries at 48th. Nonetheless, in 2015 its students scored at or slightly above the national average on the National Assessment of Educational Progress.[1]

Central to our research strategy is the state's unusual system of funding its schools. Starting in 1933, and reinforced by policy actions in 1975, the state government has had responsibility for funding a

[1] See results here: http://www.ncpublicschools.org/accountability/policies/naep/2

"sound basic education" for all students. The state currently meets its responsibility by distributing funds to the state's 115 school districts in three ways: position allotments (about 70 percent), dollar allotments (about 10 percent) and categorical programs (about 20 percent). With the exception of teacher assistants for which there are dollar allotments, support for staffing is provided primarily in the form of position allotments.

The position allotments work as follows. The state uses formulas to allocate positions (not dollars) to districts in the form of slots for specific categories of staff such as teachers, school leaders, and various types of support personnel including social workers and guidance counselors largely based on district-level student enrollments, as measured by average daily membership. The allocation formulas are detailed and clear. For example, for the year 2009-10, the state provided funds to cover 1 teacher per 18 students in grades K-3 and 1 teacher per 22 students in grades 4-6.[2] For each of the major staffing categories, districts have an incentive to use all the slots in the designated manner and to hire the most qualified people they can recruit, irrespective of salary level, because the state pays the staff according to a statewide salary schedule. Employing higher salaried teachers thus places no cost burden on the local school or district. For statewide budgetary purposes, the state predicts the costs of the slots based on the average statewide salaries per slot.

The one exception to this funding approach is for teacher assistants (TAs).[3] For them, the state provides a certain dollar amount of funding per student in the early grades to each district, thus limiting the number of full-time equivalent (FTE) TA positions a district can hire using state appropriations. The salary for teacher assistants is set at the rate for a specified state worker classification. The fungibility of

[2] The initial allocations are made based on the higher of actual average daily membership (ADM) for the prior year or the projected ADM for the current year, with adjustments made at the end of the year based on actual numbers. The allotments are not exactly the same as class sizes since they include teachers for supplemental subjects such as art and music. In recent years, the state has specified a maximum class size for each of the grade groupings. The grade groupings differ a bit from year to year. In some years, for example, kindergarten may be separated from grades 1-3 or grade 3 may be treated differently from grades K-2.

[3] Dollar allotments are also used for central office administration, textbooks, and classroom materials, supplies and equipment.

the TA dollars has historically been a bit unclear. In recent years, the legislature has clearly specified that the TA funding can be used only for teacher assistants.

The state does not adjust any of the staffing allocations or TA allotments to take account of the differing fiscal conditions of the local districts. Instead, the state provides some additional support through a variety of categorical programs. The largest categorical programs are for at-risk student services/alternative schools, children with disabilities, vocational education, and supplemental funding for low-wealth counties. Smaller programs include funding for limited English proficient students and for small counties. A 2008-09 analysis of the low-wealth funding program, which is distributed to 70 of the state's 100 countries, found that 8 percent of that funding was used for TAs and 37.3 percent for teachers and instructional support (Highlights of the North Carolina Public School Budget, 2010). Some of the other categorical programs may also be used for teachers or TAs.

The other two sources of funds are local and federal revenues. Although the state is by statute responsible for funding a sound basic education, the counties are allowed to supplement the state funds. Over time, the local share has grown and in 2012, the final year of our sample, it was 24 percent (North Carolina Public School Forum, 2012). In most cases, the counties are coterminous with the school districts but, largely as a historical anomaly, a few counties contain more than one district.[4] The locally elected school boards have to request funding from their county commissioners, because they, not the school boards, have the power to levy taxes. Districts use most of the local revenues provided for education for two main purposes – salary supplements to add to the amounts set by the state salary schedule for teachers, and funding for facilities. Some counties also use small amounts to hire additional staff and to pay for supplies. Counties differ in both their willingness and their ability to raise additional funds for education. In recent years, some districts have used leftover funds that the county

[4] There are 100 counties in North Carolina and 115 school districts.

accumulated from federal American Recovery and Reinvestment Act (ARRA) funds after the 2008 recession to offset cuts in positions funded by the state.

Federal funds are targeted toward specific populations such as low-income or handicapped children. The funds must be used for the purposes allotted and cannot be transferred. As of 2008-09, 6.5 percent of total personnel were paid from federal funds. That included about 6700 teachers and instructional support staff, 108 central office administrators, 4300 teacher assistants, and 1400 noncertified personnel. By 2011-12 that share had risen to 10.7 through a combination of reductions in state funding and increases in Title 1 funding under the NCLB act and in funding through IDEA for handicapped students.

One significant consequence of the slot-based system of funding by the state of North Carolina is that the districts are required to report staffing levels by category at the school level, not only for state-funded positions but also for positions funded by local countries or the federal government. As we elaborate further below, these detailed data serve as the basis for our analysis of the effects of staffing levels.

III. Relevant Literature on the Key Staffing Categories

We highlight here the literature related to our two main categories of primary interest: teachers and teacher assistants. In addition, we refer briefly to health and to other categories that we include in the various models. We start with teachers because it is the category for which there is the most prior interest, but our main interest is in the role of teacher assistants.

A. Teachers

Our measure of teachers is the number of regular teachers divided by the total number of students in each school. Although not specifically a measure of class size, because we do not link teachers and students to particular classes, changes in this measure within a school over time should be closely related to changes in class size.[5] Hence the relevant literature for predicting the effects of such a change is the extensive literature on the association between class size and educational outcomes.

At a theoretical level, class size often matters because of interactions among peers in the classroom. As spelled out in a well-known article by Edward Lazear (1999), instruction in the classroom is like a public good, but one in which negative externalities arise when the behavior of one or more students impede the learning of their classmates. This perspective predicts that lower class sizes may be most desirable in the lower grade levels and in schools serving disadvantaged children. The logic is that because younger children have shorter attention spans than older children the likelihood of disruption for any given class size is greater in the lower grades. An analogous argument applies for disadvantaged students who, because of their home situations, may come to class less ready to learn than their more advantaged counterparts.

The best-known and most credible evidence on class size comes from the Tennessee STAR class size reduction experiment in the 1980s. In this experiment, students in grades K-3 who were randomly assigned to small classes (13-17 students) were compared to those in regular classes (22-25 pupils) and also to those in regular classes with a teacher's aide. Krueger's 1999 analysis of the experimental data, in which he corrects for non-random attrition, finds significant and sizable test score gains in the first year that pupils are in smaller classes and smaller but still positive gains in subsequent years as they progress through school. Notably the effects are larger for more disadvantaged students. More recent studies that follow the treatment and control pupils into their 20s (Chetty et al., 2011; Dynarski et al., 2013) find that the pupils in the small classes were more likely to attend college and to have other positive outcomes.

[5] In fact, as we noted earlier in footnote 1, a per-student allocation of 18 regular teachers translates into a class size somewhat larger than 18 because the allotment includes special subjects such as art and music. The state intends an allotment of 18 students per teacher to translate into an average class size of about 21 students.

Other evidence from quasi-experimental studies that use a variety of strategies to isolate the effects of class size generate more mixed results. One strategy pioneered by Hoxby (2000) relies on variation in class size driven by demographic changes that lead to differences in the size of student cohorts. Applying this method to elementary school pupils in Connecticut, she finds no effects of class size on student achievement. Using data for Texas, Rivkin et al. (2005) find small, but inconsistent, class size effects and Cho et al. (2012) find quite small effects for students in grades 3 and 5 in Minnesota. In contrast, some of the non-U.S. studies find relatively large effects of class size reductions. A study using Israeli data by Angrist and Lavy (1999) makes use of rules relating to maximum class sizes and finds large effects of class size reductions on students in grades 4 and 5. Piketty (2004) and Bressoux et al. (2009) also find large class size effects in France, with larger effects for low-achieving students (see summary of evidence in Gibbons and McNally (2013)). Emerging from this discussion is the prediction that reductions in the number of teachers per student – which in turn imply larger class sizes – are likely to have negative effects on student outcomes at the elementary level.[6]

B. Teacher Assistants (TAs)

Nationally, teacher assistants account for close to 12 percent of the total elementary and secondary school labor force (Occupational Employment Statistics, 2014). Most districts in the nation historically required that TAs have at least a high school diploma, but the educational requirements to serve as a teaching assistant differ across states, and often, as is the case in North Carolina, also across districts and schools. The 2002 federal No Child Left Behind Act (NCLB) raised the standard for TAs

[6] We do not explore effects at the middle school level for two reasons. One is that the predicted effects are far less clear. On the one hand middle school students may have longer attention spans because they are older, and therefore be less likely to be disruptive, making larger class sizes more manageable. On the other hand, their vulnerability and hormonal changes may cause them to be more disruptive, which would call for smaller classes. A second, and more important, reason is that the link between the measure of the total number of regular teachers in a school per student, and class size is far weaker in middle schools. Changes in that measure at the middle school level need not lead to changes in the sizes of classes in math and English, which are the subjects in which students are tested. Instead they may show up in the form of changes in the number and variety of other course offerings in the school.

working in a Title I school (a school with more than 35 percent of its students qualifying for federal Title I funding), requiring them to have either two years of higher education or an associate's degree and to work with a "highly qualified teacher." About 70 percent of North Carolina's districts have adopted this higher standard for TAs, with the other 30 percent implementing it just for their Title 1 schools. North Carolina does not require TAs to have any form of state professional license.

TAs perform a variety of roles in classrooms. These include preparing classroom activities, working on instruction with individuals and small groups, performing clerical tasks, managing student behavior, and helping to evaluate student work (Kerry, 2005, NCATA). Using teacher assistants in any of these roles has the potential to free up time for teachers to focus on their main task of teaching and to make it easier to differentiate instruction within classrooms. Further, the presence of a TA in a classroom has the potential to mitigate some the negative externalities within a classroom highlighted by Lazear (1969). The precise mix of activities that teacher assistants engage in differs from classroom to classroom, of course, depending on differences in teachers' ability to make good use of their assistants and in the skills of the teacher assistants themselves. The test-based accountability provisions of NCLB appear to have placed more pressure on teacher assistants to participate directly in instruction, for which they may or may not be qualified.

The research on teacher assistants is far more limited than that on teachers. One 2001 study has garnered attention because it uses data from the highly touted Tennessee STAR experiment referred to above. This study (Gerber et al., 2001) addressed three questions: 1) Does the presence of a full-time teacher aide in the classroom in grades K through 3 advance students' academic achievement? 2) If so, does the effect depend on how long the student attends class with an aide? 3) Do some functions of aides have greater impacts on student achievement than others? The authors found that teacher aides had little, if any, positive effects on students' academic achievement, with the one exception being students who were in classrooms with an aide for 2-3 years. Nor did the authors find that any specific type of TA activity had an effect on student achievement.

A thorough review of the literature as of 2009 provides a more mixed and nuanced picture (Alborz et al., 2009a and 2009b). This review was based on 35 high-quality studies, mainly from the U.S., England and Wales, most of which examined teacher assistant support in primary schools. Of the eight studies that examined actual or perceived impacts of TAs on literacy and language, seven "suggested that trained and supported teacher assistants working on a one-to-one basis or in a small group can help primary aged children with literacy and language problems to make significant gains in learning" (Alborz et al., p. 1). Only two studies looked at numeracy, yielding mixed results. Nine of the studies looked at the impact on teaching practices, such as the approaches that teachers took to organizing the classroom and facilitating learning. The impacts varied greatly. In some cases the presence of aides did not have much effect on teaching practices, but in others it actively facilitated student learning. Also, there is some evidence that the presence of motivated support staff increased satisfaction, and reduced stress levels of teachers in mainstream classrooms (Alborz et al., technical report, p. 16). Overall, the authors conclude that teacher assistants can have a positive impact on pupil progress provided they are properly trained and supported.

None of this literature provides a particularly strong case for the efficacy of teacher assistants, although it suggests the effects are likely to be larger for disadvantaged students than for other students. The absence of recent studies and, in particular, studies that are grounded in the U.S. context of strong pressure on teachers to meet test-based accountability standards, represents a significant gap in knowledge. A primary aim of this paper is to start filling that gap by examining the relationship between exogenous changes in the number of teacher assistants at the school level and changes in student achievement.

C. Other staffing categories: Health care providers

Ideally we would have a clear and consistent measure of school-level health providers over time, but, unfortunately we do not. In practice, we define health providers as staff at the school level who are providing mental and allied health services, plus speech pathologists and audiologists. This group includes social workers (all of whom must have a degree in social work and be licensed by the NC Department of Public Instruction) but it does not include guidance counselors, a group that is deemed by the state to be providing school support services rather than health services. We do not include nurses because our data cover only about half of all school nurses in the state, those who are employed by local education agencies rather than other public agencies. Nor do we include psychologists, because the school-level information on psychologists is not consistent over time.

The state provides clear guidelines for what services are supposed to be provided by each health care category. The main task for all of them is to reduce the health-related barriers to learning that individual children bring to the classroom, by attending to their mental and physical health needs. In addition, these professionals are charged with addressing broader health issues such as the control of communicable disease, and the creation of positive classroom and school environments. Social workers, in particular, are expected to assess and evaluate students to inform the design of appropriate interventions; provide crisis intervention services related to family violence, substance abuse, and behavioral disorders; and engage in advocacy that seeks to ensure all students have equal access to education and services to enhance their local academic progress (See NC School Social Workers' Association).

Although the link between the health of children and student achievement is widely recognized, little research exists on how health professionals in schools might affect student achievement. A major challenge for researchers is the difficulty of ruling out the reverse causation that would arise, for example, if such personnel were assigned to schools judged to be most in need. Carrell and Carrell (2007) deal with this problem by using cross-semester variation in the counselor-student ratios in schools across a large Florida school district in which the exogenous variation arises because of the relatively random availability of graduate students from the University of Florida's counseling program for internships and practicums. Reback (2010a) deals with it by using Alabama's discrete cutoff for funding additional half-time appointments. Both studies find positive effects of school counselors on student behaviors such as reduced disciplinary or weapons-related incidents, but neither finds any effects on student achievement, perhaps because of the temporary nature of the changes.

In a far more ambitious study, Reback (2010b) uses national data from various state sources along with national survey data on students to explore the effects of school-site mental health services for children in third grade. His cross-state descriptive analysis provides evidence that students in states with more aggressive mental health counseling policies perform better both in terms of higher achievement and fewer behavioral problems than students in other states with weaker policies, even after controlling for many student background characteristics, including their test scores in kindergarten. In the second part of the study, which is more causally oriented, he uses a difference-in-differences approach based on the national Schools and Staffing Survey (the SASS) and finds that the availability of counseling services generates positive effects: it reduces the fraction of teachers who report that students are misbehaving or that their instruction suffers due to student misbehavior. This study has nothing to say directly about student achievement.

Other researchers have examined the extent to which school-based health centers affect a variety of student outcomes, but these studies typically focus on high schools, not the elementary level of interest in the current study. One recent working paper (Lovenheim et al., 2016), for example, uses national district-level data to analyze the effects of opening school health centers and finds reductions in teen births but no effects on graduation rates. Once again the existing evidence is limited, but it is consistent with the compelling logic that children who are experiencing mental health problems or who are dealing with stress that arises from parental job loss or their home environment will find it hard to learn.

D. Miscellaneous other staffing categories

School-level data are also available on other categories of staff that could potentially generate higher levels of achievement or improvements in other student outcomes. These categories include guidance counselors, principals and other school leaders, and noncertified academic staff.[7] In North Carolina, guidance counselors are trained to incorporate the state's Guidance Standard Course of Study through large and small group activities, focusing on students' growth and development.[8] Although the presence of guidance counselors in a school may well influence student achievement, the link is likely to be more tenuous than with the other three categories of professional staff.

Similar arguments apply to the category of school leaders, such as principals. Although adding leaders in a school may make students better behaved, the link with achievement is likely to be tenuous. As for non-certified academic support, which includes tutors, interpreters, therapists, and non-certified instructors, adding more staff could potentially have a more direct impact on student achievement, but we have chosen to treat that category the same way we treat guidance counselors and school leaders to retain the focus on teachers, teacher assistants, and health personnel. All of our achievement models control for an aggregate of these three categories, although we explore their separate effects in one of the models explaining behavioral outcomes.

[7] We excluded from all the analysis an additional category of non-regular teachers on the ground that they were working with special populations.

[8] With the rise in student assessment required by federal and state laws, many guidance counselors in North Carolina appeared to be spending significant amounts of time simply coordinating standardized testing within the school. But new guidelines from the state in 2014 decreed that they should be spending at least 80 percent of their time providing direct services to students, not counting any time they spend in test coordination.

IV. Data and Methods

We focus on three main categories of staffing positions: regular teachers, teacher assistants, and health staff, along with a miscellaneous category that we refer to as "other staff." For the reasons explained above, each of the three main types of staff has the potential to affect student outcomes. The outcome variables (measured and analyzed at the school level) include standardized math and reading scores for all students in grades 3-5 in elementary schools, percentages of students meeting state-defined proficiency levels in math and reading, and behavioral outcomes such as absence rates and suspensions.

We use school-level data from 2001 to 2012 for 1,094 public elementary schools with a total sample size of about 10,400 school-by-year observations for our test score analysis.[9] Lack of data limits our analysis of behavioral outcomes to the shorter 2006 to 2012 period. We exclude charter schools as well as schools with atypical grade configurations, such as grades 4-8 or grades K-8, which might require unusual staffing patterns. For each category of staff we have information on the number of (full time equivalent) slots broken down by whether the slots are paid for by state, local, or federal funds. Figures 1, 2, and 3 depict the number of positions in each of the three main categories per 100 students across our full analytic sample of elementary schools.

Without exception, the state government funds the largest share of the positions in each category. The changes over time in state-funded staffing levels in the categories of teachers and teacher assistants are central to our analytical strategy. The recent declines in those categories largely reflect a combination of recession-related pressures on the state budget and subsequent policy decisions by a Republican legislature after 2010 committed to reducing the size of government. In contrast, health care workers (as we measured them) have generally risen during the period perhaps partly as a response to the economic pressures on families related to the recession.

[9] Not all schools have 12 years of data because during the sample period new schools open and some close.

The state allocates slots (or in the case of teacher assistants a pot of money) to districts and not to schools. Districts may respond to changes in state-funded slots by changing levels of locally funded positions or by reallocating slots among schools.[10] Based on our discussions with district officials, it appears that districts use various district-specific formulas or policies for allocating positions across schools. For example, they might allocate slots on a straight per-pupil basis, or they might use various allocation schemes that target slots across schools to give disadvantaged schools more resources. We do not know much about the methods they use, but we do have evidence that at least in some cases their allocations have changed significantly over time. To illustrate, we show in Figure 4 the allocations across elementary schools for one district, Durham County, for the years 2004 and 2012. The graph shows that with the decline in teacher slots, the county flattened the distribution of teaching slots relative to the shares of disadvantaged students in each school (as measured by the percent eligible for free or reduced-price lunch on the horizontal axis). In addition, the district appeared to have also modified somewhat the allocation of teacher assistant positions. Within-district allocation decisions of this type pose a statistical challenge because they mean that changes in a school's slots over time may be related to student outcomes, which could lead to biased estimates of the effects of staffing on student outcomes. Hence, our statistical modeling strategy must address this and other related concerns.

A. Statistical Model and Identification Strategy

Our goal is to leverage changes over time in staff positions to determine how staffing patterns affect student outcomes at the school level. Letting *s* denote the school and *t* denote the year, we write the basic model as:

(1) $$Y_{st} = \alpha + \beta_s Staff_{st} + \beta_x X_{st} + \beta_z Z_{st} + \delta_s + \varphi_t + \varepsilon_{st}$$

[10] While districts are not allowed to transfer teaching slots to other categories, until recently they appear to have had some flexibility to shift funding allocated for teacher assistants to teachers.

where Y_{st} is a school-level outcome in year *t*. We analyze several types of outcome variables: average reading and math test scores, percentages of students proficient in math and reading, and a few behavioral outcomes. Of primary interest are the estimated coefficients, β_s, on each of four variables included in the $Staff_{st}$ vector: teachers, teacher assistants, health care workers, and other staff. Each staffing variable is the total full-time equivalent (FTE) staff of the specified type in school *s* in year *t* regardless of whether they are funded by the state, local, or federal governments.

As control variables, the model includes a vector (X_{st}) of time-varying student characteristics at the school level such as the share of students eligible for free or reduced-price lunch; a vector (Z_{st}) of teacher quality variables, such as the proportion of inexperienced teachers; school fixed effects (δ_s) and year fixed effects (φ_t). The inclusion of school fixed effects means we are looking at effects of within-school changes over time. The year fixed effects control for any statewide policy changes in the nature of the tests or basic statewide trends in the economy that might affect student outcomes. The model also includes a random error term (ε_{st}).

The statistical problem is that any one of the staffing variables could potentially be correlated with the error term. Even with all the control variables in the equation, it could be that a school with a high or low unexplained error term could have more or fewer staff in a particular category in a given year. That could occur for example if policymakers at the district level used student outcome variables in determining how to allocate staffing among schools. Alternatively, with cutbacks in total resources at the district level, policymakers may find it easier to absorb the cutback by reducing staffing in some schools than in others, perhaps because of differential rates of staff turnover that could be correlated with student outcomes.

To avoid the bias associated with such reactions, we estimate two-stage least squares (2SLS) models that leverage variation in state and federally funded staffing slots at the district level. This variation is outside the control of district policymakers and, importantly, is not determined by school-specific factors. By using district aggregates of state and federal staffing levels as instruments in the first stage of a 2SLS procedure, we are in effect restricting our analysis to changes in staffing at the school level that are predicted by factors external to individual schools. Thus, our instrumental variables, 2SLS, approach takes the following form:

$$(2) \qquad Staff_{st,p} = \alpha + \phi SA_{dt} + \lambda FA_{dt} + \kappa X_{st} + \gamma Z_{st} + \delta_s + \varphi_t + \upsilon_{st}$$

$$(3) \qquad Y_{st} = \alpha + \beta_s \widehat{Staff}_{st} + \beta_x X_{st} + \beta_z Z_{st} + \delta_s + \varphi_t + \varepsilon_{st}$$

Since there are four endogenous staffing variables, we estimate four first-stage models that look like equation 2 in structure. Here, in equation 2, $Staff_{st,p}$ represents the observed staffing level in school *s* and year *t* for one type (*p*) of staff (such as teacher assistants). The vectors SA_{dt} and FA_{dt} denote state and federal allocations of staffing positions for all four staffing categories to district *d* in year *t,* respectively. That is, for each of the four endogenous variables, we include in the first stage all instruments (i.e., all state and federal allocations for all the staffing types) along with our other controls. Those four first-stage models isolate plausibly exogenous variation in each of our four staffing variables. The second-stage model (equation 3), then uses this exogenous variation in staffing levels to estimate causal effects of staff changes at the school level on the various educational outcomes (Y_{st}).

This approach is based on two main assumptions. The first is that the exogenous, district-level aggregates are strong predictors of the school-level staffing variables. The first-stage results we report in

Appendix Table A1 corroborate that assumption.[11] The second is that the only way the exogenous district-level staffing variables affect the outcome variables is through their impact on the quantity of staffing at the school level.

One potential concern, for example, is that the estimates might be biased if some districts offset changes in state or federal funding with changes in local county-based funding. That is, what if they respond to reductions (increases) in state-funded positions by increasing (decreasing) locally funded positions? While that is clearly a possibility, it does not pose a problem because such behavior would affect the number of locally funded staffing positions, not the state or federally funded positions, which serve as the exogenous predictors in our 2SLS models. In contrast, that type of behavior could well bias the estimates in a model estimated by ordinary least squares.

Another possible concern is that changes in state or federally funded staffing may affect student outcomes at the school level through changes in the quality, in contrast to the quantity, of the staff. We are able to address this concern for teachers by including in the model three variables designed to control for their quality (Z_{st}). These time-varying, school-level variables include their average licensure test scores, the proportion of teachers with more than three years of experience, and the share who are National Board Certified. Research studies based on North Carolina data confirm that each of these measures of teacher quality are predictive of higher student test scores (Clotfelter et al., 2006, 2007).[12] The potential problem is that we have no similar measures for the quality of other school personnel,

[11] In addition, the results in Appendix Tale A1 illustrate that the subset of instruments with the tightest conceptual kinship with the particular endogenous staffing category of interest carry the bulk of the predictive power for that staffing category. For example, state and federal allocations of teacher slots predict school-level numbers of teachers much better than school-level numbers of health care staff, as they should.

[12] We rejected an alternative strategy of trying to measure teacher quality with salary data. Such salary differences would reflect both the mix of teachers within a school – with more experienced teachers paid more – and the willingness of the local county to supplement the pay of teachers over and above the statewide salary schedule. While, to some extent, higher teacher salaries may reflect quality differences in teachers, they may also reflect cost-of-living differences, or differences in the salaries needed to attract a given quality teacher to a particular part of the state.

including teacher assistants, and hence simply have to assume that the quality of teachers also serves as a proxy for the quality of other staff in the school. That assumption seems reasonable given that districts/schools are required to pay teacher assistants a state-specified salary. Only if the recession affected the supply of TAs willing to work at the given salary would the quality be affected – but why that would differ significantly for teachers and TAs is unclear. Moreover, the inclusion of year fixed effects in our models rules out the impact of statewide recession-related effects.

Finally, one might be concerned that changes in federal and state staffing levels might be correlated with changes in the characteristics of the students that in turn are likely to be correlated with student outcomes. We have accounted for that possibility by controlling for time-varying student characteristics at the school level (X_{st}). These variables include the racial mix of students, the proportions eligible for free and reduced-price lunch or who have special needs, and the percent female (see Table 1).

V. Findings

A. Descriptive Statistics and Basic Results

Table 1 provides descriptive statistics for all the variables in our test score analyses. The test scores of the students (standardized by grade and year) refer only to students in grades 3, 4, and 5 because those are the only students who are tested. All the staffing variables, however, apply to the whole school.[13]

Consider first the data for the full sample of elementary schools in Table 1. The mean teacher-to-student ratio is 0.052, that is, 5.2 teachers per 100 students. If each regular teacher had her own

[13] Average test scores for the full sample are not precisely zero in part because that sample excludes tested students in schools that offer non-standard sets of grades.

class, that would translate into an average class size of about 19 students, albeit most likely with smaller class sizes in the early grades and larger ones in the upper grades. The mean for teacher assistants is slightly more than half that, at 2.9 per 100 students. Health providers are much less numerous, with only about 0.2 per 100 students, or about 1 provider per 500 students. The "other" combined category of guidance counselors, school leaders, and non-certified academic staff is somewhat larger, with 0.7 per 100 students or about 1 per 140 students.

The student characteristics indicate that slightly less than half the students are eligible for free or reduced-price lunch (FRPL), 27 percent are black, 9 percent are Hispanic, and 10 percent have special needs. The teacher characteristics indicate that teachers have slightly above average licensure scores (i.e., Praxis scores), about 10 percent are national board (NBPTS) certified, and about 78 percent have more than three years of experience. The next two columns of Table 1 show comparable information for subsamples of students – minority students (defined as black, Hispanic, and students classified in other non-white categories) and white students. The staffing variables are almost identical across these two groups of students and are similar to those for the full sample.

The final two columns of Table 1 provide comparable information for schools with above average proportions of low-income students (indicated by high-FRPL) and those with below average proportions (indicated by low-FRPL).[14] The staffing variables are generally comparable to the full sample, although the high-FRPL schools have slightly higher ratios of teachers and teacher assistants, but lower quality teachers as measured by their lower average Praxis scores and lower percentages of Board Certified and experienced teachers.

Table 2 reports the full results for reading and math test scores, with our preferred 2SLS results for the full model in columns 1 and 3. For comparison purposes, the table also includes OLS equations

[14] We group schools into these categories based on the unweighted average proportion of FRPL students in a school over the time period of our panel. High-FRPL schools have more than 52 percent of their students eligible for FRPL and low-FRPL schools less than 52 percent.

for the full model (columns 2 and 4) as well as both 2SLS and OLS results for a modified model that excludes the health care staffing variable (columns 5-8). The first-stage estimates for the 2SLS full model are reported in Appendix Table A1. As shown in that Table, the exogenous variables (the 8 district aggregate staffing variables) enter the first stage equations with reasonable coefficients and high F-statistics that range from 23 for the “other staff” category to 343 for the teacher assistant category.

Focusing initially on the full model, we see that the estimated 2SLS coefficients for all of the staffing variables are positive, although several are not statistically significant. Before discussing and interpreting the magnitudes of those coefficients, we make three general observations. The first is that, not surprisingly, the 2SLS estimates for each of the staffing variables are larger than those from the OLS equation in both reading and math. This pattern is consistent with the hypothesis that more resources are directed to the schools with the lower performing students, which would bias the OLS coefficients down. Related is that as a result of the noise introduced by replacing actual staffing variables with predicted values, the 2SLS standard errors are substantially larger than those of the OLS estimates, which leads to less precise estimates.

The second observation is that the estimated coefficients of the health variables seem quite large (although for reasons we discuss below not quite as large relative to the effects of teachers and teacher assistants as they appear in this table). That surprising finding along with those from a falsification test we present later make us question the validity of the findings for the health variable. One possible approach is to exclude that variable from the model, as we have done in the final four columns of the table. A comparison between the full model and this other model indicates that excluding the health variable has small, but limited effects on the coefficients of the other staffing variables. Nonetheless, we prefer the full model because conceptually the health variable should be included in the model, and it could be correlated with some of the other staffing variables. Some evidence of the type of correlation about which we are concerned comes from a comparison of the magnitudes of the 2SLS coefficients of the teacher assistant variables in the full and the no health models. When the outcome is reading scores, excluding the health variable coincides with an increase in the coefficient on teacher assistants from 0.854 to 0.938, suggesting that the size of the health care staff is positively correlated with the number of teacher assistants. Similar logic justifies the inclusion in all models of the category of "other staff." By controlling for this other staffing category we avoid any bias that might arise from correlations between that category and the other staffing categories.[15]

The third observation is that all the time-varying student and teacher characteristics enter the equations as expected. Hence they appear to serve as reasonable statistical controls for any changing characteristics of the students or teachers that might be correlated with changes in staffing levels. For example, higher proportions of students eligible for free and reduced-price lunch or of minority students predict lower average test scores, and higher proportions of girls predict higher test scores. Among the teacher qualifications variables, teacher experience plays the largest and more consistent role: an increase in the proportion of teachers with more than 3 years of experience predicts higher test scores.

B. Magnitudes of the Staffing Effects

Table 3, which replicates part of Table 2, allows us to highlight and interpret the staffing effects based on the full 2SLS model. As expected, the regular teacher variables enter with positive coefficients, although we cannot rule out the hypothesis that more teachers per student (and hence smaller class sizes) have no impact on student reading scores. Even in math, the larger positive coefficient is only marginally significant. The larger magnitude of the estimated effect for math than for reading is consistent with findings from other studies that suggest that teachers exert larger impacts on test scores in math than in reading, presumably in part because of the larger out-of-school factors that affect

[15] We estimated some models in which we split the "other" category into its components of guidance counselors, school leaders, and non-certified academic support staff but find some implausibly large negative coefficients that we believe reflect some unusual correlations in the data. Later in the paper, we present some disaggregated results for the non-test score outcomes based on a shorter sample.

student reading (Clotfelter, Ladd and Vigdor,2006; 2007). The results for teacher assistants are much clearer. The models show that teacher assistants contribute in positive ways to student test scores in reading but not in math, with the reading coefficient about two-thirds that of the coefficient for regular teachers. The health workers enter with positive and statistically significant coefficients for both subjects, and the coefficients for the “other staff” variables are positive but not distinguishable from zero.

To determine the effect of one additional staff member in each category per 100 students we need to multiply the reported coefficients by 0.01. For teachers an increase of that magnitude would mean going from 5.2 teachers per 100 students to 6.2 teachers per 100 students, which would be equivalent to a reduction in average class size of about 3 students (bearing in mind, however, that some classrooms may initially be larger than others). Taking our estimates for regular teachers at face value, the estimates suggest that such a change would have an impact of about 0.01 standard deviations in reading and 0.023 standard deviations in math. For teacher assistants, the estimates indicate that one additional teacher assistant per 100 students would increase reading scores by about 0.009 standard deviations with essentially no impact on math. In the concluding section of this paper we compare these magnitudes to those reported elsewhere in the literature and to the relative costs of hiring teachers and teacher assistants.

Although the much larger coefficients for health providers (8.4 in reading and 25.4 in math may appear to suggest that the health personnel are the most productive, with one additional staff member per 100 students leading to a 0.08 to 0.25 standard deviation predicted change in student achievement, that conclusion would be misleading. The reason is that such a change would be far outside the range of our data, and would only make sense if the relationship remained linear throughout. Hence, it may be useful to consider the relative effects of a common percentage change, say 10 percent, in each of the categories, starting from the average. That would require us to multiply coefficients by 0.005, 0.003, and

0.0002 for teachers, teacher assistants and health personnel, respectively. The results of doing so are shown in square brackets under the standard errors for each estimate. For example, based on the preferred model and taking the coefficients at face value, the effect (in standard deviation units) of a 10 percent increase in teachers would be 0.006 in reading and 0.012 in math, of a 10 percent increase in teacher assistants 0.002 in reading and 0.000 in math, and for health personnel, about 0.002 in both subjects. We note that a 10 percent increase in teachers from the average would reduce the class size from about 19 to about 17 students.

C. Subgroup and Subsample Patterns

The literature we reviewed earlier suggested that smaller class sizes and more teacher assistants may be more effective for disadvantaged students than for other students. We look for differences by disadvantage first by comparing effects for minority students (i.e., black, Hispanic, and all other non-white students) to white students, based on the presumption that minority students are likely to be less advantaged than their white counterparts. The first four columns of Table 4 most clearly support the conclusion that teacher assistants are more effective for minority students than for white students in both reading and math. We note that the coefficients for minority students are large and statistically significant both for reading (coefficient of 2.13) and math (coefficient of 1.73) and that these effects are far larger than the estimated effects for white students in reading (coefficient of 0.99) and the insignificant effect in math. Tests indicate that these differences by student race are statistically significant for both subjects. Similar patterns emerge for regular teachers in that the estimated coefficients for minority students are larger (and more statistically significant) than the corresponding estimates for white students, implying that smaller class are more productive for minority than for white students. In this case, however, the differences across the subgroups are not statistically significant.

Ideally, we would have liked to repeat the analysis for subgroups divided by the free and reduced-price lunch status of the student, but we are missing the necessary data for many students.

Instead, we use school-level data on the mix of students to divide the sample of schools into those with above average or below average proportions based on their percentages of such students averaged over the full period. The final four columns report the results for the staffing variables. In this case, few differences emerge. Although there is some suggestive evidence that the health workers have larger effects in the more affluent schools (the ones with low shares of FRPL students) that difference is not statistically significant.

D. Academic Proficiency Results

We report one final set of test score results (see Table 5). In this case, the outcome is the proportion of tested students in each school who score at or above subject- and grade-specific proficiency levels. The averages for these composite proficiency rates are 74.2 percent in reading and 79.2 percent in math. In contrast to the other test score measures that incorporate all student test scores, this one focuses attention on students at the margin of success.

Throughout the period of this study, North Carolina schools, like schools in all other states, were subject to the federal accountability provisions of No Child Left Behind (NCLB). This federal law put pressure on individual schools to raise the proportions of their students who tested at proficient levels or above. The focus on proficiency rates encouraged schools to pay close attention to the students who were on the borderline of passing the tests, perhaps to the detriment of students who are likely to score way below or well above the proficiency cut point (Ladd & Lauen, 2010). Additional teacher assistants could potentially play a key role in helping students close to passing to perform well enough on the test to reach the proficiency level. The reason is that teacher assistants – assuming they have sufficient training – can work with small groups of students who need additional help to get over the hurdle as the teacher continues to work with the bulk of the class or they can free up time for the teacher to work more intensively with such students.

The positive and statistically significant estimated coefficients for the teacher assistants in both math and reading and in both forms of the model provide strong confirming evidence that teacher assistants play this role.

E. Impacts on Student Behavioral Outcomes

In addition to affecting test scores, staffing levels may affect other student outcomes of policy interest. In Table 6, we are able to examine effects on three behavioral outcomes – absences, tardies, and within-school suspensions – but only for the shorter period of 2006 to 2012.[16] Prior research has shown that such measures may serve as reasonable proxies for what some people refer to as socio-emotional or "non-cognitive" skills that contribute to student flourishing (e.g., Jackson, 2012). For example, a low absence or tardy rate may signify that a student is motivated to show up when expected. Moreover, attendance as early as grade six has been shown to predict the likelihood that a student will persevere and graduate from high school (Allensworth et al., 2007; Balfanz et al., 2007). Finally, student suspensions are often indicative of student misbehaviors that may interfere with the learning of not only the misbehaving student but also the learning of others through peer effects.

One potential limitation of the data available for all three outcomes is that it is reported by the schools. That raises the possibility that different schools may have somewhat different reporting policies which could, potentially, bias estimates of the effects of staffing levels. The inclusion of school fixed effects in our models helps somewhat. Nonetheless, it could still be that the more staff there are in a school over time, the more likely they are to report higher levels of one or more of these outcomes regardless of the true levels. Given our prediction that higher staffing would reduce the true incidence of all three of these behaviors, this reporting bias would work against detecting effects of these staffing categories on our set of behavioral outcomes.

---

[16] We also have data on out-of-school suspensions but find that they are generally not affected by the levels of staffing so we do not discuss them here.

We estimate models that include the same set of explanatory variables as the models we used to explain variation in test scores with one exception. For these models we have divided the "other staff category" into two parts, with one part including school leaders and the other the residual of guidance counselors and non-certified academic staff. We separate out the school leader category, which includes school principals and assistant principals, because a standard role for assistant principals is to address issues related to student behavior.

The absence rate is measured as the total number of absences divided by the total number of students in the school. The average rate across our sample is 4.75 days. Because of the non-normal distributions of the tardy rate and the in-school suspension rate, we treat the dependent variable in each case as a zero-one variable which takes on the value one if the school's rate is above the 75th percentile rate for the schools in the sample.[17] The sample for the tardy rate is far smaller than that for the other two variables because of incomplete data. Thus, we can show how changes in staffing levels of different types affect the probability of a high tardy or of a high suspension rate.

Relatively clear findings emerge for the teacher and teacher assistant staffing variables as well as for school leaders. A 10 percent increase in teachers (and hence smaller classes) reduces the average absentee rate by about 0.15 days per year (see estimate in square brackets), which represents about a 3 percent decline in the average absentee rate. Consistent with this finding, more teachers also lead to a statistically significant reduced probability of a high rate of in-school suspensions. We have no good explanation, however, for the positive sign for teachers in the tardy rate equation. Once again teacher assistants make a statistically significant contribution, although in this case the effects are relatively small, in the form of a lower absentee rate and a lower probability of a high tardy rate. In contrast to teachers, more teacher assistants have no apparent impact on the in-school suspension rate. Finally,

[17] The 75th percentile cutoffs at the elementary level are 2.7 times for tardies and 0.32 days for suspensions. We have experimented with different cut points but they do not affect the basic patterns shown in Table 5.

more school leaders – typically in the form of assistant principals, but measured in terms of full-time equivalents – also reduce the absentee rate and the tardy rate. The effect of a 10 percent increase in school leaders on the absentee rate is the same order of magnitude as a 10 percent increase in teachers.

F. Summary of Findings

In terms of filling gaps in the existing literature, the most important findings to emerge from this analysis relate to teacher assistants. We find clear evidence that teacher assistants have positive effects on student test scores in reading, with the positive effects being statistically significant not only for the full sample, but also for the racial subsamples and the subsamples of schools defined by poverty status. These positive effects in reading are consistent with some of the extant literature we cited earlier. The impacts on math test scores are much less clear, with positive effects appearing only for minority students. In addition, however, we find that teacher assistants help to boost school-level proficiency rates in both math and reading, and that teacher assistants reduce absentee rates and tardy rates. Thus, we have a strong and quite consistent story about the importance of teacher assistants.

More teachers (and hence smaller class sizes) also have a number of positive effects, although some of the effects are not statistically significant. The largest and most robust effects on test scores are for minority students. For these groups, smaller class sizes at the school level are associated with higher scores in both reading and math. More teachers also lead to lower student absentee rates and a lower probability of high rates of in-school suspension.

An increase in health staff also leads to higher test scores, although, somewhat unexpectedly, possibly more so for students in the more affluent (low FRPL) schools than in those serving more disadvantaged students. Evidence suggests that an increase in health care workers may also increase proficiency rates in math. Although these results are interesting and clearly worthy of further study, we do not want to oversell them because, as we show in the next section, the health care variable does not pass a falsification test, which casts doubt on the causality of the estimates. Finally, school leaders matter in that an increase in school leaders leads to lower absentee rates and lower probability of high tardy rates.

VI. Falsification and Robustness Checks

Our goal throughout has been to estimate causal relationships between school staffing and student outcomes in elementary schools. The combination of school fixed effects and time-varying demographic characteristics of schools, supplemented by a two-stage least squares (2SLS) strategy that relies on variation in state and federal staffing allocations at the district level, should rule out most, if not all, confounding effects at the school level. Nonetheless it is worthwhile to do some follow-up checks of our analysis.

The first check is a falsification test for our test score models in which we substitute for each of the staffing variables in the current year (*t*), the value of the variable in the following year (*t+1*). If our models are reasonable, we would not expect to find the same positive and statistically significant relationships between the future staffing variables and student test scores in the current year (*t*) that we find for the current staffing period variables. Table 7 reports the results of this test for both the preferred model and the model with no health worker variable. Both the teacher and the teacher assistant variables pass the test. None of the estimated coefficients of either variable is positive and statistically significant in either form of the model. The health worker variable is more problematic, however, in that it enters the preferred model with a large positive and statistically significant coefficient. Our interpretation of this finding is that the correlation over time in the health care staffing variable means that we should not interpret the estimates of health staffing that we reported in Tables 3 and 4 as causal effects. Fortunately, whether or not we control statistically for health care staffing has little effect on the coefficients of the teacher and teaching assistant variables of primary interest.

Second, we re-estimate the models with a smaller balanced sample of schools. The criterion for inclusion in this case is that a school must be operating and testing students in every year from 2001 to 2012. Given that North Carolina is a rapidly growing state with the need for many new schools, this sample adjustment reduces the set of schools from more than 1,090 in the main sample to about 690 or by about 37 percent. Moreover, the balanced sample is not representative of the full set of elementary schools in the state since it would most likely underrepresent schools in growing areas. The results for the balanced sample are reported in Table 8. Although the coefficients of the teacher variables are small and not significant, the pattern and magnitudes of the coefficients for the teacher assistant variables are similar to those for the full sample.

VII. Conclusions

It is difficult to determine how staffing levels within schools affect student outcomes such as test scores or rates of absenteeism given that school staffing levels are typically not exogenous to the school. The resulting challenge for empirical research is to rule out the possibility of reverse causation, wherein student outcomes affect staffing levels, which in turn would lead to biased estimates of staffing effects. The estimates would be upward biased if staffing levels were typically higher in schools serving more advantaged and higher-performing students and downward biased if staffing levels were higher in schools serving more disadvantaged and lower-performing students.

One of the contributions of this study is our use of a new strategy for addressing such bias-inducing reverse causation. In particular, the specific manner in which the state of North Carolina funds its schools allows us to instrument for changes in staffing at the school level with state and federally determined changes in slot allocations to districts. This two-stage approach combined with school fixed effects and time-varying demographic and teacher quality variables at the school level provides plausibly causal estimates of the effects of staffing levels, such as teachers and teacher assistants, on student outcomes.

In terms of substantive findings, the most important finding is the strong and consistent evidence on the role of teacher assistants, a staffing category that has been growing over time, but that has been woefully understudied. Positive effects of teacher assistants emerge for most of the outcome measures and across most of the specifications that we present. Moreover, the evidence is consistent with the hypothesis that teacher assistants are more productive in terms of academic achievement for minority students than for white students.

Further, findings regarding our teacher variable provide new support for the conclusion that class size matters in elementary school, at least over the implied range of class sizes observed, but more so for math scores than for reading, and more so for minority students than for white students. Due to data limitations, our positive findings that relate health staff to measures of academic achievement and behavior should at best be interpreted as a first step and one that highlights the need for additional research on the effects of a broader range of health care workers, including nurses, in schools.

We can put our estimated effects for the teacher variables into perspective by comparing the implied class size effects to other estimates of class size in the literature (see Gibbons and McNally (2013)). As we noted earlier, the most well-known study of class size is the Tennessee class size experiment, which reduced class sizes from about 22-25 students down to 13-17 students in grades K-3. A 2006 review by Diane Schanzenbach of the many studies of that experiment concluded that the test score impact was about 0.15 standard deviations. Given the average reduction was 8 students (from about 23 students per teacher down to about 15 students per teacher), the teacher-to-student ratio would have risen from 0.043 to about 0.066, or about 0.023 units in our metric. In Table 9, we apply a change of this magnitude to the coefficients for math and reading from our preferred model in Table 3 and to the analogues coefficients for minority students in Table 4.

Although some of the implied effects may appear quite small, a few points are worth noting. First, in comparing our estimates to the effects of the STAR class reductions, one should bear in mind that those reductions applied only to the early grades for which one might expect larger effects. Second, other studies using different methods such as Hoxby (2000) found no effects at all and another study using the same method she used (Cho et al., 2012) from Minnesota found effects that are comparable in size to our estimates (i.e., effects of 0.04 and 0.05 standard deviations for class size reductions of 10 students). Using the California class size reduction, Jepson and Rivkin (2009) find effects of 0.06 in reading and 0.10 in math which are comparable to our estimates for minority students. In light of this literature, we conclude that our estimates appear to be reasonable, albeit substantially smaller than effects that emerged from the Tennessee STAR context.

We have no studies with which to compare the magnitudes of our estimated achievement effects of teacher assistants. The best we can do is to compare the productivity of teacher assistants to that of teachers. If we compare our preferred estimate of 0.854 from Table 3 for teacher assistants in reading to the (statistically insignificant) estimate of the teacher effect for reading of 1.24, we would conclude that teacher assistants are about 70 percent as effective as regular teachers in reading. With reference to the coefficients in Table 4 for minority students, we would conclude that teacher assistants are about 75 percent as effective as regular teachers in reading and 41 percent as effective in math with minority students. That would imply that teachers and teacher assistants would be equally good investments if the salary of a typical TA were somewhere between 40 and 71 percent of the salary of a regular teacher, provided that the only purpose for both is to raise student achievement.

We use these comparisons primarily to indicate that our preferred estimates of the effects of staffing categories are reasonable and to emphasize our conclusion that staffing levels matter. We remind the reader, however, that student achievement as measured by test scores in math and reading is not the only goal of schooling, and that teachers, teacher assistants, and health providers may also contribute to student outcomes in ways that we have not measured here. We have provided some hints of these other outcomes in our analysis of behavioral outcomes, but more research is needed to present a full picture of how school staff contribute to valued outcomes for students.

REFERENCES

Alborz, A., Pearson, D. Farrell, P. and Howes, A. (2009a). "The Impact of Adults Support Staff on Pupils and Mainstream Schools, A Systematic Review of the Evidence" A research brief from the EPPI-Center, Social Science Research Unit, Institute of Education. University of London. DCSF-EPPI—09-01.

Alborz, A., Pearson, D. Farrell, P. and Howes, A. (2009b). *The Impact of Adults Support Staff on Pupils and Mainstream Schools. Technical Report.* Research Evidence in Education Library. London: EPPI-Center, Social Science Research Unite, Institute of Education, University of London. Report no. 1702T. http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=w423SGddbGM%3d&tabid=2438&mid=4540

Allensworth, E.M. and. J.Q. Easton. (2007). *What matters for staying on tract and graduating in Chicago Public Schools*. Chicago, IL: Consortium on Chicago School Research at the University of Chicago.

Angrist, J., & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics, 114(2),* 533-75.

Balfanz, Robert, Lisa Herzog, and Douglas J. Mac Iver. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist, 42 (4): 223-235.*

Bressoux, P., F. Kramarz and C. Prost. (2009). Teachers' Training, Class Size and Students' Outcomes: Learning from Administrative Forecasting Mistakes. *Economic Journal*, 119: 540-61.

Carrell, S. & Carrell, S. (2006). Do lower student to counselor ratios reduce school disciplinary problems? *Contributions to Economic Analysis & Policy, 5(1),* Article 11.

Chaudhary, L. (2009). Education Inputs, Student Performance and School Finance Reform in Michigan. *Economics of Education Review, 28,* 90-98.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics, 126(4),* 1593-660.

Cho, H., P. Glewwe, and M. Whitler. (2012). Do Reductions in Class Size Raise Student's Test Scores? Evidence from Population Variation in Minnesota's Elementary Schools. *Economics of Education Review, 31*, 77-95.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26(6),* 673-682.

Clotfelter, C. T., Ladd, H.F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, 41(4),* 778-820.

Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management, 32(4),* 392-717.

Gerber, S. B., Finn, J. D., Achilles, C. M., & Boyd-Zaharias, J. (2001). Teacher aides and students' academic achievement, *Educational Evaluation and Policy Analysis, 23(2),* 123-143.

Gibbons, S. G., & McNally, S. (2013). The effects of resources across school phases: a summary of recent evidence. Center for Economic Performance. DEP Discussion Paper No. 1226.

Guryan, J. (2001). Does Money Matter? Regression Discontinuity Estimates from Education Finance Reform in Massachusetts, NBER Working Paper 8269, May.

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature, 24(3),* 1141-77.

Hoxby, C. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics, 115(4),* 1239-85.

Hyman, J. (2017). Does money matter in the long run? Effect of school spending on educational attainment. *American Economic Journal: Economic Policy,* forthcoming.

Jackson, C. K., R.C. Johnson, and C. Persico (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *Quarterly Journal of Economics,* 131(1), 157-218.

Jackson, C. K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina. NBER Working Paper No. 18624. Access: http://www.nber.org/papers/w18624

Jepsen, C. and S. Rivkin (2009). Class Size Reductions and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size, *Journal of Human Resources,* 44(1), 223-50.

Krueger, A. (1999), Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics, 114(2),* 497-532

Kerry, T. (2005). Towards a typology for conceptualizing the roles of teaching. *Educational Review, 57(3), 20-30.*

Krueger, A., (1999) ''Experimental Estimates of Education Production Functions.' *Quarterly Journal of Economics*, vol. 114, pp. 497–532.
Ladd, H.F. and D.L. Lauen (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management, Vol 29(3),* pp. 426-450.

Lazear, E.P. (2001). Educational Production. *Quarterly Journal of Economics,* 116 (3): 777-803.

Papke, L. E. (2005), The Effects of Spending on Test Pass Rates: Evidence from Michigan, *Journal of Public Economics,* 8: 821-39.

Lovenheim, M. F., Reback, R. & Wedenoja, L. (2016). "How does access to Health Care Affect Teen Fertility and High School Dropout Rates? Evidence from school-based health center openings?" National Bureau of Economic Research: working paper 22030.

North Carolina School Social Workers' Association (no date). (www.ncsswa.org)

Piketty, T. and M. Valdenaire. (2006). L'impact de la taille des classes sur la reussite scolaire dans les ecoles, collµeges et lycees francais: Estimations µa partir du panel primaire 1997 et du panel secondaire 1995, Ministµere de l'Education Nationale, Les Dossiers n173.

Public School Forum of North Carolina, 2012. Local School Finance Study.

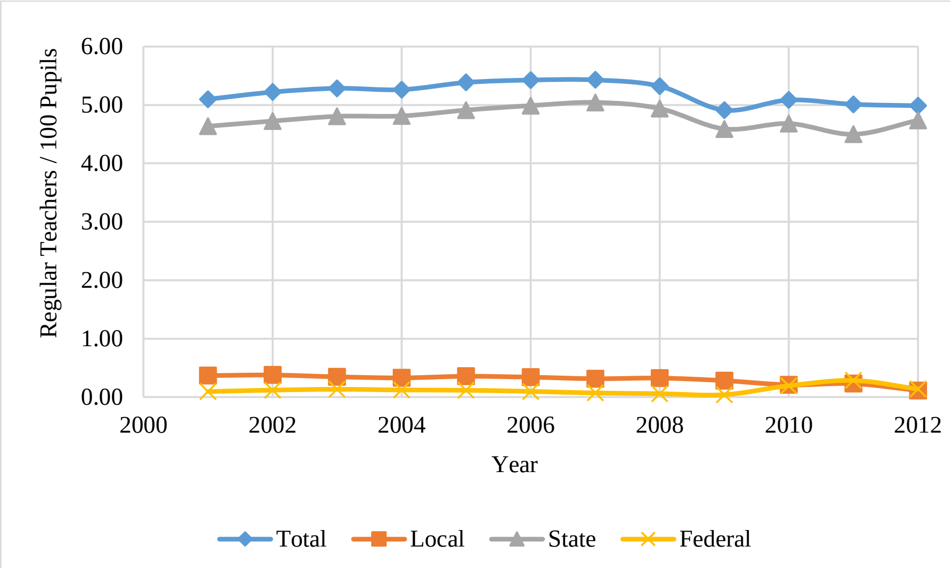
Reback, R. (2010a). Noninstructional spending improves noncognitive outcomes: Discontinuity evidence from a unique school counselor financing system. *Education Finance and Policy* 5(2), 105-137.

Reback. R. (2010b). Schools' Mental Health Services and Young Children's Emotions, Behaviors and Learning. *Journal of Policy Analysis and Management*, 29(4), 698-725.

Rivkin, S., E. Hanushek and J. Kain (2005), Teachers, Schools, and Academic Achievement, *Econometrica, 73(2),* 417–58.

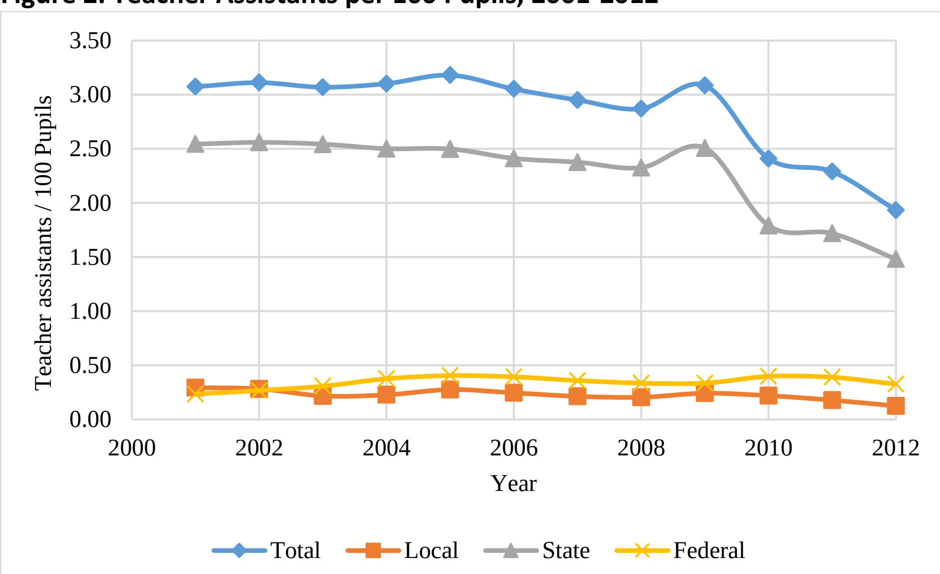
Schanzenbach, D. (2006) What have Researchers Learned from Project STAR? Bookings Paper on Education Policy: 205-28.

Figure 1. Regular Teachers per 100 Pupils, 2001-2012



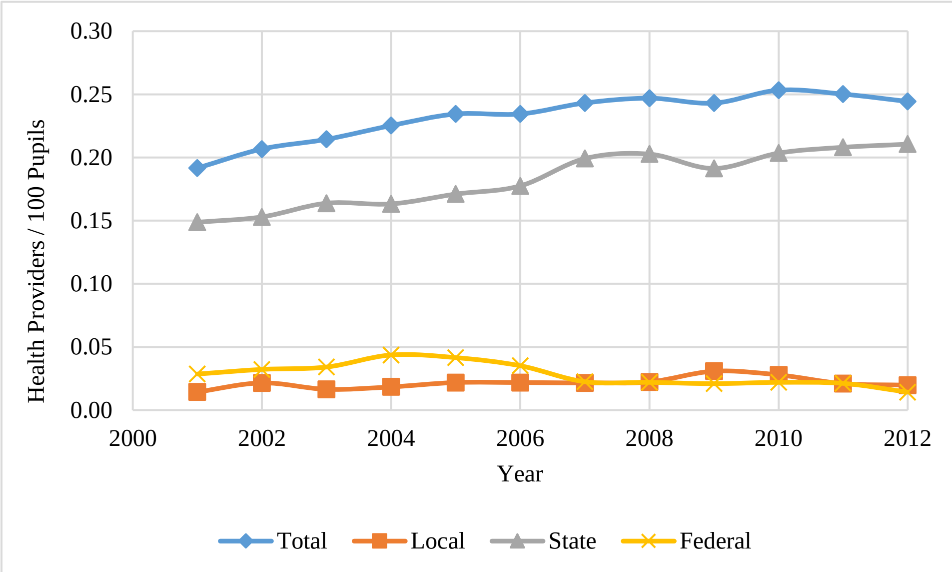
Source: Authors' calculations for all elementary schools in sample.

Figure 2. Teacher Assistants per 100 Pupils, 2001-2012



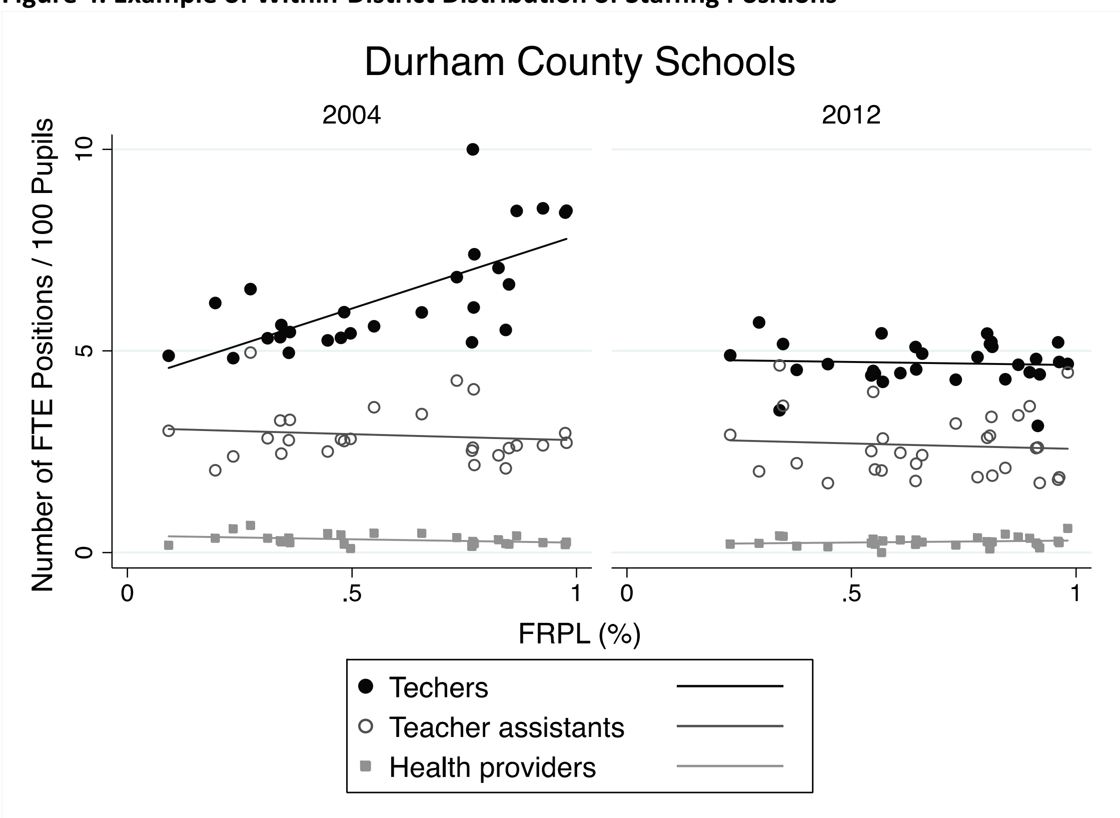
Source: Authors' calculations for all elementary schools in sample.

Figure 3. Health Providers per 100 Pupils, 2001-2012



Source: Authors' calculations for all elementary schools in sample.

Figure 4. Example of Within-District Distribution of Staffing Positions



Notes: Sample includes traditional public elementary schools in Durham County. Solid lines represent linear fits of the underlying data; FRPL = Free or reduced-price lunch.