This article 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 idiosyncrasies of that state’s approach to funding schools make it possible for us to produce plausibly causal estimates of the effects of teacher assistants (TAs) on student outcomes.

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. Hanushek’s work is the source of the common mantra that “money doesn’t matter.” But more recent work that makes use of funding changes induced by state school finance court cases challenges that conclusion and indicates a more positive role for additional funding (e.g., Hyman, 2017; Jackson et al., 2016). Moreover, other recent studies confirm positive effects of school funding, especially in the lower grades and often for students in high-poverty schools than for White students and students in more affluent schools. We conclude that teacher assistants are a cost-effective means of raising student achievement, especially in reading.

Keywords: achievement, economics of education, educational policy, elementary schools, econometric analysis, policy analysis, regression analyses, quasi-experimental analysis
disadvantaged communities (e.g., Chaudhary, 2009; Guryan, 2001; Papke, 2005).

A second smaller, but growing, body of literature focuses attention on expenditures for specific programs or nonteacher staffing 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; Lovenheim et al., 2016; Reback, 2010a).

Motivating the present study is the significant reduction in state-funded staffing levels for TAs in North Carolina that took place during the Great Recession and that have continued to the end of our study period. We exploit variation in staffing levels over the period 2001–2012 to estimate causal effects on student outcomes at the elementary level. Our main finding is that TAs enhance student performance in reading and math, with the clearest and most widespread effects on reading. We also find that TAs boost achievement outcomes more for traditionally underserved groups of students than for other students.

The article proceeds as follows. Section “North Carolina Context and School Funding” describes the North Carolina funding context and section “Key Staffing Categories” summarizes the relevant literature. We describe our data and methodology in section “Data and Method,” and present our results in section “Findings.” The article ends with a concluding discussion of the main findings in section “Summary and Discussion.”

North Carolina Context and School Funding

North Carolina has more than 10 million people distributed among several thriving metropolitan areas centered around cities 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—at the end of our sample period, in 2012, it ranked 39th among the states in terms of average annual income and 12th in terms of its overall poverty rate. Its instructional expenditures per pupil puts it at 39th 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. 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 “sound basic education” for all students. The state currently meets its responsibility by distributing funds to the state’s 115 school districts by allotments for positions or programs provided in three forms. These include position allotments for specific categories of personnel including teachers, dollar allotments for specific purposes including TAs, and categorical programs. Importantly, the personnel allotments to districts are generally used for the designated category of staffing although in various years, districts have been given some flexibility to move funds from one category to another, such as from TAs to teachers in Grades K–3.

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 school 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 2009–2010, the state provided funds to cover one teacher per 18 students in Grades K–3 and one teacher per 22 students in Grades 4–6. 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 primary funding for TAs differs in that it is based on dollar allotments to each district based on the number of pupils in the early grades (i.e., K–3). The districts use this funding to hire TAs with salaries within a state-specified range. Recent budget documents make clear that districts cannot move dollars out of the TA allotment (Department of Public Instruction, North Carolina, 2018; Nordstrom, 2017); however, in some earlier years, districts were allowed to
transfer a portion of TA funding to teacher slots if they chose to do so—but were not allowed to transfer funds into most other allotments, including those for technology, behavioral support, special education, or driver training. Districts can fund additional TAs out of other state grants, federal grants, or local funds.

The state does not adjust any of the base staffing allocations or TA allotments to take account of the differing conditions of the local districts. Instead, the state provides additional support to districts through a variety of categorical programs. The largest categorical programs are for “at-risk student services” and alternative schools, children with disabilities, vocational education, and supplemental funding for low-wealth districts. Smaller programs include funding for limited English proficient students and for small counties. A 2008–2009 analysis of the low-wealth funding program, which is distributed to 70 of the state’s 100 counties, found that 8% of that funding was used for TAs and 37.3% for teachers and instructional support (Highlights of the North Carolina Public School Budget, North Carolina Department of Public Instruction, 2010). Some of the other categorical programs, most notably the allotment for children with disabilities, were used for TAs.

School districts also receive funding from local counties and the federal government. Although the state is by statute responsible for funding a sound basic education, counties are allowed to supplement state funds. Over time, the local share has grown and in 2012, the final year of our sample, it was 24% (Public School Forum of North Carolina, 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. 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 primary purposes—salary supplements to add to the amounts set by the state salary schedule for teachers, and funding for facilities. Federal funds are targeted toward specific populations such as low-income or disabled students. The funds must be used for the purposes allotted and cannot be transferred. As of 2008–2009, 6.5% of total K–12 personnel in North Carolina were paid out of federal funds.

One significant consequence of the allotment-based system of funding by the state of North Carolina is that 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 counties 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.

Key Staffing Categories

In our empirical models, we include four categories of staffing positions: regular teachers, TAs, health care providers, and a miscellaneous category that we refer to as “other staff.” We highlight in this section the literature on those four categories of staffing, with most attention to the literature on TAs. In our empirical models, we specify each category at the district level, aggregated across the set of elementary schools in our sample within a district, as the full-time equivalent (FTE) staff by type per student.

Teachers

Our measure of teachers is the number of positions identified as regular teachers divided by the total number of students in each district (in a given year). Although not specifically a measure of class size, because we do not link teachers and students to particular classrooms, changes in this measure within a district over time should closely relate to changes in class size. 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 Lazear (2001), instruction in the classroom is like a public good, but one in which negative externalities arise when the behavior of one or more students impedes the learning of their classmates.

The best-known and most credible evidence on how changes in class size affect student achievement 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 with those in regular classes (22–25 pupils) and also to those in regular classes with a teacher’s aide. Krueger’s (1999) reanalysis of the experimental data, in which he corrects for nonrandom 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.

Other evidence from quasi-experimental studies that use a variety of strategies to isolate the causal effects of class size generate more mixed results. Hoxby (2000) finds no class size effects on student achievement in Connecticut; Angrist et al. (2019) find no effects in Israel; Rivkin et al. (2005) find small, but inconsistent, class size effects in Texas; and Cho et al. (2012) find small effects for students in Grades 3 and 5 in Minnesota. Based on this literature, we would predict that more teachers per student (as a proxy for smaller class sizes) would improve the outcomes of disadvantaged students, with average effects on all students less clear. Unfortunately, the limited amount of within-district variation over time in the context of our preferred models constrains our ability to detect effects of teachers or of class size. Nonetheless, we include teachers in all our models to avoid the possibility that our estimates of the effects of TAs are confounded by class size effects.

**TAs**

Nationally, TAs account for close to 12% of the total elementary and secondary school labor force (U.S. Department of Labor, 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 TA 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 working in a Title I school (a school with at least 40% of its students qualifying for federal Title I funding), requiring them to have either 2 years of higher education or an associate’s degree and to work with a “highly qualified teacher.” About 70% of North Carolina’s districts have adopted this higher standard for TAs, with the other 30% implementing it just for their Title I 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). Using TAs 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 of the negative externalities within a classroom highlighted by Lazear (2001). The precise mix of activities that TAs engage in differs from classroom to classroom depending on differences in teachers’ ability to make good use of their assistants and in the skills of the TAs themselves. The test-based accountability provisions of NCLB may have placed more pressure on TAs to participate directly in instruction, for which they may or may not be qualified.

The research on TAs is far more limited than that on teachers. Krueger (1999) concluded that the provision of a full-time teacher aide had modest and inconsistent effects on student achievement. In a reanalysis of the Tennessee STAR data, Gerber et al. (2001) 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 to 3 years. Nor did the authors find that any specific type of TA activity had an effect on student achievement. A thorough 2009 review of the literature on TAs mainly from the United States, England, and Wales suggests a more mixed and nuanced picture (Alborz et al., 2009a, 2009b). In particular, of the eight studies that examined the effects of TAs in primary schools on measures of literacy, seven “suggested that trained and supported TAs 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., 2009a, p. 2). Effects on math achievement were more mixed. Thus, one
reasonable hypothesis is that TAs may be particularly helpful in attending to differentiated needs of students related to reading.

Recent and compelling evidence of the potential for positive effects of TAs emerges from a randomized trial of two types of teacher aides—those with and without teaching degrees—in sixth-grade classrooms in Denmark (Andersen et al., 2018). The authors find positive and significant intent-to-treat effects on test scores in reading, but not in math, with effect sizes larger for students from disadvantaged backgrounds. Although the effects were similar in magnitude across the two types of teacher aides, the authors speculate that the mechanisms differed. The aides with degrees functioned more like teachers and served, according to the authors, as a flexible form of class size reduction. The aides without degrees appeared to work more as parts of professional teams focused on reducing classroom behavioral problems and complementing pedagogical efforts by the lead teacher.

Overall, this literature suggests that TAs may have positive effects on student outcomes, with the test score effects likely to be larger for reading than for math, and larger for disadvantaged students than for other students. The absence of recent 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 article is to fill that gap by examining the relationship between arguably exogenous changes in the number of TAs and changes in student achievement in North Carolina.

Health Care Providers

Our measure of school-level health care providers is not as comprehensive as we would have liked. 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 excludes 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 school psychologists because school-level information on psychologists is not consistent over time.

According to state guidelines, the main task for all the health care workers is to reduce 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.

Although the link between the general health of children and student achievement is relatively clear (Currie, 2009; Ding et al., 2009), little research exists on how, if at all, health professionals in schools might affect student achievement and other student outcomes. 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. Recent studies of the introduction of school-based health centers indicate that they generate mixed effects on academic outcomes. One national study (Lovenheim et al., 2016) finds that their introduction substantially reduced teen birth rates but only slightly reduced high school dropout rates.

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 school counselors, principals and other school leaders, and noncertified academic staff. In North Carolina, school 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. Two high-quality causal studies—one by Carrell and Carrell (2006) and the other by Reback (2010b)—both 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 used to identify effects.

Similar arguments apply to the category of school leaders, such as principals. Principals influence staffing, curriculum, and climate, all of
which may shape student achievement (Coelli & Green, 2012; Dhuey & Smith, 2014). However, within-district variation over time in leadership position allotments is, unsurprisingly, very limited over our period and thus not well-suited to function as a stand-alone staffing category in our analytic approach. As for noncertified academic support, which includes tutors, interpreters, therapists, and noncertified 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 school counselors and school leaders to retain the focus on teachers, TAs, and health personnel. All of our achievement models control for an aggregate of these three categories.

**Data and Method**

Our goal is to determine how each of our four staffing categories affects student achievement outcomes in math and reading in elementary school, with a primary focus on TAs. We begin with student- and school-level data that we aggregate to the district level from 2001 to 2012. This procedure generates 1,149 district-by-year observations. In each year, we include all public elementary schools in North Carolina with traditional grade configurations, defined as pre-K or K to Grade 5 or to Grade 6.\footnote{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.} As state-based allocations of staff are made to districts, the identifying variation we exploit operates at the district level, and thus we anchor our analysis at that level.

As we have explained, the state allocates slots (or dollars in the case of TAs) to districts and not to schools. Thus, districts in North Carolina, and by extension the schools under their authority, have staff that are funded from one of three sources: (a) the district itself through local revenues (e.g., property taxes), (b) the state government, and (c) the federal government. Districts are free to determine how many positions they choose to fund out of local revenue. In addition, they may also influence the number of federally funded positions through their decision of whether to spend federal funds on staffing versus other inputs such as curricula, as well as decisions about whether to apply for certain kinds of federal funding. Hence, neither the locally nor the federally funded positions are exogenous to the district. In contrast, local districts have little or no control over the basic state-funded positions or funding allotments which are determined by statewide policy decisions. We note, however, that many counties also receive additional state funding in the form of low-wealth and small county supplements that they may use to pay for positions at the school level, which are typically coded as state-funded positions. We control directly for these sources of aid in our modeling approach, below.

For each category of staff, we have school-level information on the number of (FTE) positions broken down by whether they are paid for by state, local, or federal funds. Figures 1 to 3 depict the number of positions in each of the three main categories per 100 students in elementary schools across our full analytic sample of districts. The figures illustrate that the state government funds the largest share of the staffing positions in each category. The recent decline in TAs largely reflects a combination of recession-related pressures on the state budget and subsequent policy decisions by the state legislature after 2010 aimed at 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.

Districts vary in how they allocate state-funded positions among schools. Based on discussions with district officials, we found that some allocate slots on a strict per-pupil basis, whereas others use various allocation schemes that target slots across schools to give disadvantaged schools more resources. As our analysis is anchored at the district level, we view the within-district disbursement of staff across schools as part of the treatment effect of the plausibly exogenous shift in staffing levels associated with changes in state policy.

**Empirical Approach**

Our goal is to leverage changes over time in staff positions in each of the four categories to determine how staffing patterns affect student
outcomes at the district level. Letting $d$ denote the district and $t$ denote the year, we write the basic model as:

$$Y_{dt} = \alpha + \beta_{\text{Staff}_{dt}} + \beta_{\text{Z}_{dt}} + \beta_{\text{Ed}_{dt}} + \delta_d + \phi_t + \epsilon_{dt},$$

(1)

where $Y_{dt}$ is a district-level outcome in year $t$. We analyze several types of outcome variables: average reading and math test scores and percentages of students proficient in math and reading. Of primary interest are the estimated coefficients, $\beta_s$, on each of four variables included in the $\text{Staff}_{dt}$ vector: teachers, TAs, health care workers, and other staff. Each staffing variable is the total FTE staff of the specified type in district $d$ in year $t$ regardless of whether they are funded by the state, local, or federal government.

As control variables, the model includes a vector ($X_{dt}$) of time-varying student characteristics at the district level such as the share of economically disadvantaged students; a vector of teacher quality variables ($Z_{dt}$), such as the proportion of experienced teachers; and a set of time-varying measures of local economic or other relevant conditions at the district level ($E_{dt}$), such as the local unemployment rate, the presence of charter schools, and low-wealth and small county aid per pupil. All models also include district fixed effects ($\delta_d$) and year fixed effects ($\phi_t$). The inclusion of district fixed effects means that we are using within-district changes in staffing over time to estimate effects on our outcomes of interest. 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 ($\epsilon_{dt}$). The inclusion of an extensive set of control variables, along with the district and year effects, should go a long way toward generating causal effects of the staffing variables.

Nonetheless, concerns about bias may arise because any one of the staffing variables could potentially be correlated with the error term. That problem could occur, for example, if policymakers at any of the three levels of government determined staffing levels based on the performance of the district’s students. Although much of the state government’s basic staffing allotments are

FIGURE 1. *Teacher assistants per 100 students, 2001–2012.*

Note. Authors’ calculations based on students in public schools serving elementary grades in North Carolina. Consult text for additional details on sample construction.

FIGURE 2. *Regular teachers per 100 students, 2001–2012.*

Note. Authors’ calculations based on students in public schools serving elementary grades in North Carolina. Consult text for additional details on sample construction.

FIGURE 3. *Health providers per 100 students, 2001–2012.*

Note. Authors’ calculations based on students in public schools serving elementary grades in North Carolina. Consult text for additional details on sample construction and types of staff counted as health providers.
solely a function of district enrollment levels, that may not be true for staffing paid for by either the federal or local government.

Hence, we supplement the basic ordinary least squares (OLS) model of Equation 1 with a reduced-form (RF) version in which each of the four staffing variables is replaced by its state-funded component. This component is outside the control of district policymakers and, importantly, is not determined by district-specific factors that may correlate with student achievement. By purging the staffing variables of endogenous components that may be associated with local and federal decisions, this RF specification estimates “intent-to-treat” effects of state funding on student outcomes for each category of staffing.

In addition, we use a two-stage least squares (2SLS) approach to estimate the causal effect of additional staff (of a given type) on student achievement outcomes. By using district aggregates of staffing positions funded by the state as instruments in the first stage of the 2SLS setup, we once again purge the influence of most of the hard-to-measure factors that might shape both staffing levels and student outcomes. Our instrumental variables (IV), 2SLS approach takes the following form:

\[
\begin{align*}
\text{Staff}_{d(t)} &= \alpha + \phi S_{At} + \kappa X_{d,t} + \gamma Z_{d,t} \\
&\quad + \lambda E_{d,t} + \delta_{d,t} + \phi_{d,t} + \nu_{d,t} \\
Y_{d,t} &= \alpha + \beta_1 \text{Staff}_{d,t} + \beta_2 X_{d,t} + \beta_3 Z_{d,t} \\
&\quad + \beta_4 E_{d,t} + \delta_{d,t} + \phi_{d,t} + \epsilon_{d,t}.
\end{align*}
\]

As there are four endogenous staffing variables, we estimate four first-stage models that take the form of Equation 2 with \(\text{Staff}_{d(t)}\) representing the observed staffing level in district \(d\) and year \(t\) for one type \((p)\) of staff (such as TAs). The vector \(S_{At}\) denotes state-funded staffing positions for all four staffing categories in 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., state-funded positions aggregated to the district level for each of the staffing types) along with all other controls including general purpose state aid that districts might use for positions that would be identified as state-funded positions.\(^{16}\) See Table 1 for a full list of 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 on the various educational outcomes \(Y_{d,t}\)

In all of our models including our 2SLS models, we weight observations by the number of test-takers (in Grades 3 to 5 or 6 as appropriate) in each district so that our estimated effects of staffing can be interpreted as average effects on students, not on districts. In addition, we cluster standard errors at the district level to address the potential for student outcomes in a district to be correlated over time.\(^{17}\)

This IV approach is based on two main assumptions. The first is that the exogenous, district-level aggregates of the state-funded positions are strong predictors of the total amounts of district-level staffing by staffing type. The first-stage results we report in Supplementary Table S1 in the online version of the journal corroborate that assumption.\(^{18}\) The second is that changes in the number of state-funded staffing positions within districts over time influence outcomes such as student performance only through changes in staffing levels. This assumption is commonly referred to as the exclusion restriction and can be thought of in two related parts (Angrist & Pischke, 2009). The first is that the instrument is as good as randomly assigned. We provide evidence of the plausibility of this assumption in Table 2 in which we regress lagged measures of achievement and demographic characteristics on our instruments. Overall, we fail to detect meaningful associations between these measures and our instruments, assuaging concerns that state-funded staffing flows may be correlated with prior patterns of student achievement or shifting sociodemographic characteristics of districts.\(^{19}\) Moreover, in all our specifications, we directly control for a range of time-varying student characteristics \(X_{d,t}\). These variables include the racial mix of students, the proportions who are economically disadvantaged or have special needs, and the percent female.

The second part of this assumption, which is that the instrument has “no effect on the outcomes other than through the first-stage channel” (Angrist & Pischke, 2009, p. 117), is not directly testable. However, the main threats to the full fulfillment of the exclusion restriction in our setting are indirect, namely compensating behaviors on the part of districts in response to state-funded
| Variable                      | Full sample | Non-White students | White students | High poverty | Middle poverty | Low poverty |
|-------------------------------|-------------|--------------------|----------------|-------------|----------------|-------------|
|                               | M (SD)      | M (SD)             | M (SD)         | M (SD)      | M (SD)         | M (SD)      |
| Test scores                   |             |                    |                |             |                |             |
| Reading score (SD)            | 0.020 (0.190) | -0.355 (0.144)      | 0.307 (0.208)  | -0.302 (0.162) | -0.036 (0.116) | 0.165 (0.146) |
| Math score (SD)               | 0.026 (0.204) | -0.341 (0.161)      | 0.308 (0.230)  | -0.296 (0.177) | -0.036 (0.139) | 0.178 (0.154) |
| Share proficient in reading   | 0.769 (0.115) | 0.652 (0.141)       | 0.858 (0.085)  | 0.673 (0.143)  | 0.755 (0.111)  | 0.807 (0.097)  |
| Share proficient in math      | 0.817 (0.096) | 0.720 (0.118)       | 0.893 (0.070)  | 0.739 (0.123)  | 0.803 (0.095)  | 0.854 (0.072)  |
| Staffing categories           |             |                    |                |             |                |             |
| Teacher assistants/student    | 0.028 (0.012) | 0.026 (0.012)       | 0.029 (0.011)  | 0.035 (0.008)  | 0.025 (0.014)  | 0.030 (0.009)  |
| Regular teachers/student      | 0.051 (0.004) | 0.051 (0.004)       | 0.051 (0.004)  | 0.048 (0.006)  | 0.052 (0.004)  | 0.051 (0.003)  |
| Health staff/student          | 0.002 (0.001) | 0.002 (0.001)       | 0.002 (0.001)  | 0.002 (0.001)  | 0.002 (0.001)  | 0.002 (0.001)  |
| Other staff/student           | 0.006 (0.002) | 0.006 (0.003)       | 0.006 (0.002)  | 0.008 (0.005)  | 0.006 (0.002)  | 0.006 (0.002)  |
| District characteristics      |             |                    |                |             |                |             |
| Economically disadvantaged (%)| 0.478 (0.133) | 0.507 (0.140)       | 0.455 (0.123)  | 0.716 (0.103)  | 0.526 (0.069)  | 0.362 (0.090)  |
| Limited English proficient (LEP, %) | 0.061 (0.045) | 0.067 (0.049)       | 0.056 (0.041)  | 0.069 (0.074)  | 0.062 (0.045)  | 0.058 (0.036)  |
| Black (%)                     | 0.269 (0.169) | 0.334 (0.166)       | 0.218 (0.154)  | 0.454 (0.206)  | 0.321 (0.158)  | 0.159 (0.086)  |
| Hispanic (%)                  | 0.096 (0.060) | 0.106 (0.065)       | 0.089 (0.056)  | 0.106 (0.104)  | 0.097 (0.060)  | 0.093 (0.046)  |
| White (%)                     | 0.565 (0.193) | 0.479 (0.175)       | 0.631 (0.180)  | 0.314 (0.144)  | 0.521 (0.185)  | 0.678 (0.126)  |
| Other (%)                     | 0.070 (0.068) | 0.081 (0.089)       | 0.062 (0.042)  | 0.126 (0.198)  | 0.062 (0.029)  | 0.070 (0.037)  |
| Female (%)                    | 0.493 (0.010) | 0.493 (0.010)       | 0.492 (0.010)  | 0.495 (0.018)  | 0.493 (0.010)  | 0.492 (0.008)  |
| Special needs (%)             | 0.114 (0.030) | 0.112 (0.030)       | 0.116 (0.029)  | 0.101 (0.029)  | 0.112 (0.026)  | 0.120 (0.033)  |
| Unemployment rate (%)         | 7.204 (2.728) | 7.247 (2.711)       | 7.170 (2.739)  | 9.048 (3.086)  | 7.381 (2.692)  | 6.569 (2.466)  |

(continued)
| Variable                                           | Full sample | Non-White students | White students | District subsamples |
|----------------------------------------------------|-------------|--------------------|----------------|--------------------|
|                                                   | (1)         | (2)                | (3)            | (4)                | (5)                | (6)                |
| Median household income (US$1,000s)                | 44.239 (8.654) | 44.605 (9.046)   | 43.956 (8.327) | 33.48 (5.337)     | 42.201 (6.321)     | 49.292 (8.715)    |
| Charter schools (#)                               | 0.665 (0.999) | 0.781 (1.042)   | 0.575 (0.955)  | 0.007 (0.081)     | 0.631 (0.835)     | 0.852 (1.220)     |
| Charter school students (%)                        | 0.006 (0.013) | 0.006 (0.013)   | 0.005 (0.013)  | 0.001 (0.013)     | 0.007 (0.015)     | 0.005 (0.010)     |
| Low-wealth aid (dollars per student)              | 102.1 (156.7) | 108.8 (166.7)   | 97.0 (148.4)   | 356.2 (166.0)     | 97.9 (150.5)      | 53.3 (101.2)      |
| Small county aid (dollars per student)            | 21.2 (133.4)  | 18.0 (129.0)    | 23.6 (136.5)   | 95.2 (328.9)      | 22.7 (119.3)      | 3.2 (41.1)        |
| Teacher characteristics                           |             |                   |                |                    |                    |                    |
| Average Praxis score                               | 0.061 (0.178) | 0.020 (0.204)   | 0.093 (0.148)  | -0.307 (0.205)    | 0.027 (0.105)     | 0.187 (0.105)     |
| >3 years of experience (%)                         | 0.799 (0.062) | 0.793 (0.063)   | 0.803 (0.061)  | 0.808 (0.068)     | 0.795 (0.065)     | 0.801 (0.055)     |
| Turnover rate (%)                                  | 0.212 (0.056) | 0.221 (0.055)   | 0.205 (0.055)  | 0.218 (0.072)     | 0.215 (0.058)     | 0.206 (0.047)     |
| Number of students Grades 3–6                      | 9,560 (10,117) | 11,105 (10,646) | 8,367 (9,519)  | 1,667 (983)       | 9,516 (9,242)     | 11,311 (11,424)   |
| District (PK/K-6)                                  | 20,053 (21,336) | 23,348 (22,422) | 17,511 (20,093) | 3,547 (2,067)     | 19,981 (19,394)   | 23,690 (24,226)   |
| Number of observations                             | 1,149        | 1,149             | 1,149          | 277               | 578               | 294               |
| Number of districts                                | 102          | 102               | 102            | 26                | 50                | 26                |

Note. Sample includes students in traditional public elementary schools in North Carolina from 2001 to 2012 with grade configurations PK–5, PK–6, K–5, and K–6 aggregated by district. The “health staff” variable includes speech specialists, social workers, and health services personnel; it does not include psychologists or nurses. The “other staff” variable includes school leaders such as principals, school counselors, and other noncertified academic support staff such as tutors. Means for the full sample and the district subsamples are weighted by the total number of test-takers (Grades 3–6) in the district, whereas means for the student subgroups are weighted by the number of test-takers in that student group in a district. For Columns 4 to 6, we group districts into subsamples based on the average percentage of economically disadvantaged students for each district over the study period: Districts in the “high” group fall in the top quarter of the distribution of the average share of economically disadvantaged students; districts in the “low” group fall in the bottom quarter; districts in the middle constitute the middle 50% of that distribution. District-level data on the local unemployment rate come from the Bureau of Labor Statistics (BLS) and data on median family income come from the U.S. Census Bureau SAIPE (Small Area Income and Poverty Estimates).
staffing changes that may also influence student achievement.

As a test of the degree to which we observe such offsetting behavior with respect to staffing for TAs, Figure 4 plots year-to-year changes in TAs funded by the state against year-to-year changes in locally funded TAs. We focus on the years surrounding the recession, 2007 to 2011, and fail to observe a systematic relationship between declines in state-funded TAs and increases in locally funded TAs. Thus, it seems that the majority of districts, including those with notable declines in state-funded TAs, do not meaningfully offset such declines with locally funded positions of the same type. This provides indirect evidence that most districts are not pulling dollars from other inputs to counteract declines in state-funded TAs. Therefore, our estimates of the effects of TAs on student outcomes ought to pick up little of such indirect effects.

Another possible concern is that changes in state-funded staffing may affect student outcomes at the district level through changes in the quality, in contrast to the quantity, of the staff. We address this concern for teachers largely by including in the model variables designed to control for their quality (Z_{it}). These time-varying variables are their average licensure (i.e., Praxis) test scores, the proportion of teachers with more than 3 years of experience, and the average turnover rate of teachers at the school level. Research studies based on North Carolina data confirm that teachers with higher Praxis scores and more experience are predictive of higher student test scores (Clotfelter et al., 2006, 2007). Recent research documents that teacher turnover leads to higher proportions of teachers with weak qualifications and lower student test scores (Ronfeldt et al., 2013; Sorensen & Ladd, 2020). Two time-varying, county-level variables also help to control for quality, in this case the quality of all staffing categories not just those of teachers. These are the local unemployment rate and median family income. The logic is that changes in both variables affect the opportunity costs of working in the public schools. When the local unemployment rate rises (or median family income falls), for example, the salaries for teachers and staff, which are relatively fixed, look more attractive than when the local labor market is tighter and is offering higher wages (Nagler et al., 2020). As a result, schools may be able to hire more qualified employees.

Finally, if sizable numbers of families were systematically moving their children out of traditional public schools in favor of charter schools or were moving between districts in response to aggregate performance measures, we might inappropriately attribute changes in achievement caused by the cross-sector or cross-district movement of families to changes in staffing. We address possible concerns about cross-sector shifting by including in our models controls for the presence of charter schools, as measured at the district-year level by the number of charter schools serving elementary school grades and the proportions of students in charter schools. First established in North Carolina in 1996, charters reached the legislated limit of 100 schools by 2011, at which time the limit was raised. During the sample period, most districts had very few charters (with the average over the full sample period being only about two-thirds of a charter school, as shown in Table 1), although some had several.

**Descriptive Statistics**

Table 1 provides descriptive statistics for the main analytic sample. The test scores of the students (standardized by grade and year) refer only to students in Grades 3, 4, and 5 or 6 because those are the only students who are tested. All the staffing variables, however, apply to all the students in the analytic sample for each district.
Consider first the data for the full sample of elementary schools in Column 1. The mean teacher-to-student ratio is 0.051, that is, 5.1 teachers per 100 students. If each regular teacher had her own class, that would translate into an average class size of about 19 to 20 students, albeit most likely with smaller class sizes in the early grades and larger ones in the upper grades. The mean for TAs is slightly more than half that, at 2.8 per 100 students. Health providers are much less numerous, with only about 0.2 per 100 students, or about one provider per 500 students. The “other” combined category of school counselors, school leaders, and noncertified academic staff is somewhat larger, with 0.6 per 100 students or about one per 167 students.

Table 1 also includes a number of district-level variables that capture characteristics of students, the wider economy, and the two other sources of general state aid, as well as a few teacher characteristics. The student characteristics indicate that slightly less than half the students are economically disadvantaged,23 27% are Black, about 10% are Hispanic, and about 11% have special needs. The teacher characteristics indicate that teachers have average licensure test scores (i.e., Praxis scores) slightly above the normalized value of 0 for all teachers and about 80% of the teachers have more than 3 years of experience. The average teacher turnover rate at the school level is 21%.24 The next two columns of Table 1 show comparable information for subgroups of students—students of color (defined as Black, Hispanic, and students classified in other non-White categories) and White students. The staffing variables are quite similar across these two groups of students as well as to those in the full sample.

The final three columns of Table 1 provide comparable information for subsamples of districts grouped by the average share of economically disadvantaged students.25 The staffing variables are generally comparable with those in the full sample, although the high-poverty districts have a slightly higher ratio of TAs to students, but a lower ratio of teachers to students and the teachers are of lower quality as measured by their lower average Praxis scores.

### TABLE 2
Assessing the Validity of the Instrument

| Independent variable                                      | Teacher assistants | Teachers | Health staff | Other staff |
|-----------------------------------------------------------|--------------------|----------|--------------|-------------|
| Lagged reading test scores \( (t - 1) \)                 | 0.005 (0.005)      | 0.001 (0.004) | 0.001 (0.001) | 0.001 (0.001) |
| Lagged math test scores \( (t - 1) \)                   | 0.003 (0.004)      | 0.001 (0.004) | 0.000 (0.001) | 0.001 (0.001) |
| Lagged LEP % \( (t - 1) \)                               | -0.009 (0.019)     | 0.014 (0.018) | 0.003 (0.004) | -0.007 (0.006) |
| Lagged economically disadvantaged % \( (t - 1) \)       | -0.003 (0.004)     | -0.003 (0.004) | -0.000 (0.001) | -0.001 (0.002) |
| Lagged Black % \( (t - 1) \)                             | -0.006 (0.008)     | -0.008 (0.006) | 0.001 (0.002) | 0.006* (0.004) |
| Lagged Hispanic % \( (t - 1) \)                          | 0.022 (0.016)      | -0.006 (0.017) | -0.003 (0.003) | 0.006 (0.005) |
| Lagged other % \( (t - 1) \)                             | -0.009 (0.016)     | -0.017 (0.019) | -0.003 (0.004) | -0.006 (0.009) |
| Lagged female % \( (t - 1) \)                            | 0.018 (0.012)      | -0.001 (0.011) | -0.002 (0.002) | -0.006 (0.006) |
| Lagged special needs % \( (t - 1) \)                     | 0.007 (0.013)      | 0.004 (0.011) | 0.003 (0.002) | 0.011 (0.008) |
| Number of observations                                   | 1,149              | 1,149     | 1,149        | 1,149       |
| Number of districts                                      | 102                | 102       | 102          | 102         |

Note. Sample includes students in traditional public elementary schools in North Carolina from 2001 to 2012 with grade configurations PK–5, PK–6, K–5, and K–6 aggregated by district. Staffing variables are measured in per-student terms. All models include district and year fixed effects. Standard errors appear in parentheses and are clustered by district. LEP = limited English proficient.

* \( p < .1. ** p < .05. *** p < .01. ** p < .05. *** p < .01. **
Findings

Core Results

Table 3 reports the main results for reading and math test scores, contrasting estimates of staffing effects that emerge from our three models: OLS, RF, and 2SLS. All estimates come from specifications that include all of the control variables (see Supplementary Table S2 in the online version of the journal). In the first row of Table 3 are the estimates for TAs, the main staffing category of interest in this study, followed by estimates for teachers, health staff, and other staff.

Consider first the OLS results for reading and math (see Columns 1 and 4, respectively). For reading, the OLS coefficients for TAs, regular teachers, and health staff are all positive and statistically significant and for math all are positive, although the TA coefficient is just short of statistical significance. Taken at face value, these positive coefficients imply that these three types of staffing contribute to student achievement. In contrast, the category of other staff enters with small and insignificant positive coefficients.

The reported coefficients can be interpreted as the effect on student achievement of a one unit increase per 100 students in the specific staffing category, but the relative magnitudes can be misleading because of the differing average staffing levels across categories. The very low average ratio of health care workers per student means, for example, that such an increase for health care workers represents a far larger percentage increase than it does for teachers or for TAs. Thus, to help with the interpretation of the coefficients and to make them more comparable across staffing categories, we show in square brackets below each estimate in this table and in subsequent tables the implied effects of a 10% increase in the staffing variable starting from the average. For example, based on the OLS model, a 10% increase in TAs is predicted to increase reading scores by 0.004 standard

### Table 3

**Reading and Math Test Scores: OLS, RF, and 2SLS Results, 2001–2012**

| Independent variable | Reading | | | Math | | |
|-----------------------|---------|---------|---------|-------|---------|---------|
|                       | OLS     | RF      | 2SLS    | OLS   | RF      | 2SLS    |
| Teacher assistants    | 1.426*** | 1.860*** | 1.725*** | 0.854 | 1.355** | 1.261*  |
| (0.356)               | (0.400) | (0.369) | (0.520) | (0.660) | (0.672) |         |
| [0.004]               | [0.005] | [0.005] | [0.002] | [0.004] | [0.004] |         |
| Regular teachers      | 2.094**  | 0.675    | 0.698    | 3.654** | 1.317    | 1.659    |
| (0.991)               | (0.808) | (1.193) | (1.509) | (1.245) | (1.819) |         |
| [0.011]               | [0.003] | [0.004] | [0.019] | [0.007] | [0.008] |         |
| Health staff          | 9.884**  | 8.161**  | 11.204** | 18.702*** | 18.152*** | 24.891*** |
| (4.572)               | (3.898) | (5.229) | (6.244) | (5.147) | (7.189) |         |
| [0.002]               | [0.002] | [0.002] | [0.004] | [0.004] | [0.005] |         |
| Other staff           | 0.929    | 0.125    | 0.192    | 2.124  | 2.068    | 2.009    |
| (1.562)               | (1.418) | (1.424) | (1.858) | (1.730) | (1.791) |         |
| [0.001]               | [0.000] | [0.000] | [0.001] | [0.001] | [0.001] |         |
| Number of observations | 1,149  | 1,149    | 1,149    | 1,149  | 1,149    | 1,149    |
| Number of districts    | 102    | 102      | 102      | 102    | 102      | 102      |

Note. Sample includes students in traditional public elementary schools in North Carolina from 2001 to 2012 with grade configurations PK–5, PK–6, K–5, and K–6 aggregated by district. Each model also includes all controls listed in Table 1 as well as year and district fixed effects. Staffing variables are expressed per student. Instruments are state-funded positions of a given staffing type. See text for additional details on the IV/2SLS strategy. Each observation is weighted by the (subject-specific) number of test-taking students in the district, and standard errors (in parentheses) are clustered at the district level. OLS = ordinary least squares; RF = reduced form; 2SLS = two-stage least squares.

*p < .1. **p < .05. ***p < .01.
deviations (SD), a 10% increase in regular teachers by 0.011 SD, and a 10% increase in health staff by 0.002 SD.

Although these OLS findings appear reasonable, we suspect they may be biased, with the estimates for TAs being downward biased and for regular teachers being upward biased. The concern for teachers is that despite all the control variables in the equation, there may still be a positive correlation between the unexplained error term and the number of teachers per student. That would be the case if districts with high-achieving students (or more specifically, those with positive unexplained variation in student test scores) are located in wealthier areas willing to provide greater local funding for teachers than other areas. The bias for TAs will go in the opposite direction if districts with many low-performing students are able to attract more funding from the federal government for TAs than are other districts. In that case, the additional funding would bias downward the estimated effect of TA funding on student achievement. While we are not willing to reject the OLS results, outright, this type of reasoning leads us to prefer results that emerge from the RF and 2SLS models, especially for the estimates of TAs.

As reported in the first row of the table, the RF and 2SLS estimates for TAs are positive and statistically significant for both subjects. The fact that the estimates are somewhat larger than their OLS counterparts is consistent with our expectation that the OLS estimates of TAs would be downward biased. However, all three sets of estimates for reading are clearly in the same ballpark, suggesting that our use of district and year fixed effects along with rich time-varying controls likely to be correlated with TA levels (from all funding sources) and student achievement go a long way toward producing defensively causal estimates of the effects of TAs on student performance.

In contrast, the RF and 2SLS estimates for regular teachers are notably smaller than the companion OLS estimate and are statistically indistinguishable from zero for both reading and math. This pattern is consistent with our speculation that well-resourced, higher-performing districts may use local dollars to hire additional teachers in ways that would bias upward the OLS estimates. However, recall the relatively flat trend in state-funded teacher positions depicted in Figure 2. Over our period, the state pursued a teacher allocation strategy designed to maintain reasonable class sizes throughout the recession—and thus our instrument, state-funded teaching positions over time within a district, may be limited in its capacity to identify the effects of teachers on student outcomes. Thus, a reasonable estimate of the causal effects of teacher staffing levels on student achievement likely falls somewhere between our 2SLS/RF and OLS results. Finally, the three approaches generate very similar results for the health staff and other staff variables. In all cases, higher levels of health staff lead to higher levels of both reading and math achievement, whereas higher levels of “other staff” have insignificant effects in all cases.

While some readers may prefer the RF estimates to the 2SLS estimates on the ground that the former do not require the exclusion restriction necessary for the latter, we generally prefer the 2SLS estimates because they can be interpreted as the effect of exposure to additional staff of a particular type on student achievement. Moreover, basic patterns of findings do not differ between the RF and 2SLS sets of results. Hence, we focus attention here on the 2SLS estimates in Columns 3 and 6 of this table, and in subsequent tables we report only the 2SLS results (but see Supplementary Tables S3–S5 in the online version of the journal for a full set of RF estimates).

As already noted, the figures in square brackets below each estimate indicate the effects of 10% increases. Hence, the 2SLS results imply that a 10% increase in TAs, that is, an increase from 2.8 to 3.1 per 100 students, would increase test scores by 0.005 standard deviations in reading and 0.004 standard deviations in math, effects that are slightly larger than those from the OLS models. Because we have no information on the specific activities of TAs within classrooms, we are not able to explore the mechanisms that generate these positive effects. For example, TAs may contribute to student achievement either by their direct interactions with students or by freeing up time for teachers to reallocate their time and effort to the core tasks of teaching.
The comparable results for health care workers are 0.002 standard deviations for reading and 0.005 for math. To assuage any concerns about the inconsistency over time in the specific types of staffing positions that we were able to count as “health staff,” we have also estimated our preferred 2SLS models without that staffing variable. As shown in Supplementary Table S2 in the online version of the journal, this deletion has little influence on the estimated effects of TAs. Despite definitional challenges not only for the health staff variable but also for “other staff,” we prefer to include both variables in all the models to avoid any bias that might arise from correlations among the staffing categories.

Despite definitional challenges not only for the health staff variable but also for “other staff,” we prefer to include both variables in all the models to avoid any bias that might arise from correlations among the staffing categories.28

We conclude this section on core results by noting that most of the time-varying measures of demographic and teacher characteristics enter the equations as expected (see Supplementary Table S2 in the online version of the journal for the full models). Thus, they appear to serve as reasonable statistical controls for any district-level changes in the composition of students or teachers that might be correlated with changes in staffing levels.29 The two time-varying controls that capture changes in local economic conditions—the local unemployment rate and median household income, both defined at the county level—also enter in expected ways. These controls are important given that much of the intertemporal variation that we exploit comes from cutbacks in TAs during the recession. Given that the timing of such cutbacks is correlated with recessionary pressures across districts, their effects could be confounded by the effects of such pressures on the local labor market for school staff. Consistent with an earlier study that focused on the relationship between recessionary pressures and teacher quality (Nagler et al., 2020), the positive coefficient on the unemployment rate (and a negative coefficient on median income) suggests that schools are able to hire more qualified staff when job opportunities in the local private sector economy are limited. Thus, if the pool of TA applicants markedly changed in a transient manner during the recession along dimensions not captured by our control variables, our results may not be as applicable to times when the economy is stronger. This could potentially limit the external validity of our findings.

Robustness Checks and Additional Threats to Internal Validity

In this section, we first test the robustness of our basic findings and then explore some of the remaining potential threats to the internal validity of our approach.

The robustness checks appear in Table 4. Columns 1 and 4 of the table repeat our preferred 2SLS results from Table 3. We then examine the robustness of those results to (a) limiting the sample to a fully balanced set of districts, and (b) adding district-specific linear time trends, which control for the evolution of any remaining, hard-to-measure within-district forces possibly correlated with both changes in staffing levels and student achievement. As we move to a balanced sample of only 89 districts, our basic conclusions about the effectiveness of TAs in both reading and math hold up. With the addition of the district-specific linear time trends, the estimate for TAs in reading falls somewhat but retains its statistical significance. In contrast, the estimate for math is small and no longer statistically significant. Thus, the effect of TAs on student achievement in reading is more robust than the effect in math.

We note as well that the addition of the time trends substantially reduces the estimated effects of health staff in both subjects at the same time that it raises the coefficients for other staff. Given that the models already include a large number of time-varying control variables, all of which were included for specific theoretical reasons, we interpret these unexpected results with some caution and simply acknowledge that our estimates of the positive effects of health staff in other models are not robust to the addition of linear time trends.30

The final column of Table 4 includes a falsification test. For this test, we start with the preferred specification of our 2SLS estimator but use the share of economically disadvantaged students as the dependent variable. If our instruments have allowed us to carve out plausibly exogenous variation in staffing levels, such variation ought not to be correlated with changes in the socioeconomic characteristics of districts. We see no evidence of such concerns with respect to TAs, health staff, and other staff, but some evidence of a marginally significant negative relationship between regular
teachers and the share of economically disadvantaged students. On net, this evidence generally supports our identifying assumptions.

**Additional Results**

The literature we reviewed earlier suggests that TAs may have larger effects for traditionally underserved students than for other students. We examine such differences in Table 5 first by exploring the degree to which effects of TAs and other staff differ for students of color (i.e., Black, Hispanic, and other non-White students) as compared with White students. The first four columns of Table 5 show that the effects of TAs on test scores in both math and reading are larger for non-White students than for White students. In particular, 10% increases in TAs raise the test scores of non-White students by 0.01 SDs in reading and by 0.008 SDs in math, effects that are larger than the comparable estimates of 0.003 SD and 0.001 SD for White students.

In addition, we examine in Columns 5 to 10 the extent to which the effectiveness of TAs differs across subsamples of districts defined by their poverty rates. The results indicate that estimated coefficients for TAs in the high-poverty districts (5.04 in reading and 8.86 in math) far exceed the estimated coefficients in the other subsamples of districts. The large standard errors for the TA estimates in the high-poverty subsample, however, renders the coefficient for reading just short of statistical significance. Smaller but positive and statistically significant effects are also apparent for TAs in reading, but not in math, in both middle- and low-poverty districts. Finally, we note that regular teachers also enter with large and positive statistically significant effects in the high-poverty districts, a pattern that is consistent with prior work that finds relatively larger benefits of smaller class sizes for disadvantaged students (e.g., Dynarski et al., 2013).
### TABLE 5

**Test Score Results for Staff Variables, by Subgroup and Subsample, 2SLS Estimator, 2001–2012**

| Independent variable | Student subgroups | District subsamples |
|-----------------------|------------------|--------------------|
|                       | Non-White students | White students | High poverty | Middle poverty | Low poverty |
|                       | Reading (1) | Math (2) | Reading (5) | Math (6) | Reading (9) | Math (10) |
| Teacher assistants    | 3.726***    | 3.161***   | 1.050***   | 0.428     | 5.044     | 8.863**   |
|                       | (0.613)     | (0.991)    | (0.385)    | (0.567)   | (3.076)   | (3.703)   |
|                       | [0.010]     | [0.008]    | [0.003]    | [0.001]   | [0.018]   | [0.031]   |
| Regular teachers      | 1.009       | 2.289      | 0.483      | 1.191     | 2.104**   | 4.231***  |
|                       | (1.328)     | (1.808)    | (1.226)    | (1.830)   | (1.026)   | (1.257)   |
|                       | [0.005]     | [0.012]    | [0.002]    | [0.006]   | [0.010]   | [0.020]   |
| Health staff          | 9.673*      | 31.521***  | 10.408*    | 20.397*** | 0.423     | 9.722     |
|                       | (5.726)     | (8.779)    | (5.984)    | (7.767)   | (12.906)  | (15.018)  |
|                       | [0.002]     | [0.006]    | [0.002]    | [0.004]   | [0.000]   | [0.002]   |
| Other staff           | 0.470       | 2.054      | -0.808     | 1.302     | 0.410     | 0.065     |
|                       | (1.412)     | (1.812)    | (2.172)    | (2.386)   | (1.372)   | (1.889)   |
|                       | [0.000]     | [0.001]    | [0.000]    | [0.001]   | [0.000]   | [0.000]   |

| F statistics from first-stage models by endogenous staffing outcome: |
|------------------------------------------------------------------|
| Regular teachers | 31.10 | 31.29 | 38.82 | 38.83 | 289.56 | 288.72 | 16.58 | 16.62 | 22.09 | 21.97 |
| Teacher assistants | 63.41 | 63.49 | 37.50 | 37.60 | 72.94 | 73.00 | 158.44 | 159.47 | 31.45 | 31.55 |
| Health staff | 67.98 | 67.72 | 56.82 | 56.90 | 15.28 | 15.35 | 96.32 | 96.33 | 18.12 | 18.03 |
| Other staff | 196.07 | 198.10 | 123.24 | 123.66 | 248.88 | 251.90 | 56.71 | 56.61 | 28.08 | 27.96 |

| Number of observations | 1,149 | 1,149 | 277 | 578 | 294 |
| Number of districts | 102 | 102 | 26 | 50 | 26 |

**Note.** Sample includes students in traditional public elementary schools in North Carolina from 2001 to 2012 with grade configurations PK–5, PK–6, K–5, and K–6 aggregated by district. Each column comes from a separate 2SLS regression that includes all the control variables listed in Table 1, plus year and district fixed effects. The staffing variables are expressed per student. For the student subgroups, each district-level observation is weighted by the number of test-takers in that group in the district, whereas for the district subsamples, each district-level observation is weighted by the total number of test-takers in the district. Standard errors clustered by district are in parentheses and estimated effects for 10% changes in the staffing variables are in square brackets. For Columns 5 to 10, we group schools into subsamples based on the average percentage of economically disadvantaged students for each district over the study period: Districts in the “high” group fall in the top quarter of the distribution of the average share of economically disadvantaged students; districts in the “low” group fall in the bottom quarter; districts in the middle constitute the middle 50% of that distribution. 2SLS = two-stage least squares. **p < .1. ***p < .05. ****p < .01.
proportion of tested students who score at or above subject- and grade-specific proficiency levels. We report 2SLS results for all students and the same subgroups and subsamples as in Table 5. The averages for the composite proficiency rates are 76.9% in reading and 81.7% in math. In contrast to average test score measures that incorporate all student test scores, this measure 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 requirements of NCLB. This federal law put pressure on 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 on the borderline of passing the tests, perhaps to the detriment of students who are likely to score well above or below the proficiency cut point (Ladd & Lauen, 2010; Lauen & Gaddis, 2016). Additional TAs 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 TAs—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.

Table 6 shows that additional TAs raise the shares of all students scoring proficiently in both reading and math, and that the effects are about twice as large for non-White students as for White students. In addition, the effects are somewhat larger in reading than in math. Across the district subsamples, the largest estimates, once again, are for the high-poverty subsample, although large standard errors render the reading coefficient statistically insignificant, and the math coefficient only marginally significant. Together, these results reinforce our earlier findings about the positive impact of TAs on academic outcomes, especially for non-White students, and in reading.

### Are TAs Cost-Effective?

TAs have lower salaries than teachers, hence TAs need not be as effective as teachers to make them cost-effective. The question here is whether TAs are sufficiently effective relative to teachers to make hiring more TAs a cost-effective investment. For the following analysis, we assume that the only criterion for hiring TAs is their contribution to student test scores. This is clearly an oversimplification given that the educational goals of schools, and, more specifically, of both teachers and TAs, extend far beyond test scores (e.g., Jackson, 2012).

North Carolina has a statewide salary schedule that differentiates teacher salaries based on a teacher’s educational background and years of experience. For the following calculation, we use a monthly salary of US$4,400 which is the 2019 salary of a teacher with a master’s degree and 5 years of experience. The state also specifies a salary range for TAs (Level 56 category on the 2019 salary schedule) of US$2,100 to US$3,300. Thus, if we assume that a typical TA is paid at the midpoint of that range, the TA’s salary would be US$2,700, or about 61% of that of a teacher.

We use the upper-bound, OLS estimates of the effects of teachers on student achievement in math and reading. This provides conservative estimates of effectiveness with which to assess the relative cost-effectiveness of TAs. Thus, we compare the following estimates from Table 3:

|                        | Reading | Math |
|------------------------|---------|------|
| Teacher assistants     | 1.725   | 1.261|
| (2SLS estimate)        |         |      |
| Regular teachers       | 2.094   | 3.654|
| (OLS estimate)         |         |      |
| Ratio TA/Teacher       | 82%     | 35%  |

*Note. 2SLS = two-stage least squares; OLS = ordinary least squares; TA = teacher assistant.*

Based on these figures, TAs would appear to be a cost-effective investment for reading given the 82% effectiveness ratio is well above the 61% salary ratio, but they would not be cost-effective for math. Alternatively, a comparison of the average effectiveness ratio of 59% \((=(82+35)/2)\) across the two subjects to the salary ratio indicates that TAs are roughly cost-effective on average given their contributions to both subjects. In addition, we remind the reader that the OLS estimates for regular teachers are most likely biased upward. Taking that upward
TABLE 6  
School Staffing and Academic Proficiency, 2SLS Estimator, 2001–2012

| Independent variable | Share proficient in reading | Share proficient in math |
|----------------------|-----------------------------|-------------------------|
|                      | All students                | Non-White students      | White students            | High poverty | Middle poverty | Low poverty | All students | Non-White students | White students | High poverty | Middle poverty | Low poverty |
| Teacher assistants   | (1) 0.417** (0.213) [0.001] | (2) 1.434*** (0.268) [0.004] | (3) 0.525*** (0.187) [0.002] | (4) 1.358 (1.304) [0.005] | (5) 0.632** (0.247) [0.002] | (6) 0.632** (0.274) [0.002] | (7) 0.499* (0.261) [0.001] | (8) 1.149*** (0.387) [0.003] | (9) 0.416* (0.219) [0.001] | (10) 2.556* (1.421) [0.009] | (11) 0.411 (0.283) [0.001] | (12) 0.503 (0.403) [0.001] |
| Regular teachers     | (2) −0.020 (0.583) [0.000] | (3) 0.186 (0.653) [0.001] | (4) 0.086 (0.547) [0.000] | (5) 1.351*** (0.449) [0.006] | (6) 0.675 (0.583) [0.004] | (7) −2.501 (2.912) [0.000] | (8) 0.414 (2.920) [0.000] | (9) 1.126 (2.956) [0.000] | (10) 0.248 (3.373) [0.000] | (11) 1.377*** (3.320) [0.000] | (12) 1.243 (3.434) [0.000] |
| Health staff         | (1) 1.699 (2.912) [0.000] | (2) 1.582 (2.920) [0.000] | (3) −1.474 (2.956) [0.000] | (4) 0.514 (5.337) [0.000] | (5) 4.708 (3.732) [0.000] | (6) −5.063 (5.663) [0.000] | (7) 9.228*** (3.062) [0.000] | (8) 13.779*** (3.432) [0.000] | (9) 3.111 (3.204) [0.000] | (10) 8.384 (5.182) [0.000] | (11) 12.770*** (5.422) [0.000] | (12) −2.034 [−0.001] |
| Other staff          | (2) 0.263 (0.865) [0.000] | (3) 0.386 (0.601) [0.000] | (4) −0.547 (0.900) [0.000] | (5) 0.539 (0.545) [0.000] | (6) 2.216 (1.866) [0.000] | (7) −5.677*** (2.344) [−0.003] | (8) 1.100* (0.668) [0.001] | (9) 1.060 (0.681) [0.001] | (10) −0.190 (1.066) [0.000] | (11) 0.314 (0.626) [0.000] | (12) 1.535 (2.063) [0.000] |
| Number of observations | 1,149 1,149 1,149 277 578 294 | 1,149 1,149 1,149 277 578 294 |
| Number of districts  | 102 102 102 26 50 26 | 102 102 102 26 50 26 |

Note. Sample includes students in traditional public elementary schools in North Carolina from 2001 to 2012 with grade configurations PK–5, PK–6, K–5, and K–6 aggregated by district. Each column comes from a separate 2SLS setup that includes all the control variables listed in Table 1, plus year and district fixed effects. The staff variables are expressed per student. Each district-level observation is weighted by the number of test-taking students in the district, and standard errors are clustered at the district level. Standard errors are in parentheses and estimated effects of a 10% change in staff are in square brackets. 2SLS = two-stage least squares. *p < .1. **p < .05. ***p < .01.
bias into account would further strengthen the argument that TAs are cost-effective, even based on the limited definition of effectiveness used here.

**Summary and Discussion**

Children interact with a range of adults in schools. Research in the social sciences has largely focused on the role of teachers in shaping student achievement. Teachers are of central importance to the work of educating students, but by no means function in isolation. In this article, we expand the lens of inquiry to study the effects of TAs and other staff members on student outcomes using data from elementary schools in North Carolina.

It is difficult to determine how staffing levels affect student outcomes given that levels of staffing are typically not exogenous to schools and districts. 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 districts serving more advantaged and higher-performing students and downward biased if staffing levels were higher in districts serving more disadvantaged and lower-performing students.

We use a new strategy for addressing such bias-inducing reverse causation to estimate the effects of TAs on student test scores. 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 district level with state-determined changes in funded positions to districts. Given the large reduction in state funding for TA positions during our sample period, this two-stage approach—combined with district and year fixed effects, time-varying demographic characteristics and measures of teacher quality, as well as county-level economic controls—generates plausibly causal estimates of the effects of TAs on student outcomes. Although we have pursued a variety of strategies to limit the confounding influence of local economic changes, to the extent that the recession broadly boosted the quality of the TA pool in a transitory fashion along dimensions we cannot capture, our results might not hold in periods with a stronger economy. This points to the need for additional work on the characteristics of TAs that lead to improved student outcomes.

We find evidence that TAs have positive effects on student test scores in reading and math, with the largest, consistent, and most robust effects in reading. As we illustrated above, the effects of TAs in that subject are sufficiently large to show that hiring TAs can be a cost-effective way to improve test scores in reading. In addition, we find that TAs have larger positive effects on the academic performance of students of color and suggestive evidence that the effects are higher in high-poverty districts than in other types of districts. Finally, we also find evidence that TAs help to boost school-level proficiency rates, particularly among non-White students and those in high-poverty districts.

Our strong and quite consistent story about the importance of TAs counters the relatively more pessimistic, standard view about TAs that emerged from the highly cited Tennessee STAR experiment during the 1980s. However, our results are quite consistent with much more recent experimental evidence from Denmark that finds positive effects of TAs on reading scores (Andersen et al., 2018). The role of TAs in the United States has likely evolved since the 1980s, with perhaps more complementary use of TAs as “part of a professional team who share responsibility for a classroom rather than help with practical tasks” (Andersen et al., 2018, p. 35). Thus, although we cannot observe specific activities of TAs in our study, our results are consistent with such a view.

Increases in health staff also appear to lead to higher test scores, with the largest effects in math and in middle-poverty districts. We are careful not to oversell these health staff results because our measure of health care staffing within schools excludes some important personnel such as nurses and school psychologists. However, these estimates are suggestive of positive contributions of health care providers to student outcomes at the elementary school level.

We remind the reader that student achievement as measured here by test scores in reading and math is not the only goal of schooling, and that teachers, TAs, health providers, and other staff members may also contribute to student outcomes.
in ways that we have not measured, including in the form of achievement in other subjects, as well as behavioral and other non-test-score outcomes. More research is needed to understand the ways in which TAs and other support staff contribute to valued outcomes for students.

Acknowledgments
We thank Charles Clotfelter and Kris Nordstrom for comments on earlier drafts. We are grateful to two anonymous referees for very thoughtful suggestions that improved the paper. Winnie Biwott, Adrienne Jones, Ying Shi, and Mavzuna Turaveva provided excellent research assistance for this project.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), which is funded by a consortium of foundations. For more information about CALDER funders, see www.caldercenter.org/about-calder.

Notes
1. Statistics on annual income and poverty rates by state (and year) can be found through the U.S. Census Bureau: https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t
2. Data on educational spending come from the U.S. Census Bureau, Public Education Finances, and data on teacher salaries are available from the National Education Association (NEA), Rankings of the States 2012 and Estimates of School Statistics 2013.
3. See results here: http://www.ncpublicschools.org/accountability/policies/naep/2.
4. 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 as 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.
5. Periodically the state has permitted local education agencies some flexibility in the use of the personnel allotments—but such flexibility was typically used to “game the system.” By labeling an individual as a state-funded employee, for example, the district would receive the average state salary and benefits for that person, and then could use other local funding sources to pay for individuals making salaries below the state average.
6. Dollar allotments are also used for central office administration, textbooks, and classroom materials, supplies and equipment.
7. There are 100 counties in North Carolina and 115 school districts.
8. Some counties also use small amounts to hire additional staff including teacher assistants (TAs) 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 accumulated from federal American Recovery and Reinvestment Act (ARRA) funds after the 2008 recession to offset cuts in positions funded by the state.
9. In fact, as we noted earlier in Note 1, a per-student allocation of one regular teacher per 18 students translates into a class size somewhat larger than 18 because the allotment includes positions devoted to 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.
10. 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) finds large effects of class size reductions on students in Grades 4 and 5, while Piketty and Valdenaire (2006) 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).
11. 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 the North Carolina School Social Workers’ Association, n.d.).
12. We excluded from all the analysis an additional category of nonregular teachers on the grounds that they were working with special populations.
13. With the rise in student assessment required by federal and state laws, many school 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% of their time providing direct services to students, not counting any time they spend in test coordination.

14. Although the state has 115 districts, some have no schools offering only the grades in our specified range, and thus our final sample is based on 102 districts. Moreover, not all the districts have such schools in every year.

15. We include the number of charter schools (at the elementary level) in a given district and year as well as the share of overall district enrollment accounted for by those schools as control variables in our main model, to assuage any concerns about the effects of the (growing) charter sector on outcomes or staffing decisions of local public schools. Results from models that omit these two controls are extremely similar and available from the authors upon request.

16. We would have preferred the more direct approach of purging from our measures of state-funded positions any that were financed from the low-wealth or small county aid, but the data were not available to do so. Our strategy instead is to control statistically for that aid in both the first and second stages of the analysis. In the first stage, these controls mean that the variation in state-funded positions that provides the exogenous variation for the second stage is independent of any variation in that aid. That is, the “left-over” variation in state-funded positions ought to come primarily from the “base” allocations, which are solely a function of enrollment and allocation formulas.

17. Weighting by total enrollment at the district level produces nearly identical results given the very high correlation between the total number of test-takers and total enrollment in a district.

18. In addition, the results in Supplementary Table S1 in the online version of the journal illustrate that the instrument with the tightest conceptual kinship with the particular endogenous staffing category of interest carries the bulk of the predictive power for that staffing category. For example, state allotments of teacher slots predict district-level numbers of teachers much better than district-level numbers of health care staff, as they should.

19. Across the 36 estimates in Table 2, we would expect to observe, on average, about two statistically significant coefficients at the 5% level by chance, or about four at the 10% level. In fact, we find only one marginally significant coefficient. Recall that in our main models, we also control for a number of other covariates intended to capture local economic shocks as well as other pots of general purpose state aid that districts might use to support staffing. Thus, Table 2 further bolsters the contention that the vast majority of state-funded positions captured by our instrument are a function of the “base allocation” procedure that solely relies on enrollment and makes no attempt to account for differences in student or district characteristics.

20. 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 district (or 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.

21. In addition, recall that in Table 2 we failed to detect a statistically significant relationship between lagged measures of reading and math performance and state-funded staffing positions, assuaging such concerns.

22. Average test scores for the full sample are not precisely zero as they would be for all students in those grades throughout the state because the sample excludes tested students in schools that offer nonstandard sets of grades.

23. This measure is based on eligibility for free or reduced-price school meals.

24. Turnover is defined at the school level and is calculated at each school in each year as the percent of teachers in the school in year $t-1$ who were not there in year $t$.

25. We group schools into these categories using the unweighted average proportion of economically disadvantaged students in a district over the full time period of our panel. High-poverty districts fall in the top quarter of the distribution of those average shares; low-poverty districts fall in the bottom quarter; and the remainder fall within the middle 50% of that distribution.

26. First-stage estimates for the two-stage least squares (2SLS) models appear in Supplementary Table S1 in the online version of the journal. As reported in that table, high $F$ statistics indicate that the instruments generate appreciable independent variation in district staffing levels of various types.

27. The fact that the 2SLS estimates for TAs are somewhat smaller than the RF estimates reflects the fact that the coefficient on state-funded TAs on total TAs in the first-stage regressions is slightly greater than 1 (see Supplementary Table S1 in the online version of the journal).

28. We also estimated some models in which we split the “other” category into its components of school counselors, school leaders, and noncertified academic support staff—but have very limited variation in these
subcategories over time, leaving us with little ability to produce reasonable estimates.

29. For example, higher proportions of Black students, or students with special needs, predict lower average test scores, and higher proportions of girls predict higher test scores in reading. In terms of teacher characteristics, as proxies for teacher quality, a higher proportion of teachers with more than 3 years of experience predicts higher test scores. The negative coefficients for teacher Praxis scores in the math equations, however, were unexpected.

30. Leaving either or both of these staffing variables out of the model that includes time trends has essentially no effect on the estimated effects for TAs.

31. Ideally, we would have liked to repeat the analysis for subgroups divided by their family income level as proxied by a student’s classification as economically disadvantaged, but we are missing the necessary data for many students. Instead, we use district-level data to divide the sample of districts into high-poverty (those in the top 25%), middle-poverty (the middle 50%), and low-poverty (those in the bottom 25%) districts based on each district’s percentage of students eligible for free or reduced-price meals averaged over the whole period.

32. Although all of our models include district and year fixed effects, Ho (2008) underscored that schools face varying incentives to focus on students near the proficiency threshold and showed that the density of students near the proficiency cutoff can complicate the interpretation of results based on proficiency measures, especially measures that characterize gaps in performance. Thus, we urge caution in interpreting the relative magnitudes of effects across subgroups.

33. For comparison purposes, here are other monthly teacher salaries: BA and 5 years of experience, US$4,000; BA and 8 years of experience, US$4,300; and MA with 8 years of experience, US$4,730.

References

Alborz, A., Pearson, D., Farrell, P., & Howes, A. (2009a). The impact of adult 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] (DCSF-EPPI—09-01). University of London.

Alborz, A., Pearson, D., Farrell, P., & Howes, A. (2009b). The impact of adults support staff on pupils and mainstream schools [Technical Report: Report no. 1702T]. In Research evidence in education library. EPPI-Center, Social Science Research Unite, Institute of Education, University of London.

http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=w423SGddbQ%3d&tabid=2438&mid=4540

Andersen, S. C., Beuchert, L., Nielsen, H. S., & Thomsen, M. K. (2018). The effect of teacher’s aides in the classroom: Evidence from a randomized trial. Journal of the European Economic Association, 18, 469–505.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides’ rule to estimate the effect of class size on scholastic achievement. Quarterly Journal of Economics, 114(2), 533–575.

Angrist, J. D., Lavy, V., Leder-Luis, J., & Shany, A. (2019). Maimonides’ rule redux. American Economic Review: Insights, 1(3), 309–324.

Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics. Princeton University Press.

Bressoux, P., Kramarz, F., & Prost, C. (2009). Teachers’ training, class size and students’ outcomes: Learning from administrative forecasting mistakes. Economic Journal, 119, 540–561.

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.

Cho, H., Glewwe, P., & Whitley, M. (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. (2006). Teacher-student matching and the assessment of teacher effectiveness. Journal of Human Resources, 41(4), 778–820.

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.

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. Economics of Education Review, 31, 92–109.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. Journal of Economic Literature, 47(1), 87–122.

Department of Public Instruction, North Carolina. (2018). 2018-2019 allotment policy manual. https://files.nc.gov/dpi/documents/fbs/allotments/general/2018-19policymanual.pdf

Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? Canadian Journal of Economics, 47(2), 634–663.
Ding, W., Lehrer, S. F., Rosenquist, J. S., & Audran-McGovern, J. (2009). The impact of poor health on academic performance. New evidence using genetic markers. *Journal of Health Economics, 28*(3), 578–597.

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), 692–717.

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* (DEP Discussion Paper No. 1226). Center for Economic Performance.

Guryan, J. (2001). Does money matter? Regression discontinuity estimates from education finance reform in Massachusetts (NBER Working Paper No. 8269). National Bureau of Economic Research.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141–1177.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2), 141–164.

Ho, A. (2008). The problem with proficiency: Limitations of statistics and policy under no child left behind. *Educational Researcher, 37*(6), 351–360.

Hoxby, C. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics, 115*(4), 1239–1285.

Hyman, J. (2017). Does money matter in the long run? Effect of school spending on educational attainment. *American Economic Journal: Economic Policy, 9*(4), 256–280.

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). http://www.nber.org/papers/w18624

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

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, 114*(2), 497–532.

Ladd, H. F., & Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management, 29*(3), 426–450.

Lauen, D. L., & Gaddis, S. M. (2016). Accountability pressure, academic standards, and educational triage. *Educational Evaluation and Policy Analysis, 38*(1), 127–147.

Lazear, E. P. (2001). Educational production. *Quarterly Journal of Economics, 116*(3), 777–803.

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? (Working Paper No. 22030). National Bureau of Economic Research.

Nagler, M., Piopiunik, M., & West, M. R. (2020). Weak markets, strong teachers: Recession at career start and teacher effectiveness. *Journal of Labor Economics, 38*(2), 453–500.

Nordstrom, K. (2017). Financing education in North Carolina: A budget and tax guide. North Carolina Justice Center. https://www.ncjustice.org/publications/financing-education-in-north-carolina/

North Carolina School Social Workers’ Association. (n.d). https://ncsswa.com/

Papke, L. E. (2005). The effects of spending on test pass rates: Evidence from Michigan. *Journal of Public Economics, 8*, 821–839.

Piketty, T., & Valdenaire, M. (2006). L’impact de la taille des classes sur la reussite scolaire dans les ecoles, colleges et lycées francais: Estimations à partir du panel primaire 1997 et du panel secondaire 1995 [The impact of class size on student success in French elementary, middle, and high schools—Estimates based on the 1997 primary and 1995 secondary panels] (Les Dossiers n173). Ministère de l’Education Nationale.

Public School Forum of North Carolina. (2012). Local school finance study. https://www.ncforum.org/wp-content/uploads/2016/10/2012FinanceStudy.pdf

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. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. Econometrica, 73(2), 417–458.
Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. American Educational Research Journal, 50(1), 4–36.
Sorensen, L., & Ladd, H. (2020). The hidden costs of teacher turnover. AERA Open, 6(1). https://doi.org/10.1177/2332858420905812
U.S. Department of Labor. (2014, May). Occupational employment statistics: OES data. https://www.bls.gov/oes/tables.htm

Authors
STEVEN W. HEMELT is an associate professor in the Department of Public Policy at the University of North Carolina at Chapel Hill, and a research fellow at the Institute of Labor Economics (IZA). His research intertwines the economics of education, education policy, labor economics, and program evaluation.
HELEN F. LADD is the Susan B. King Professor Emerita of Public Policy and Economics, Sanford School of Public Policy at Duke University. She has studied a range of education policy issues, including school choice and school finance, in the United States and other countries.
CALEN R. CLIFTON is a fellow with the Strategic Data Project at the Harvard Graduate School of Education and a senior data analyst with ImpactTulsa. His research interests include teacher quality, school choice, and segregation.

Manuscript received February 15, 2019
First revision received February 25, 2020
Second revision received October 28, 2020
Accepted January 4, 2021