NATIONAL
CENTER for ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

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

Right-Sizing the Classroom: Making the Most of Great Teachers

Michael Hansen

WORKING PAPER 110 • JANUARY 2014

Right-Sizing the Classroom: Making the Most of Great Teachers

Mike Hansen
American Institutes for Research

Contents

Acknowledgements

This project was supported by a research grant from the Thomas B. Fordham Institute. I acknowledge the excellent research assistance of Tiffany Chu, and helpful guidance and feedback from Amber Winkler and Checker Finn. The analysis represented here does not necessarily reflect the opinions of any affiliated institutions, and any and all errors are mine. Author contact: 1000 Thomas Jefferson St. NW, Washington, DC 20007; e-mail: mhansen@air.org

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

Right-sizing the Classroom: Making the Most of Great Teachers
Michael Hansen
CALDER Working Paper No. 110
February 2014

Abstract

This paper examines the value of strategically assigning disproportionately larger classes to the strongest teachers in order to optimize student learning in the face of differential teacher effectiveness. The rationale is straightforward: Larger classes for the best teachers benefit the pupils who are reassigned to them; they also help the less effective teachers improve their instruction by enabling them to concentrate on fewer students. But just how much of a difference could manipulating class sizes in this way make for overall student learning and access to effective teaching? This study performs a simulation based on North Carolina data to estimate plausible student outcomes under this approach.

In the North Carolina data, I find there is currently a very slight tendency to place more students in the classes of effective teachers; but still only about 25 percent of students are taught by the top 25 percent of teachers. Intensively reallocating eighth-grade students—so that the most effective teachers have up to twelve more pupils than the average classroom—may produce gains equivalent to adding roughly two-and-a-half extra weeks of school. Even adding a handful of students to the most effective eighth-grade teachers (up to six more than the school's average) produces gains in math and science akin to extending the school year by nearly two weeks or, equivalently, to removing the lowest 5 percent of teachers from the classroom. The potential impacts on learning are more modest in fifth grade, where the large majority of teachers are in self-contained classrooms.

Results show that this strategy shows an overall improvement in student access to effective teaching, yet gaps in access for economically disadvantaged students persist. For instance, disadvantaged eighth-grade students are about 8 percent less likely than non-disadvantaged peers to be assigned to a teacher in the top 25 percent of performance. This gap in access

changes little in spite of the policy putting more students in front of effective teachers—because the pool of available teachers in high-poverty schools does not change under this strategy. Thus, this policy alone shows little promise in reducing achievement gaps.

Introduction

Being assigned to the classroom of a great teacher can make significant differences in a myriad of outcomes for an individual student, including greater academic achievement, improved odds of going to college, and higher future wages.[1] Yet, if teachers are so important, it is paradoxical why schools continue to assign the top 25 percent of instructors only about 25 percent of students.[2] The very presence of large differences in productivity across the teacher workforce implies an arbitrage opportunity—that is, overall improvements could be realized through simple reallocations of students across teachers. This paper investigates how student learning might change if more students were disproportionately allocated to top teachers based on prior performance.

This idea of reallocating students according to teachers' effectiveness is not new. Student-teacher ratios and class sizes in the United States had been on steady declines extending for nearly four decades (Ahn and Brewer, 2009), but the Great Recession has recently forced states and school districts to decide how to reduce costs and mitigate the harm to student learning, leading some to reverse on their class size reduction policies.[3] In response to these intense budget pressures, high-profile personas in education policy, including U.S. Secretary of Education Arne Duncan and Bill Gates, have promoted strategic classroom assignments in which larger classes are assigned to the most effective teachers.[4]

[1] Chetty, et al. (2011) provide evidence that teacher productivity, measured by estimated value added on standardized test scores, has a statistically significant relationship with future student outcomes such as college attendance, wages, and teenage pregnancy.

[2] These numbers are based on the observed data in North Carolina that are used for this study. See Table 3 for statistics on current patterns of assignment.

[3] California, for instance, has reversed its class-size reduction program that was originally adopted in 1996. Half of the state's 30 largest districts had student-teacher ratios of 30 or more in 2011-12, where virtually all districts in the state had ratios of 20 or less in 2008-09 (Freedberg, 2012).

[4] Secretary Duncan includes this suggestion in a list of possible ways schools could constructively adapt to the 'new normal' in education under budgetary pressures (Duncan, 2010). Bill Gates suggests this in an op-ed published in the *Washington Post* (Gates, 2011).

Similarly, Public Impact, an education research group, launched the Opportunity Culture initiative in 2011, which aims to increase students' exposure to the top 25 percent of teachers.[5]

Beyond financial pressures inducing changes on class size policies, copious research on teacher quality over the last decade has overshadowed the evidence on class size from a prior era. Class-size research generally presupposed that all teachers were equally effective. Thus, the key policy lever for increasing student achievement was reducing the number of students assigned to any one teacher. Yet, the differences in teacher productivity across classrooms are large and important predictors of student learning and other outcomes, and the magnitude of these differences *swamps* the expected effect of smaller classes,[6] suggesting that a policy focus on teacher quality rather than class size will result in the largest improvements in student performance in the nation's public schools. Still, rather than framing the solution as smaller class sizes versus better teachers, this paper combines the two approaches. It posits that student outcomes can improve by determining class sizes strategically for teachers based on their expected classroom performance.

But just how much of a difference could manipulating class sizes make on overall student learning and access to effective teaching? No district to my knowledge has purposefully allocated students in this manner.[7] In the absence of real-world implementation, this paper uses North Carolina data from fifth and eighth grade classrooms to simulate student outcomes under the assumption that

[5] This initiative provides 10 different models (plus variations and combinations of these models) on how schools can organize around their most effective teachers. Models include subject or role specialization, multi-classroom leadership, and time-technology swaps, among others. The strategy I analyze here, where principals strategically shift class sizes to give more students to the most effective teachers and fewer students to the least effective, is akin to Public Impact's Class-Size Shifting (in-person) model. However, the available documents on this model do not provide the optimality conditions for determining class size in the face of variable teacher productivity, as I do here. For more information on this and the other models in Public Impact's Opportunity Culture initiative, see http://www.opportunityculture.org.

[6] Nye, et al. (2004) estimate a 10-20 student difference in class size would be required to compensate for the learning difference between an average and highly effective teacher. See a concise review of the research on the teacher value-added literature, including a comparison to class-size reduction policies in Hanushek and Rivkin (2010).

[7] To my knowledge, some districts have tried experimenting with this or similar approaches to increase student learning through manipulating class sizes, but no large-scale, systematic implementation has been done to my knowledge.

schools strategically assign larger classes to the strongest teachers and smaller classes to the weakest. Specifically, I ask: First, to what extent do observed class size assignments already represent differences in teacher performance? And second, to what extent does a purposeful policy of allocating students to teachers on the basis of teacher effectiveness boost achievement (as well as students' access to quality teaching)? The cost of this policy is also an important consideration and will be discussed at the end.

In summary of the key findings here, I find relatively small changes in the way class sizes are assigned to teachers can potentially make educationally significant differences in student learning, though these gains appear to be sensitive to the grade level. These simulated gains are positive, but relatively small in grade 5, and could produce a net negative return if teachers required compensation to comply. Conversely, the results in grade 8 suggest large gains associated with the strategic class size policy, equivalent to roughly 2.5 additional weeks of instruction in math and science when effective teachers can teach up to 12 students above the average number. Even a less intensive level of sorting, where effective teachers only teach up to six students above the average, results in gains of nearly 2 weeks in math and science. The student learning gains here imply positive returns even when generously compensating teachers for accepting the additional student load.

Motivation

Prior Research

The variation in teacher effectiveness, which occurs both within and across schools, is large and accounts for more of the differences in student learning than any other known education intervention (e.g., Aaronson, et al., 2007; Hanushek and Rivkin, 2010). Studies have also shown that traditional characteristics of teacher preparation, such as licensure and education, are poor predictors of their classroom performance and are thus ineffective strategies in boosting workforce quality (Goldhaber and Hansen, 2013; Staiger and Rockoff, 2010). State and district policies have also begun to shift in response to these findings. For example, 36 states have revamped their teacher and principal evaluation systems since 2009, and most of these states have adopted value-added estimates, which are teacher-productivity measures based on student gains on standardized tests, as a key component of these evaluation systems (National Council on Teacher Quality, 2012).

The new, widespread availability of teachers' performance data compels one to consider how this data may be most effectively used to improve student achievement. Proposals to directly use performance data—including pay-for-performance bonuses, or selectively retaining teachers for tenure—are abundant, but are also unpopular among teachers.[8] However, using performance data *indirectly* for the same purpose may be more politically palatable. Strategically assigning larger class sizes to high-performing teachers is one such mechanism that indirectly uses performance data to improve student outcomes.

This policy mechanism counters the common policy push towards universally smaller class sizes, which are widely popular with both parents and teachers.[9] There is rigorous evidence that smaller classes boost student learning, though these findings are more narrow than commonly believed—the largest credible class-size effects appear to be due to a student's first exposure to a small class, and have only been demonstrated in the earliest grades.[10] For most students above the third grade, the evidence

---

[8] See Scholastic and the Bill and Melinda Gates Foundation, 2012.

[9] Class-size policies are well-liked by parents and teachers and are commonly believed to have positive effects on student outcomes; policies on limiting class size have been adopted in 36 states, contributing to the steady decline of student-teacher ratios in public schools nationwide over the last few decades of the twentieth century. See Ahn and Brewer (2009) for further discussion.

[10] In Whitehurst and Chingos's (2011) recent review of the evidence on class size, the largest credible estimated effect they report is that from Krueger's (1999) analysis of the Tennessee STAR experiment, conducted in grades K-3. Students randomly assigned to small class sizes, averaging 7 to 8 fewer students than their comparison classes, increased their test scores in grade 4 by 0.22 standard deviations, on average, across subjects of the Stanford Achievement Test. According to Krueger, most of the achievement gains were realized as a one-time gain at a student's initial exposure to the small class (which was in kindergarten for most study participants); the effects of successive exposures to small classes, though still positive, were roughly a quarter of the initial year's magnitude. Thus, this estimate could be considered an upper bound class-size effect, and its magnitude has not been duplicated in other rigorous studies.

points to at most a small class-size effect, if any at all.[11] (Using the North Carolina data, I likewise estimate small class-size effects in fifth and eighth grades.) Thus in effect, it would take an increase of at least 10-20 additional students in a good teacher's class to dilute his or her productivity to that of an average teacher.[12] Put another way, assigning a few extra students to the class of an effective teacher can translate to significant learning gains for these students, while only making very small reductions in that teacher's performance for everyone else in the class. Beyond that, universal class-size reduction as a policy is exceptionally costly and has the unintended consequence of lowering workforce quality in its implicit demand to increase the quantity of teachers.[13] In summary, there's little reason to prioritize universally smaller class sizes when there are smarter ways to assign classes to teachers who vary in their classroom effectiveness.

Assigning Class Sizes to Optimize Learning

The class-size shifting strategy simulated here considers the assignment of students to classes as a learning optimization problem, asking how would class sizes be assigned to different teachers if total student achievement were the objective? Consider the principal's assignment problem under variable teacher effectiveness, where his or her objective is to maximize total student learning (L) by distributing students across teachers with different levels of expected performance (τ_i). The main choice variable is

[11] Whitehurst and Chingos (2011) identify four studies showing credible, statistically significant, positive effects associated with smaller class sizes; another three studies that present mixed results (where some estimates of the class size reduction are significantly positive in some cases or for some student groups, but the overall class size effect was not statistically significant); and another two that found no significant class size effect. The authors interpret the overall evidence as suggesting a linear class-size effect "diminish[ing] with each grade in school, with a reduction of a given number of students in 5th grade expected to have about half the effect of the same number of students in kindergarten" (p. 10).

[12] Within-school standard deviations (SD) of teacher productivity range from 0.08 to 0.26 SD of student achievement in reading and 0.11 to 0.36 SD in math (Hanushek and Rivkin, 2010). In their analysis of teacher effectiveness using data from the Tennessee STAR class size experiment, Nye, et al. (2004) estimate a 10-20 student difference in class size would be required to compensate for the learning difference between an average and highly effective teacher.

[13] Jepsen and Rivkin (2009) document how the increase in demand for teachers inadvertently undermined the qualifications of the teacher workforce in California, when class-size reduction policies were adopted in the late 1990s.

the number of students assigned to each class (n_i), given teachers whose effectiveness in the classroom varies across individual teachers and is a function of class size ($\tau_i(n_i)$).

(1) $$\max L = \tau_j(n_j) \cdot n_j + \tau_k(n_k) \cdot n_k \text{ by choosing } n_j$$

$$\text{subject to: } n_j + n_k = N$$

In this two-classroom model, N students are distributed between classrooms *j* and *k*. Substituting the constraint into the model and maximizing this equation with respect to n_j results in:

(2) $$\frac{\partial L}{\partial n_j} = \frac{\partial \tau_j}{\partial n_j} n_j + \tau_j(n_j) + \frac{\partial \tau_k}{\partial n_j}(N - n_j) - \tau_k(N - n_j) = 0$$

Rearranging terms, using the identity $\frac{\partial \tau_k}{\partial n_j} = -\frac{\partial \tau_k}{\partial n_k}$ as implied by the constraint, and using asterisks to denote optimal choices gives the necessary first-order optimality condition[14]:

(3) $$\frac{\partial \tau_j}{\partial n_j} n_j^* + \tau_j(n_j^*) = \frac{\partial \tau_k}{\partial n_k} n_k^* + \tau_k(n_k^*)$$

For ease of interpretation, however, one can rearrange the terms once again to bring the teacher production functions to the left-hand side and the marginal product terms to the right-hand side.

(4) $$\tau_j(n_j^*) - \tau_k(n_k^*) = \frac{\partial \tau_k}{\partial n_k} n_k^* - \frac{\partial \tau_j}{\partial n_j} n_j^*$$

In short, an optimizing principal will allocate students across classrooms to the point where individual gains to a marginal student switching from the weak teacher's classroom to the strong teacher's classroom (on the left-hand side) are offset by the net collective changes in learning among the students already assigned to classrooms (on the right-hand side). If teacher effectiveness did not vary across classrooms such that $\tau_j(n_i) = \tau_k(n_i)$ for all n_i, then the optimizing principal would balance

[14] Barrett and Toma (2013) analyze a similar strategic allocation of students across teachers according to teacher effectiveness, but their model framework and optimality conditions differ from those presented here. Rather, Barrett and Toma assume strong and weak teachers differ in their marginal productivity at equal class sizes and that school principals optimize learning by reallocating students from weak to strong teachers to the point where teachers' marginal products per student are equal across classrooms (the convexities of teachers' production functions are assumed). The analysis presented in this paper varies in two key ways: first, it directly considers the differences in teacher productivity across classrooms (where Barrett and Toma only consider differences in the convexities of the production functions); and second, it considers the collective gains or losses incurred among all students in each class resulting from the addition of a marginal student (where they consider only marginal product per marginal student).

the collective losses in one class against the gains in the other by equalizing class sizes across teachers. Conversely, if there were no change in teacher effectiveness associated with increasing class size (i.e., $\frac{\partial \tau_k}{\partial n_k} = \frac{\partial \tau_j}{\partial n_j} = 0$), the optimizing principal would put all students in the classroom of the teacher with the greatest effectiveness, even if the difference in effectiveness was very small.

In practical terms, in the typical case where teacher effectiveness varies across classrooms and the class-size effect is small in magnitude (as is true based on the NC data used here), this approach reassigns students from the weakest to the most effective teachers in a given school teaching the same course, starting from the basis of equally-sized classrooms. As an example, consider two 8th grade math teachers' classrooms with 24 students initially assigned to each, and based on prior performance, the principal expects teacher A to outperform teacher B by 0.1333 standard deviations of student performance, and the expected effect of student learning due to the change in class size is expected to be a loss of 0.0035 standard deviations for each additional student assigned to the class.[15] Moving one student from teacher B's class into teacher A's class does three things concurrently. First, the "shifted" student gains from being reassigned to a better teacher, equivalent to the differential in teacher effectiveness between the two (0.1333 s.d.). Second, moving the student increases the class size of teacher A's classroom, resulting in a "penalty" imposed on other pupils already assigned to teacher A. The magnitude of this penalty is the product of the new class size and the class size effect (25 x -0.0035 = -0.0875 s.d.). Third, teacher B's class shrinks by one student, benefiting all students who remain with teacher B. The magnitude of this benefit is the product of the new class size and the class-size effect, which is positive in this case due to a reduction in size (23 x 0.0035 = 0.0805 s.d). The net effect of shifting this first student is the sum of these three pieces, in this example a net gain of 0.1263 s.d.; or 0.0026 s.d. when averaged over all 48 students in the grade. Depending on how many students are

[15] These values are based on the estimates generated in 8th grade math using the NC data in this analysis, see Table 2. Teacher A is one standard deviation of teacher effectiveness greater than teacher B, or an 85th percentile teacher compared to a teacher at the median.

allowed to shift between classes (and the difference in effectiveness between the teachers), the resulting class sizes become more imbalanced, where the most effective teachers lead the largest classes and the weakest the smallest.[16]

In theory at least, this disproportionate alignment of students should translate to greater learning gains overall as more students receive high-quality instruction, and the weaker teachers receive small classes that are more conducive to increasing their performance as well. How well this approach may work in practice is analyzed in the following simulation.

Data

This analysis utilizes longitudinal data from North Carolina administrative files for the 2007-08 through 2010-11 school years. The standardized test scores are those on the criterion-referenced End-of-Grade (EOG) tests administered as part of the state's ABCs accountability program. Math and reading tests are given annually to all students in grades 3 through 8. In addition, North Carolina has also administered an EOG science test to students in grades 5 and 8 only since the 2007-08 school year.[17]

The administrative data also document typical variables on student background including student demographics, which will be used as student-level explanatory variables in predicting student achievement. Next, course membership files in the data are used to identify the classes in which students receive instruction and the teachers to whom they are assigned; the variable on class size

[16] As more students are allowed to transfer from teacher B's class to teacher A's class, the net marginal gain for each transferred student declines as both the expected performance differential between the teachers shrinks and the class-size penalty in teacher A's class increasingly outweighs the benefit in teacher B's class. For instance, extending the example in the text using the estimates from 8th grade math teachers, the marginal gain for transferring the 6th student between classes (which would result in class sizes of 30 and 18 for teacher A and B, respectively) is 0.0563 s.d. overall, or 0.0012 s.d. per student. In this example, it would be optimizing to continue transferring students from teacher B to teacher A until overall learning is maximized when 10 students have been transferred, which would result in class sizes of 34 and 14 for teacher A and B, respectively.

[17] For estimation in the analysis, student test scores are standardized (i.e., converted to z-scores) based on published statewide means and standard deviations in the tests' yearly technical reports.

comes from these files.[18] Finally, teacher-level payroll files are used to determine teacher experience, a key variable predicting a teacher's classroom performance.[19] Six unique data samples were analyzed in this study, corresponding to the unique subject-grade combinations in the data for grades 5 and 8 (two grades across three subjects).

A necessary condition for a school to even consider adopting this class-size shifting strategy is the availability of at least two distinct teachers at the same school assigned to the same grade (or grade and subject in the case of secondary grades) so that the instructional content is interchangeable across instructors. Further, it's helpful to see the number of teachers available to schools to understand how much flexibility these schools may have in reallocating students across teachers. As shown in Table 1, the majority of schools serving 5th or 8th grades in North Carolina satisfy this minimum necessary condition of having at least two teachers.[20] Note, though, these numbers represent schools—not classrooms—and schools with fewer teacher options tend to be smaller schools with fewer classrooms. Across both grades, 89 percent or more of the unique classrooms occur in schools with at least two teachers.

Methods

To examine the potential gains from the class size assignment policy described here, I conduct a simulation wherein class sizes are strategically determined to optimize total student learning in the

[18]Students who are linked with multiple teachers in their course membership files (because of either switching classes mid-year, or taking multiple classes in the same subject) are linked to only one teacher in estimating value added, and that determination is made by 1) attributing the student to the teacher in the tested school (for students in multiple schools), and 2) then attributing to the one responsible for the largest portion of the school year (for students with multiple teachers in the tested school), and in the few cases where students are still linked to multiple teachers, 3) attributing the student to the teacher with the fewest other linked students. By linking students to only one teacher in this way, I avoid the need to use a more computationally intensive dosage-weighted model in estimating teacher effects.

[19] Teachers with missing experience values are recoded as 0 and flagged with a missing indicator variable.

[20] To be included in this table, a school must have at least one class (minimum of 5 students) on record in the given subject / subject-grade. In addition, 5th grade classes must be self-contained (or block classes teaching at least 2 of the 3 subjects) for inclusion; 90% of all 5th grade student observations are in such a classroom in this year. These counts are based on the course membership files in North Carolina, and it is possible records may be missing or incomplete due to data errors in some schools; hence, the varying number of total schools by subject in 8th grade.

presence of variable teacher quality based on the NC data in the 2010-11 school year. This simulation has three major components, each of which are described below.

Estimating key simulation parameters using first three years of data

I begin by estimating a teacher fixed effects model based on three prior school years (2007-08 through 2009-10). All teachers' classrooms in the given grade and subject with at least 5 valid student observations in one school year are included in the analysis.[21] I estimated using the following equation:

(5) $$y_{i,c,j,t} = \mathbf{y_{i,t\text{-}1} I_t \beta_1} + \boldsymbol{X}_{\mathbf{i,t}}\boldsymbol{\beta}_{\mathbf{2}} + Size_{c,j,t}\gamma + \mathbf{Class_{c,j,t}}\boldsymbol{\delta} + \mathbf{Exp_{j,t}}\boldsymbol{\zeta} + \mathbf{I_j}\boldsymbol{\tau} + \varepsilon_{\mathrm{i,c,j,t}}$$

In this equation, current student learning ($y_{i,c,j,t}$) in the given grade and subject is the dependent variable, where the subscripts represent an individual student *i*, in class *c*, assigned to teacher *j*, at time *t*.[22] The value of current student learning is predicted using the following explanatory variables: a vector of students' prior-year test scores in both reading and math, which are interacted with yearly indicator variables ($\mathbf{y_{i,t\text{-}1} I_t}$)[23]; a vector of student characteristics ($\boldsymbol{X}_{\mathbf{i,t}}$), which include indicators for gender, racial categories, eligibility for free or reduced-price lunch (FRL), special education status, and limited English proficiency; a linear term on class size ($\mathrm{Size}_{c,j,t}$); a vector of variables capturing classroom composition effects resulting from the mix of students in the classroom ($\mathbf{Class_{c,j,t}}$), including mean prior test score (in the same subject as the dependent variable), the percentage of students in the class that are FRL eligible, and the count of students in the class with prior test scores below the 20th

[21] Note that I use all classrooms at this point to generate the estimates of interest, with the exception of dropping classes with 5 or fewer students to avoid the inclusion of specialty classes. I investigated the possibility of using a higher threshold (12 students), and found little difference in the sample of students/classrooms included in the sample, and chose to use the more inclusive sample. I will be limiting the sample in the simulation to those classrooms that have two or more available teachers in the same school.

[22] For estimation in the analysis, student test scores are standardized (i.e., converted to z-scores) based on published statewide means and standard deviations in the tests' yearly technical reports; hence, the value-added estimates should be interpreted to mean relative (not absolute) gains in student achievement. This model varied slightly when estimating this equation with science test scores as the dependent variable. Because science is not tested in every year, students' prior-year math and reading scores were used as the only predictors, and consequently the teacher fixed effects estimates should be considered "quasi-value-added."

[23] Students with missing values in their same-subject prior test score are dropped from the analysis sample. Those with missing opposite-subject prior test scores (e.g., missing prior reading scores when the outcome of interest is math scores) but have valid same-subject prior test scores are imputed based on other observed variables in the data.

percentile; a vector of categorical variables representing a teacher's experience ($\mathbf{Exp}_{j,t}$)[24]; and a vector of indicator variables representing student-teacher links ($\mathbf{I}_j$) used to estimate the teacher fixed effects.

The coefficients generated in this model that will be used later in conducting the simulation are those representing the estimated class size effect ($\hat{\gamma}$), within-teacher differences in a classroom's peer effects ($\widehat{\boldsymbol{\delta}}$), within-teacher average returns to experience ($\widehat{\boldsymbol{\zeta}}$), and the teacher value-added estimates spanning the three years of data ($\hat{\boldsymbol{\tau}}$). The class-size estimates and the variation of teacher value-added estimates are presented in Table 2. As expected, the estimates on class size show relatively small negative coefficients associated with increases in class size (ranging from -0.0052 to 0.0), while the standard deviation of teacher effects are many multiples larger (ranging from 0.06 to 0.19 standard deviations). I do not claim any of these estimates as causal, though they are within the range of other credible estimates of class size (Chingos, 2013) and teacher quality (Hanushek and Rivkin, 2010), and I use them as parameters of interest for the simulation as though the relationships were causal.

Imprecise value-added estimates (i.e., small samples associated with a given teacher) are shrunken towards the mean of the teacher distribution (within grade and subject) via the empirical Bayes adjustment. For teachers observed in multiple grades over the three-year period, separate value-added estimates for a teacher are averaged across grades to result in a single prior value-added estimate per teacher (per subject). In elementary grades, where teachers in self-contained classrooms teach all subjects to their students, expected teacher performance is not determined separately by subject, but jointly. So in these cases, I simply average all prior value-added estimates across subjects.

Evaluating current classroom assignments based on expected performance

As a starting point of comparison, it is helpful to evaluate whether and how North Carolina class sizes are allocated through the lens of varying teacher effectiveness; after all, there is little to be gained from this approach if it's already part of the status quo. I assume a principal forms his or her expectation

[24] The omitted category is 0 years of experience, and separate indicator variables are established at the following intervals: 1-2 years of experience, 3-4 years, 5-9 years, 10-14 years, 15-19 years, and 20 or more years.

of a teacher's future teacher productivity based on the combined value of the teacher's experience and past value-added performance. The experience predictor is simply the corresponding value of the experience coefficients ($\hat{\boldsymbol{\zeta}}$), as estimated in Equation (5), that is associated with the teacher's level of experience for the coming school year. Because teachers' empirical Bayes adjusted value-added estimates ($\hat{\boldsymbol{\tau}}_{\boldsymbol{j}}$) based on the three prior years are imperfect predictors of future performance, the principal does not give these prior estimates full weight, but scales them according to their predictive coefficient on future performance (notated as $\hat{\lambda}$).[25] Thus, the principal's expectation of a teacher's performance is computed as:

$$(6) \qquad \hat{\theta}_{j,t+1} = \mathbf{Exp}_{\mathbf{j,t+1}} \hat{\boldsymbol{\zeta}} + \hat{\lambda} \mathbf{I_j} \hat{\boldsymbol{\tau}}_{\boldsymbol{j}}$$

To evaluate the degree of sorting that naturally occurs in the data, I will look at assignment patterns of class sizes relative to this measure of expected performance. Several metrics will be computed to examine this issue:

1. The within-school correlation of class size and expected teacher performance, which takes a simple average across all schools of the school-level correlation coefficient between class size and expected teacher performance.
2. The proportion of leftover students—I define 'leftover students' as additional students in teachers' classes above the average class size in the school-grade; this metric simply reports whether these students are assigned to teachers in either the top or bottom 25% of the distribution of expected performance.
3. The proportion of students assigned to teachers in the top 25% of the distribution of expected performance, overall and also among FRL-eligible students only.

[25] I obtain these scalar weights by using a version of Equation (5) that produces 2-year value-added estimates based on the 2007-08 and 2008-09 school years (instead of producing 3-year estimates). These 2-years estimates are then used as regressors in a value-added regression predicting student learning in the 2009-10 school year, and the estimated coefficients on the prior value-added estimates are retained as the scalar weights. This approach provides out-of-sample weights to predict future teacher performance with prior multi-year value-added estimates. For elementary grades, I use the average value of this scalar across subjects to create the expected performance measures.

Simulating classroom assignments and gains

All of the parameters generated above (in Equations 5 and 6) are necessary to simulate how principals could strategically reassign their students. However, in order to compute the learning gains that may result from this approach, I still need to understand how teachers actually perform in the 2010-11 target school year, as this will vary from the value of their expected performance. To do this, I calculate each teacher's actual value-added in the 2010-11 school year, net of any class size, classroom composition, and teacher experience effects. This is estimated in a two-stage process. The first stage adjusts students' test scores in the 2010-11 school year based on the classroom and teacher characteristics to which they were exposed (using the corresponding coefficient estimates from Equation 5):

$$(7) \qquad y^{*}_{i,c,j,2011} = y_{i,c,j,2011} - Size_{c,j,2011}\hat{\gamma} - \mathbf{Class_{c,j,2011}}\hat{\boldsymbol{\delta}} - \mathbf{Exp_{j,2011}}\hat{\boldsymbol{\zeta}}$$

This is a necessary adjustment, as these conditions are constant across all students within a class and are therefore collinear with a single-year teacher fixed effect for teachers observed in just one class.[26] The second stage then uses this adjusted student learning as the dependent variable in a value-added regression that predicts single-year teacher value-added in a fixed effects model:

$$(8) \qquad y^{*}_{i,c,j,2011} = \mathbf{y_{i,2010}}\boldsymbol{\beta_1} + \boldsymbol{X}_{\mathbf{i,2011}}\boldsymbol{\beta_2} + \mathbf{I_j}\boldsymbol{\tau} + \varepsilon_{\mathrm{i,j,2011}}$$

In this model, the prior test score ($\mathbf{y_{i,2010}}$), student characteristics ($\boldsymbol{X}_{\mathbf{i,2011}}$), and teacher assignment ($\mathbf{I_j}$) vectors are equivalent to those described in Equation (5). The resulting teacher effect estimates ($\hat{\boldsymbol{\tau}}_{\boldsymbol{j},\mathbf{2011}}$) from this regression are directly used in the simulation below to compute student learning under alternate classroom assignments.

Now with all of these various elements calculated, I can now move to strategically manipulate class sizes across teachers in relation to the expected difference in teacher productivity across

[26] Some, but not all, teachers in the samples instruct multiple classes in a single year, and therefore the class-size, composition, and experience effects would not be collinear for them. I use this adjustment approach uniformly, regardless of the number of classes they taught in 2010-11.

classrooms. Teachers' school-, grade-, and subject-specific assignments are taken as given in the data. However, I limit the sample to classrooms that can be strategically manipulated due to the presence of two or more available teachers (as previously documented in Table 1), in addition to those having more than 5 student observations associated with the teacher so that the teacher will have an estimated teacher effect in the 2010-11 school year.

I execute the simulation by randomly ordering all students in a given school-grade, and then assigning them one at a time to available teachers. For elementary grades, teachers are assigned to all subjects jointly since the classrooms are self-contained; in this case, the marginal class-size effect is averaged across all subjects when optimizing expected student gains in the simulations. For middle grades, teacher assignments are additionally determined by subject as well since instruction is departmentalized, and multiple classes can be assigned to the same teacher.[27,28]

The unconstrained assignment process proceeds as follows: Knowing each teachers' expected productivity functions (given in Equation 6), the principal orders teachers from most to least effective. The principal then successively assigns students to the most effective teacher first until the marginal gain (relative to the next alternative teacher) for an individual student going to the most effective teacher is smaller than the collective losses in learning among students already assigned to that teacher's class. At this point, the next marginal student is assigned to the teacher who is 2nd most

[27] The number of classes an 8th grade teacher is assigned can be variable in practice, so I allowed this to be manipulated in the simulation as well. The number of classes per teacher is constrained to be the maximum observed in the actual data (up to six) across any teacher within the same school-grade-subject combination in the actual data. For instance, if there are two unique 8th grade math teachers in a school, and one teaches three classes while the other teaches one in the actual data, either teacher could potentially be assigned up to three classes in the simulations. I additionally impose a constraint that limits the total number of unique classes within a school-grade-subject to be less than or equal to the original number of classes observed, to ensure that learning gains do not come from artificially lowering class sizes by assigning the same number of students to more simulated classes. To see how sensitive the results were to allowing the number of classes per teacher to be manipulated, I re-ran the simulation requiring teachers to be assigned the same number of courses as observed in the actual data, and large gains were still evident, though the magnitudes were roughly 75% of the values shown in the 8th grade results of Figure 1 across all subjects.

[28] The simulation here does not distinguish between different courses in the same school-grade-subject combination. For example, 8th grade math, pre-algebra, and algebra may all occupy unique course codes in the data, but I treat them interchangeably in the interest of removing artificial constraints in the assignment process, as these course codes are somewhat endogenous to the process of assigning students to teachers and classrooms.

effective. The assignment of successive students will then bounce between these two classrooms as the individual gains are balanced against the collective losses up to the point where the benefit of the next student being assigned to the 3rd most effective teacher is greater than the losses to students in either of the other two classrooms, thereby introducing the 3rd teacher into the assignment process. This process continues, bringing in additional teachers as necessary in descending order of effectiveness, until all students are assigned to classrooms, and the final distribution of class sizes across teachers will reflect the ordering of teacher productivity (large classes for the most productive, small classes for the least productive).

In this simulation, I impose a constraint on the upper limit of the number of students that an effective teacher can accept over the school-grade average (e.g., up to three additional students over the average). The process for constrained optimization of total student learning is similar to that described above. The only difference is to introduce a new teacher when either the optimality condition holds across classrooms, as described in the unconstrained process above, or when the new three-student constraint binds. Yet, even under this scenario with additional constraints, the optimal class sizes will reflect expected differences in teachers' performance.

After all the class assignments are determined, several outcome measures are calculated for students based on those assignments. The calculated outcome measures are described in further detail below:

1. Mean change in student learning – This is calculated with the following equation:

$$\Delta y_{i,c,j,2011} = \Delta \mathbf{I_j} \hat{\boldsymbol{\tau}}_{\boldsymbol{j},\mathbf{2011}} + \Delta \mathbf{Exp_{j,2011}} \hat{\boldsymbol{\zeta}} + \Delta Size_{c,j,2011} \hat{\gamma} + \Delta \mathbf{Class_{c,j,2011}} \hat{\boldsymbol{\delta}} \tag{9}$$

   This equation calculates the change in learning entirely as a function of the change in assigned teacher and classroom characteristics, measured as the simulated value less the original value in the observed data.[29] Note that this expression assumes the other elements of student learning (namely, $\mathbf{y_{i,2010}}\widehat{\boldsymbol{\beta}}_{\mathbf{1}} + \boldsymbol{X}_{\mathbf{i,2011}}\widehat{\boldsymbol{\beta}}_{\mathbf{2}} + \hat{\varepsilon}_{\text{i,c, j,2011}}$) are constant within a student, regardless of teacher assignment and thus drop out of the equation for calculating the change in learning.

2. Proportion of students assigned to top-25% teachers – Teachers appearing in the 2010-11 data are flagged as top 25% teachers based on their estimated value-added spanning the three years prior to the 2010-11 school year. The proportion of students assigned to them is calculated as the count of assigned students to these teachers in the numerator over the total number of students in the 2010-11 school year.
3. Proportion of FRL students assigned to top-25% teachers – This metric is analogous to the one described above, with the exception of counting assignment rates among FRL-eligible students only, and the denominator also only includes FRL-eligible students.
4. Proportion of students assigned to teachers with >= 5 years of experience, MA or higher degree attainment – These metrics are analogous to item #2 above except for the use of alternate indicator variables to flag teacher characteristics, and the FRL versions are analogous to item #3 above.

This classroom manipulation is carried out at varying degrees of "intensity," which represent the number of additional students allowed to be reallocated into the highest-performing teacher's class (relative to the average class-size in the given school-grade-subject combination).[30] These levels range

[29] Note that while the strategic assignment of class sizes to teachers is based on *expected* performance for the 2011 year, the calculation of the change in student learning is based on the *realized* teacher performance for that year.
[30] Note that all legal class size limits are ignored for these particular simulations. According to the National Council on Teacher Quality's Tr3 Database, 28 states have explicit limits on class size in at least one grade, and in states that do not directly limit class size, district-level collective bargaining agreements commonly include such provisions. Only 10 states include explicit restrictions (several other states have recommendations, but not limits) about some element of the distribution of class sizes; for example, where the average class size in a grade may not exceed a certain amount. North Carolina, the state on which this analysis is based, has legal class size limits for grades K-3 only, but any individual class may not exceed the state's funding allotment ratio of teachers to students by more than three students. Shifting a few students across classes appears to be permissible in most states (assuming actual class sizes are not already meeting or exceeding class size limits). Still, laws in some states may need to be amended before schools could adopt aggressive shifting policies that move larger numbers of students, creating systematically

from 0 additional students (i.e., class sizes are equal across teachers in a school, with the exception of any remainder that cannot be divided across the number of classrooms) to 12 additional students (e.g., the best teacher is assigned up to 32 students, if the equal class sizes were expected to be 20). Keep in mind that the additional students are being shifted over from other teachers in the same school and grade—meaning that a weaker coworker may be teaching as few as 8 students at the 12-additional-student level. This simulation is performed over three iterations at each intensity level, as the results vary slightly across each random iteration, and the outcomes are averaged across the three rounds to compute an expected value. [31]

Methodological Limitations

The simulation results rest on several methodological assumptions, which limit the generalizability of the results. Namely, the analysis assumes the difference in student outcomes are determined solely by the difference of estimated teacher productivity and classroom composition effects (including class size); and the simulation assumes a teacher's performance is similarly insensitive to changes the classroom composition and size, beyond that which is accounted for with in the included observable characteristics. On this point, one might consider how shifting students of varying *qualities* may result in differing outcomes overall; yet, I focus specifically on shifting student quantities only. These simplifying assumptions are necessary to make the simulation feasible overall. Either of these assumptions, however, could be violated in practice, which would result in some additional variation in outcomes.

Additionally, the analysis estimates class-size effects that are assumed to be linear and constant across all teachers, though it's possible they may either be non-linear or imprecisely estimated; these

---

unequal class sizes. North Carolina's class size laws would need to be changed to carry out the highest levels of strategic sorting described here.

[31] Three iterations were used because of the computational intensity of the assignment process across all schools in the state administrative data. For a subset of the data, 20 iterations were used to approximate 90-percent confidence intervals through dropping the largest and smallest values. These confidence intervals were extremely narrow, and provided confidence that averaging over three iterations was giving a sufficiently precise result.

two properties of the assumptions warrant discussion. First, the linear class-size effect I use here is justified in both the sample and prior research.[32] I do not, however, interact the class-size effect with teacher effectiveness measures; were it the case that individual teachers' class-size effects were negatively correlated with their effectiveness, as Woessman and West (2006) have speculated, the results presented here would understate the potential learning gains to this strategic class size approach.[33] And second, it's possible that the class-size effects estimated here may either under- or over-estimate real class-size effects, though this is not a serious threat to the usefulness of the approach presented here. On the one hand, if class-size effects were actually smaller than estimated (i.e., zero), teacher assignment itself would be the only key component in calculating student learning gains. Under such a scenario, the optimal allocation of students would place all students with the best teacher in the grade, and would simply result in a much smaller teacher workforce. On the other hand, if class-size effects were consistently larger, say on the order of magnitude as those presented in Krueger (1999), the corresponding implication is that the variation in teacher quality is much larger.[34] In this case, strategically shifting students based on these larger-variation teacher effects results in even stronger gains from implementing the policy, though fewer students would be shifted across classes before achieving the optimal level of learning.

Results

[32] I fit regressions using a non-linear class size function on the analysis sample, and found no consistent evidence of a non-linear relationship. Whitehurst and Chingos (2011) also interpret the prior research evidence on class size to suggest a linear class-size effect.

[33] Using international data, Woessman and West (2006) find class-size effects are smaller in magnitude in countries with more qualified teacher workforces, and larger in countries with less qualified teachers. They speculate this finding may suggest an interaction between class-size and teacher productivity where class-size effects are larger for less effective teachers but near zero for the most productive teachers. If true, the strategic assignment examined here becomes even more productive by reducing class size where it matters most for student learning, and increasing class sizes where the extra students make little difference.

[34] Recall the slightly positive sorting between class size and teacher quality; if class-size effects are actually larger in magnitude, true teacher effects must show greater variability than what is estimated in the analysis sample. The simulation was conducted with larger imposed class-size effects to investigate this and found net effects that were larger for transferring small numbers of students, though these gains plateaued after transferring just a few students.

Results of Baseline Assignment Observed in North Carolina Data

To document how students are actually allocated to teachers in the North Carolina data, Table 3 illustrates the baseline condition of class size assignments with respect to prior teacher performance in the 2010-11 school year. It contains three noteworthy points. First, NC schools have some naturally occurring variation in class size within the same school (see row 2). The reported metric is average class size deviation from the mean within school-grade; for example, a value of 2.738 for 5th grade math implies large 5th grade math classes in a given school had about 3 extra students than the school's average for the grade, while smaller classes in the same school teaching an equivalent course had 3 fewer students. In other words, using the average 5th grade math class size of about 22 students as a reference, a school may typically have one larger class with 25 students and another smaller class of 19 students. I find that average class size deviation is consistently larger in 8th grade (ranging from 3.8 to 5.7) than in 5th grade (1.7 to 3.0).[35]

Second, the within-school correlation coefficients between expected teacher performance and class size (row 3) are slightly positive across all of the samples, ranging from 0.01 to 0.13.[36] This suggests that some strategic assignment of students to more and less effective teachers may already be occurring to a limited extent—but it appears to be very slight. A helpful way of conceptualizing this relationship between class size and effectiveness is to think of a single "leftover" student.[37] That is, if classes were equally assigned to all teachers at the beginning of the year and then a single student arrives on the first day of school, to whom should the one leftover student be assigned? If schools were maximizing

[35] Note that the average class-size and average class-size deviation values vary across subjects in 5th grade, though these are self-contained classrooms. This occurs because some schools will maintain separate records for each subject that a self-contained instructor teaches, and the class size values may vary slightly across the subjects, even though there is no notable difference in the classroom lists provided in the administrative data.

[36] I formally estimate this relationship between class size and expected teacher performance using a regression with school fixed effects, and the positive associations seen here are statistically significant. The positive association between these variables is consistent with the evidence presented in Barrett and Toma's (2013) analysis using data from 10 districts in Kentucky, again suggesting that some strategic class-size determination may already be occurring to a limited extent.

[37] For the purpose of this table, a leftover student is any student assigned to a teacher above the school-grade average class size.

achievement, this leftover student should naturally go to the teacher expected to perform the best—somebody has to accept the larger class, so assigning the student to where he or she would be best-served optimizes overall learning. Row 4 of the table shows that in practice, fewer than half of these leftover students are actually assigned to teachers in the top 25%. By way of comparison, Row 5 shows only slightly smaller proportions going to teachers identified in the bottom 25% (8th grade math and reading are slightly more likely to be assigned to these teachers). Hence, while there is a small tendency to assign larger classes to better teachers in the data, it appears very slight indeed.

And finally, the percentage of students assigned to teachers in the top 25% of effectiveness ranges from 24% to 29% (row 6).[38] However, the percentage of FRL-eligible students assigned to top-quartile teachers (in row 7) ranges from 22% to 26%, roughly 2 percentage points lower across most of the six samples, revealing a gap in access to effective teaching for economically disadvantaged students. This gap means that disadvantaged students tend to get relatively weaker teachers, which can reinforce pre-existing achievement gaps.[39]

This last finding warrants further discussion. Note that these gaps in access may arise in two ways: 1) due to the uneven distribution of effective teachers across schools (high-need schools tend to have a harder time attracting the most effective teachers), and/or 2) due to disadvantaged students being disproportionately assigned to the weakest teachers in a school. A recent study of access to effective teaching in 29 school districts reports the across-school access gap to effective teachers is the larger of the two, accounting for 60 percent or more of the total access gap (Isenberg, et al., 2013). The class-size shifting strategy analyzed here *only* addresses problems of within-school sorting by directly

[38] Designations for top-25% teachers are based on teachers' expected performance measures, which combine prior value-added estimates with the effect from a teacher's experience, among all teachers available to teach in the 2010-11 school year.

[39] Though not reported in Table 1, a gap in access also arises when using more traditional teacher credentials. For instance, roughly 71% of 5th grade students are assigned to teachers with five or more years of teaching experience, but 69% of FRL students have such assignments. The corresponding numbers for 8th grade are 72% and 70%, respectively. If measuring the proportion of students assigned to teachers with a master's degree or higher, the proportions are 33% overall and 32% among FRL students in 5th grade, and 34% overall and 32% among FRL students in 8th grade.

increasing student access to the most effective teachers within a school. Because most of the gap in access occurs across schools, this strategy is not expected to remediate the entire access gap for disadvantaged students.

Simulated Learning Gains Under Strategic Allocation

The simulated gains in average student learning are presented in Figure 1, which include 6 graphs—one for each tested subject (across columns) in both 5th and 8th grades (down the rows). Each graph in the figure presents the mean changes in student learning on the vertical axis. The horizontal axis represents the maximum number of additional students allowed in an effective teacher's classroom; recall the value of 0 represents equally-sized classrooms, and movement to the right means that students are more intensively reallocated to the top-performing teachers and away from the weakest.

Figure 1 shows small simulated gains across all subjects that generally increase as the simulation allows for more students to be assigned to the most effective teachers. As expected, net gains in student learning are realized as the best teachers teach larger classes and the weakest teachers have progressively smaller ones. Note, however, that the magnitudes of the student learning gains on the vertical axis differ across grades, and represent greater gains in the 8th grade estimates than in 5th grade.

An interesting pattern is the shape of these learning gains—they are not linear but generally curve at higher intensity levels, showing a diminishing return to the policy. In other words, the first few students shifted from the weakest teachers to the strongest represent the largest potential gains associated with the policy. Allowing still more students to be shifted across classes still results in net student learning gains, but these marginal gains are smaller than those expected from the initial set of shifted students. In fact, across all grades and subjects (with the exception of 8th grade reading), over 75% of the potential gain from allowing up to 12 students to be assigned to the best teachers' classes is already realized when allowing just six students to be shifted. Hence, even shifting small numbers of students may generate the bulk of the potential gain from this strategy—an important nuance for schools that may wish to implement this approach.

Still, casual observation of these figures may lead one to conclude that, because the gains appear small, they may not be worth the effort to change; but this is not the case. While magnitudes of simulated gains are indeed modest, these are *average* learning gains across all students, not just those who are moved. In practice, the gains are accrued primarily to individual students who are shifted into stronger teachers' classrooms. But when averaged across all other students in both classrooms who stay put, the average effect seems more modest. Yet even these average gains can be consequential in the aggregate. In the larger 8th grade estimates, the potential gains of 0.02 standard deviations in math and science, with as many as 12 additional students, amount to an additional 2.5 weeks of schooling.[40] While reassigning 12 students to the best teacher is likely not always feasible, even 6 additional students can result in gains of 0.015 standard deviations, which is almost two weeks of additional schooling in 8th grade math and science. Similar levels of student shifting translate to nearly a week of additional schooling in 8th grade reading. These are considerable increases obtained by simply reorganizing classes in ways that don't differ much from current practice (NC classes already show disparities in class size near this level)—principals need only be more strategic about it. In the 5th grade estimates, the potential gains from this strategy are more modest. Allowing up to 12 additional students for the best teachers can produce math and science achievement gains of 0.005 standard deviations, and this equates to roughly two additional days of schooling.

The largest potential improvements from this strategy come from the 8th grade results. Why such a large difference between grades? I speculate this is driven by the fact that elementary classes are self-contained, so when students are assigned to a teacher, they receive instruction in all subjects from

[40] A gain of 0.02 SD of student achievement is approximately 7% of the average annual gain for the grade 7 to 8 transition, based on nationally normed tests (Hill, et al., 2008). These values and the others that follow in this discussion are converted to days of learning based on a 180-day school calendar for ease of interpretation here.

him or her. This self-contained approach has three consequences. First, the most consequential difference reflects the fact that effective teachers in 5th grade only have one class of students, and are therefore exposed to fewer additional students in total, whereas effective teachers in 8th grade may teach multiple sections of the same class, all of which can be larger than the school average. Second, value-added estimates from self-contained grades are less stable from year to year than those from departmentalized grades that are based on multiple classes (McCaffrey, et al., 2009), so a principal's predictions of elementary teachers' performance (assuming value-added measures are a major criterion of that prediction) are going to be less reliable. And third, we know that teacher performance is positively correlated across subjects (Goldhaber, et al., forthcoming), but averaging imprecise performance estimates for the same teacher across subjects dilutes the potential gains. In practice, elementary school principals may have more reliable performance data on teachers (separate from value-added estimates used here) and may therefore be able to realize greater gains than what are simulated, but this is speculative.

Based on the theory laid out above, one could reasonably predict that the potential gains from class-size shifting strategies are largest in the cases where the expected difference in productivity between teachers in the same school is large. I find that prediction to be true, though the full results are omitted for brevity (available upon request). To investigate this, I first flagged schools in the analysis samples where the teachers with either the highest or lowest expected performance were significantly different from other available teachers in the school grade; flagged schools constituted between 51 to 70 percent of the original samples of schools with two or more teachers shown in Table 1. I then calculate the gains from strategically allocating class sizes based on these subsamples only. The resulting estimates showed general patterns similar to what is pictured in Figure 1, though the magnitudes of the simulated gains were 20-30% larger across all of the intensity levels, in both grades and all three subjects.

Inherent Risk in Strategic Allocation

An important caveat to these findings is that the gains portrayed in Figure 1 represent averages, and are not guaranteed for any given school. At the heart of this strategic class-size approach is a gamble—principals are taking a small net class-size “penalty” when classes are disproportionately assigned at the beginning of the year in exchange for higher *expected* student gains among certain teachers. So while the expected payoff is positive, there’s a chance principals inadvertently reallocate students amongst teachers that results in an inefficient outcome—for example, the most effective teacher based on past performance may have a bad year due to personal issues, or a rookie teacher may prove more effective than anticipated. Consequently, some principals will inadvertently place more students with teachers who end up performing worse than those tasked with smaller classes, resulting in losses of learning (compared to what would have been the case with equal class sizes). However, on average, the overall gains for students are expected to be positive.

The riskiness of this approach is demonstrated in Table 4, which presents the interquartile range of simulated learning changes at select intensity levels. These ranges represent the 25th and 75th percentiles of mean changes in learning, when averaged at the school-grade-subject level. In other words, 50 percent of all schools manipulated in this simulation fall within this range, the remaining 50 percent fall equally above and below these ranges. Note that the lower numbers in these ranges (i.e., the 25th percentile) are negative in all cases, showing a non-trivial amount of schools are actually worse off through strategic assignment, at least in this one year of simulations. If schools were to use this strategy in multiple years and across multiple grades, however, one could reasonably expect overall outcomes to converge to the mean expected value. Additionally, the risk increases (i.e., the interquartile ranges widen) as more additional students are allowed to sort into effective teachers’ classrooms, which is expected. In general, though, the range is mostly expanding on the upside as more students are shifted across classrooms in the simulation. The upside risk is particularly evident in the case of 8th grade math and science, where the 25th percentile values in both cases show relatively little movement as the number of additional students increases, while the 75th percentile show large upward movements representing larger potential gains.

Simulated Access to Effective Teachers

The net gains in student learning shown above are a direct result of placing more students in effective teachers' classrooms, so next I investigate how much this changes student access to effective teachers. Figure 2 presents various metrics representing the patterns of assignment. It shows the proportions of students assigned to teachers with three different characteristics (across the columns): teachers in the top 25% of effectiveness (first column), teachers with 5 or more years of experience (second column), and teachers holding a master's degree or higher (third column).[41] Each graph includes three lines: one showing an overall level of exposure, and then separate lines documenting access by students' FRL eligibility status.[42]

As shown in Figure 2, greater proportions of students overall are assigned to teachers with these characteristics as more students are strategically shifted across classrooms. Yet, these metrics do not show the same rate of increase—the proportions climb most quickly for top-25% teachers, but more slowly for more experienced teachers and those with graduate degrees. This is expected, as prior performance is the measure that is directly used in determining when to shift students across classrooms, and these other characteristics are only weakly correlated with it.

Yet, look beyond the overall trends and instead focus on the lines by FRL status, which show a slightly different story. While both FRL and non-FRL lines slowly increase when additional students are allocated to more effective teachers, they stay roughly parallel. This same pattern is observed across all

[41] Prior teaching experience and holding a graduate degree are not equally predictive of future performance (see Goldhaber and Hansen, 2013); however, they both may signal greater commitment to the teaching profession and are more common metrics of student access to credentialed teachers.

[42] In 8th grade, teacher assignments are made by subject, and these graphs vary slightly across subjects, though the qualitative findings and persistent gaps are consistent across subjects. The 8th grade measures represented in Figure 2 are calculated from the 8th grade math sample. Teacher assignments in 5th grade are not subject specific.

three measures of teacher quality. Thus, class-size shifting does not appear to mitigate the relative gap in access compared to non-disadvantaged students, although it does improve disadvantaged students' absolute level of exposure to qualified teachers overall. By construction, strategic assignment as implemented here will increase student access to effective teachers within schools; so, this persistent gap is due to the fact that effective teachers are not equally allocated across schools. Consequently, this class-size shifting strategy alone cannot reduce pre-existing inequalities, and some other intervention would be necessary to remediate student access gaps in teacher quality entirely.

Feasibility Issues

The feasibility of this class-size shifting strategy ultimately rests on whether teachers are willing to participate, which raises issues related to compensation, which brings with it the question of whether such a strategy would be feasible in constrained education budgets. Additionally, this approach implies a consequential shift in the status quo of schools in multiple other dimensions. Both financial considerations and other various considerations are addressed in this section.

Cost Considerations

Not all teachers would eagerly accept additional pupils, but many appear open to the prospect of earning more pay in return for taking on more pupils.[43] For instance, a teacher compensation survey in Washington state (2006) found 83 percent of educators would prefer an additional $5,000 in compensation versus the alternative of having two fewer students in their classes (Goldhaber, et al., 2010). And a recent nationwide survey by *Education Next* and Harvard's Program on Education Policy and Governance posed the question slightly differently, asking teachers whether they would prefer higher wages (an extra $10,000 per year) or smaller classes (3-student reduction). Forty-two percent of

[43] It's worth clarifying that teachers should not be compelled to take more students on if they are not willing. If teachers were forced to take on larger classes, one could logically expect them to intentionally underperform in order to avoid this consequence, which obviously undermines the intent of the strategy.

teachers chose the higher salary compared to 47 percent who preferred smaller classes. But these responses appear to be driven by teachers' current class sizes, with those in high-class-size states tending to choose reductions and those in low-class-size states favoring the money.[44] So, depending on the current level of class sizes, a proposal to increase them for the most effective teachers could potentially be met with some enthusiasm. Parents also appear to be supportive of such an approach, with 73% preferring a large class taught by an effective teacher to a regular class-size and a randomly chosen teacher, according to survey results from the FDR Group (Farkas and Duffett, 2012).

Is extra compensation for teachers necessary to implement this strategy? Perhaps not. Low-intensity shifting, where only a few additional students are assigned to the highest-performing teachers, may be implemented without requiring additional pay. As shown in Table 3, some naturally occurring class-size variation already exists within schools—on the order of three additional students in 5th grade and five in 8th grade. Presumably most of these class-size differences are not compensated. Hence, if principals operated a little more strategically by intentionally giving say, three to six more students to the best teachers, their schools could potentially experience a disproportionate share of the potential gains under the policy without additional costs. It may also be possible to make a large class a more desirable assignment by providing non-monetary rewards in compensation for the larger class; for example, recognizing teachers as "lead teachers" for the grade or subject, giving them first choice when it comes to non-teaching assignments (or reducing them altogether), allowing increased time for lesson preparation, or providing them increased access to teacher aides.

Additionally, one of this strategy's unique features is that it provides a way of paying high-performing teachers more—under the cover of giving them more students. It also suggests a low-resistance path around the constraints of contracts and salary schedules since it is readily defended as

[44] Matthew Chingos presents these results and his further analysis in a blog post, available at: http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2013/01/30-class-size-chingos, accessed June 21, 2013.

"extra pay for extra work." The policy outlined here simply suggests being selective about which teachers receive the extra students (and extra pay). In order to fund this, dollars could be re-purposed in multiple ways, including lower spending on instructional specialists (who may prove less essential given the expected boost in student achievement) and lower pay for teachers assigned fewer students (though it is unclear how popular this approach may be).[45] Finally, money may be found in some schools where dwindling class sizes for the least effective teachers might compel principals to simply let their class sizes drop to zero (i.e., let them go and not replace them). In this case, average class sizes may increase across for all teachers in the school, but with strategically determined class sizes this need not imply a loss of learning.

Finally, even if entirely new funds were needed to compensate teachers for participating, the strategy may still be worth it—at least in 8th grade. Based on the estimates presented in Chetty, et al.'s (2011) analysis of long-run effects of teachers, the simulated student learning gains calculated here can be translated into the value of future student wages (discounted to the present), which are shown in Figure 3 (the black lines).[46] This figure includes two additional lines, representing the average per-student cost associated with paying teachers a bonus in proportion to the additional students in their classes. The two bonus lines pictured here depict bonuses of $2,000 (in blue) and $4,000 (in green) for each additional student. For example, an effective teacher agreeing to teach 25 students, when she would otherwise have had a 20-student class under equally-sized classrooms, is given a $10,000 bonus under the $2,000 level. Note that bonuses are assumed to be paid for *extra* students, while the cost is averaged across *all* students.

[45] Public Impact's Financial Planning Summary explores a variety of sources where schools may reallocate current spending to finance their "Reach Extension" models in a cost-neutral way. These are just a few of them.

[46] The authors estimate the marginal impact of a one standard deviation increase in teacher quality on the net present value of students' future earnings at age 28 at 0.9%, or roughly $4,600 per grade (Chetty, et al., 2011, p. 39). Given that the simulated gains in student learning calculated here are based primarily on increasing students' exposure to effective teachers, these estimates are applied to the results to approximate the expected student benefit of strategic shifting.

While these cost and benefit lines are not technically accrued to the same account (the benefit accrues to students, the cost to the taxpayer), comparing them shows what the expected return might look like.[47] The horizontal distance between the benefit and the cost lines captures the return: if the benefit line exceeds the cost line, the return is positive, and vice versa for a negative return. Based on these figures, raising taxes to pay for teacher bonuses in 5th grade provides essentially a zero return at the $2,000 level, and a negative return at the $4,000 level. However, returns are universally positive in 8th grade even when providing generous additional compensation for teachers.

Other Feasibility Considerations

If schools are serious about formalizing this strategy, and implementing it more intensively than just reallocating a few students, several additional items warrant discussion. First, we have practical constraints to implementation. Strategic shifting has two necessary conditions: 1) there must be multiple instructors in a school leading separate classrooms with equivalent content, and 2) principals need relevant performance data to determine which teachers are better and by how much. If these conditions are not met, implementation may be difficult or impossible. For example, some schools engage in tracking students by ability level, thus qualitatively differentiating otherwise similar classes; this complicates the shifting strategy as students are not as readily moved from one class to another (though some marginal students may be transferred without seriously affecting the tracking). Similarly, if only crude performance categories are available to school decision-makers, or if performance measures were not available in a timely way, the strategy could be severely undermined.[48]

---

[47] It may seem counter-intuitive to directly compare future private benefits with current public costs, but this is an inherent property of public education, which is an inter-generational investment converting wealth from current taxpayers to human capital in students. If current costs exceed the value of the future benefit to students, this is obviously a bad investment.

[48] Note that for the purposes of the simulation, value-added data and experience were the only measures available to me to predict classroom performance. In practice, however, principals may have access to and use a wide range of performance measures to form their expectations of teacher performance.

Second, there are state laws, district policies, and bargaining agreements that may stand in the way of implementation. Some laws or agreements already have compensation clauses built in for teachers with class size overages, which could facilitate the adoption of more strategically determining who gets the larger class and compensation. But, this strategy cannot be implemented (beyond the hypothetical leftover student) where class sizes are already pressing their legal limits. In most other circumstances, though, schools are likely able to adopt an informal version of shifting a few students across classrooms without any major policy change. More intensive levels of shifting will likely require changes to laws, policies, collective bargaining agreements, or even perhaps all three in some locales.

Third, it's worth considering how this strategy, though seemingly subtle, may disrupt school environments. This will implicitly set up a hierarchical ordering that differentiates teachers based on class size—reflecting teacher effectiveness—which can change interpersonal dynamics in a profession that, at present, is largely flat. Whether the creation of this hierarchy is helpful to teachers (as it creates a type of career ladder) or harmful (as it raises barriers between colleagues) is debatable. This hierarchy could also constitute a publicly visible signal of teacher effectiveness to parents (where they are presumably less informed at present), which could therefore induce a change in parent behavior in light of this information.

Fourth, on an individual level, teachers' behavioral responses to the number of students in their class also warrant discussion. More students for the best teachers each year could inadvertently burn them out prematurely. Smaller classes for low-performing teachers with equivalent pay may unintentionally create a perverse incentive, in which workers may intentionally shirk on effort in order to improve their working conditions. The possibility for these adverse responses speaks to the importance of making sure teachers are willing to participate, and are compensated accordingly (either through monetary or other non-monetary method) for the change in class size (which is a consequential change in working conditions for both strong and weak teachers).

And finally, class-size shifting at high intensity levels could also introduce some logistical issues for physical facilities. To accommodate large variations in class size, schools would need physical classrooms that reflect that variation, with some large spaces and some small spaces (or perhaps dividers that could flexibly partition instructional spaces). This stands in contrast to the common school blueprint, which is typically designed around the model of equally sized classrooms.

Conclusion

Can student achievement improve by simply reallocating pupils among teachers? These results say yes. In both 5^{th} and 8^{th} grades, the adoption of even a modest strategic class-size shifting policy can potentially improve student learning in math, reading and science. Though the simulated gains are relatively small in 5^{th} grade, gains in 8^{th} grade are large enough to result in a net-positive return even when generously compensating teachers for accepting the additional student load. As a result, this strategic approach to class size shows promise, particularly in contexts where teacher assignments are subject-specific, as a single policy to reduce students' exposure to weak teachers, improve student access to effective instructors, and reward effective teachers.

These findings have several implications for state and local policymakers relative to equity and efficiency. First, this approach shows little promise in improving the equity of the public education sector. As shown in Table 3, gaps currently exist in students' access to teacher quality—whether that is measured as value-added performance or with more traditional credentials. Although Figure 2 shows some improvements for economically-disadvantaged students as more students are allocated to higher performing teachers, the gaps in access persist. These gaps represent differences in teacher quality *across schools*. Thus, class-size shifting *within* schools alone cannot bridge them, although policymakers could consider other strategies in tandem with strategic shifting to reduce these gaps, including: differential pay or bonuses for teaching in high-need schools; directly transferring high- or low-performing teachers across schools; or the use of technology to transmit effective instruction across multiple schools.[49]

More promising are the findings related to efficiency, which refers to the overall productivity of the workforce in terms of student learning, given the current assignment and stock of teachers in the system. Reallocating students can enhance efficiency in both elementary and middle grades, though the gains are particularly notable in the latter. These improvements are delivered primarily through increasing student exposure to effective teachers, and reducing the class sizes of less-effective teachers. Combined, these assignments translate to better outcomes for students in both larger and smaller classes, presumably in both cognitive and non-cognitive measures.[50] The simulated learning gains were notably larger in the subsample of schools where teachers' prior performances were significantly different.

The simulated learning gains in 8th grade are large enough to warrant special mention. For those districts committed enough to aggressively pursue class-size shifting, they could realize the equivalent of an additional two and a half weeks of instruction in 8th grade math and science. Yet, even less intensive strategies only shifting six additional students in effective teachers' classes—class-size disparities near what I observe in 8th grade in the North Carolina data—result in gains equivalent to almost two extra weeks of instruction.

It may be useful to compare these gains with other proposed policy interventions. Another policy to promote teacher quality directly removes the lowest performing teachers from the classroom

[49] A recent study by Grissom, et al. (2013) describes an equity-improving involuntary transfer policy in Miami-Dade schools in which low-performing teachers in disadvantaged schools were relocated to schools serving lower-risk students. Glazerman, et al.'s (2013) recent analysis of the Talent Transfer Initiative showed similar equity-improving results from incentivizing high-performing teachers to teach in low-performing, disadvantaged schools.

[50] Students assigned to high achieving students benefit on several dimensions, as Chetty, et al. (2011) show. Those in smaller classes also benefit; the cognitive evidence was discussed previously, and Dee and West (2011) estimate positive returns on non-cognitive dimensions, such as school engagement, due to smaller class sizes in 8th grade. The class size strategy laid out here increases the overall variation in class sizes, increasing sizes for the most effective teachers while simultaneously decreasing sizes for the least effective. Inasmuch as students are expected to be exposed to both large and small class sizes as they progress through grades, students may actually benefit more from either assignment than what would have otherwise been expected.

(keeping class sizes unchanged). The simulated gains in 8th grade math and science achieved by shifting just 6 additional students into the classrooms of effective teachers are equivalent to the expected effect of removing the lowest 5% of teachers in these subjects—*and this could potentially be achieved without actually removing them!*[51] Rather, class-size shifting enables the lowest-performing teachers to become more effective than they may be otherwise by allowing them to individualize instruction in smaller classes. Another comparison policy is class-size reduction: the simulated gains of nearly 0.015 standard deviations of student achievement in 8th grade math and science for shifting 6 additional students is equivalent to the expected gain from class size reductions of 4 to 6 students (based on the Table 2 estimates in these subjects). Cost estimates for a one-student reduction in class size range from $161 to $420 per student (Chingos, 2013), similar to the range of costs required for the bonuses for 6 additional students in Figure 3. Yet, the strategic shifting is simulated to produce results 4 to 6 times larger than direct class-size reduction.

The sizable simulated gains in 8th grade contrast with the lesser gains estimated for 5th grade. Recall that these differences are likely due to self-contained classrooms in this grade. What if schools didn't use self-contained classrooms? The results may be very different. Using a simulation, Goldhaber, et al. (forthcoming) estimate gains in student learning under subject-specialized elementary grades on the order of roughly 0.04 or more standard deviations of student achievement, or roughly 10 times the 5th-grade gain estimated from the class-size shifting strategy presented here. If elementary grades were departmentalized, as recommended in Jacob and Rockoff (2011), pairing the gains from departmentalization with those from a class-size shifting strategy would presumably increase the expected gains even more.

[51] Based on the expected mean of the teacher distribution after removing the bottom 5% (calculations come from Greene, 2000, p. 899). These values assume one standard deviation of teacher quality is equivalent to 0.14 standard deviations of student achievement, which is the mean of the standard deviations across 8th grade math and science in Table 2. Hanushek (2009) estimates that the cumulative effect of removing the lowest-performing 6 to 10 percent of teachers from the profession would be sufficient to bring overall achievement levels in the United States to those in Canada.

In sum, strategically allocating more students to more effective teachers can produce significant gains in student achievement and modest overall gains in student access to effective teachers.[52] Fortunately, pursuing this policy on a limited basis would likely require very small changes in practice, requiring only minor changes in the status quo of how students are allocated across teachers, and could yield meaningful gains in student learning. On the other hand, if schools are willing to formalize rewards for taking on extra students and to allow class sizes to vary more than usual, the potential gains to student learning, in 8th grade at least, appear to outweigh the modest effort required to make it happen.

[52] Of course, the specific class-size shifting strategy analyzed here is one of many potential ways that more students may be exposed to high-quality teaching. Other strategies include digital learning to supplement high-quality classroom instruction without lowering class sizes, or leadership roles that allow effective teachers to more directly manage and supervise other teachers' efforts. Public Impact's Opportunity Culture initiative provides a host of alternative approaches to extend the reach of high-performing teachers.

References

Aaronson, Daniel, Lisa Barrow, and William Sander, (2007). “Teachers and Student Achievement in the Chicago Public High Schools,” *Journal of Labor Economics* 25: 95–135.

Ahn, June and Dominic J. Brewer, (2009). “What Do We Know About Reducing Class and School Size?” in *Handbook of Education Policy Research*, eds. Gary Sykes, Barbara Schneider, and David N. Plank. New York, NY: Routledge.

Barrett, Nathan and Eugenia F. Toma, (2013). “Reward or Punishment? Class Size and Teacher Quality,” *Economics of Education Review*, 35: 41-52.

Chetty, Raj, John Friedman, and Jonah Rockoff, (2011). “The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood,” National Bureau of Economic Research working paper #17699.

Chingos, Matthew M. (2013). “Class Size and Student Outcomes: Research and Policy Implications,” *Journal of Policy Analysis and Management* 32(2): 411-438.

Dee, Thomas S. and Martin R. West, (2011). "The Non-Cognitive Returns to Class Size," *Education Evaluation and Policy Analysis* 33(1): 23-46.

Duncan, Arne, (2010, Nov. 17). “The New Normal: Doing More with Less,” Secretary Arne Duncan's Remarks at the American Enterprise Institute. http://www.ed.gov/news/speeches/new-normal-doing-more-less-secretary-arne-duncans-remarks-american-enterprise-institut

Farkas, Steve and Ann Duffett, (2012). “How Americans Would Slim Down Public Education,” Washington, DC: Thomas B. Fordham Institute.

Freedberg, Louis, (2012, May 31). “Class Size Reduction Program Continues to Unravel,” EdSource. http://www.edsource.org/today/2012/class-size-reduction-program-continues-to-unravel/8730#.UchwLPn5GSp

Gates, Bill, (2011, Feb. 28). “How teacher development could revolutionize our schools,” *Washington Post*. http://www.washingtonpost.com/wpdyn/content/article/2011/02/27/AR2011022702876.html

Glazerman, Steven, Ali Protik, Bing-ru Teh, Julie Bruch, and Jeffrey Max, (2013). *Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment*, (NCEE 2014-4003). Washington, DC: National Center of Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goldhaber, Dan, Michael M. DeArmond, and Scott DeBurgomaster, (2011). "Teacher Attitudes About Compensation Reform: Implications for Reform Implementation," *Industrial Labor Relations Review* 46(3): 441-463.

Goldhaber, Dan and Michael Hansen, (2013) "Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance," *Economica* 80(319): 589-612.

Greene, William H., (2000). *Econometric Analysis*. Upper Saddle River, NJ: Prentice-Hall.

Hanushek, Eric A., (2009). "Teacher Deselection," in *Creating a New Teaching Profession*, ed. Dan Goldhaber and Jane Hannaway. Washington, DC: Urban Institute Press.

Hanushek, Eric A. and Steven G. Rivkin, (2010). "Generalizations about Using Value-Added Measures of Teacher Quality," *American Economic Review* 100 (2): 267-271.

Hill, Carolyn, Howard Bloom, Alison Black, and Mark Lipsey, (2008). "Empirical Benchmarks for Interpreting Effect Sizes in Research," *Child Development Perspectives*, 2(3): 172–177.

Isenberg, Eric, Jeffrey Max, Philip Gleason, Liz Potamites, Robert Santillano, Heinrich Hock, and Michael Hansen (2013). *Access to Effective Teaching for Disadvantaged Students* (NCEE 2014-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Jacob, Brian A. and Jonah E. Rockoff, (2011). "Organizing Schools to Improve Student Achievement: Start Times, Grade Configurations, and Teacher Assignments," The Hamilton Project Discussion Paper 2011-08. Washington, DC: Brookings Institution.

Krueger, Alan B., (1999). "Experimental Estimates of Education Production Functions," Quarterly Journal of Economics, 115(2): 497–532.

McCaffrey, Daniel F., Tim R. Sass, J.R. Lockwood, and Kata Mihaly, (2009). "The Intertemporal Variability of Teacher Effect Estimates," *Education Finance and Policy*, 4(4): 572-606.

National Council on Teacher Quality, (2012). "*State of the States 2012: Teacher Effectiveness Policies,"* Washington DC: National Council on Teacher Quality.

Scholastic and the Bill and Melinda Gates Foundation, (2012). "Primary Sources: 2012 – America's Teachers on the Teaching Profession," New York, NY: Scholastic.

Staiger, Douglas O. and Jonah E. Rockoff, (2010). "Searching for Effective Teachers with Imperfect Information," *Journal of Economic Perspectives* 24(3): 97-118.

Whitehurst, Grover J. and Matthew M. Chingos, (2011). "Class Size: What Research Says and What it Means for State Policy," Washington, DC: Brookings Institution.

Woessmann, Ludger and Martin West, (2006). “Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS,” European Economic Review, 50(3): 695–736.

Tables and Figures

Table 1. Choice Sets of Teachers Available for School Principals

	5th grade	8th grade		
		Math	Reading	Science
1 teacher	186	189	148	235
2 teachers	201	143	106	140
3 teachers	279	130	93	130
4 teachers	293	108	87	85
5+ teachers	451	176	280	113
Total	1,410	746	714	703

Note: Count of unique schools serving the particular grade / grade-subject combination. The number of available teachers in the rows represent unique teacher identifiers.

Table 2. Estimated Class-size Effects and Teacher Value-added Variation

	Grade 5			Grade 8		
	Math	Reading	Science	Math	Reading	Science
Class size	-0.0052***	-0.0020***	-0.0047***	-0.0035***	0.0000	-0.0024***
	(0.0005)	(0.0005)	(0.0005)	(0.0002)	(0.0003)	(0.0003)
Standard deviation of teacher effects	0.1513	0.0801	0.1927	0.1333	0.0612	0.1500

Note: ***: $p<0.01$. Estimates derived from teacher fixed effects model across three years of data prior to 2010-11 school year using the North Carolina administrative data. Included covariates include prior test scores in reading and math, student characteristics, class size, classroom composition, and teacher experience.

Table 3. Snapshot of Observed Class Size Assignment in North Carolina for 2010-11

	Grade 5			Grade 8		
	Math	Reading	Science	Math	Reading	Science
1. Average class-size	21.9	21.6	22.6	21.6	21.3	23.9
2. Average class-size deviation from mean within school	2.738	3.073	1.743	5.583	5.680	3.815
3. Within-school relationship between teacher performance and class size	0.071	0.134	0.046	0.028	0.015	0.014
4. Proportion of leftover students assigned to top 25% teachers	0.243	0.249	0.259	0.303	0.288	0.380
5. Proportion of leftover students assigned to bottom 25% teachers	0.228	0.221	0.246	0.318	0.296	0.369
6. Proportion of students assigned to the top 25% of teachers in prior value added	0.258	0.287	0.237	0.251	0.244	0.254
7. Proportion of FRL students assigned to top 25% of teachers in prior value added	0.235	0.260	0.217	0.232	0.243	0.226

Note: Statistics derived from student-teacher assignments based on 2010-11 school year.
How to read this table: The six samples (corresponding to each grade and subject combination) used in the study are presented across the columns of the table. Measures representing current class size assignments (down the rows) are calculated in each of the samples. In Row 6, for example, 25.8% of students in the 5th grade math sample are assigned to teachers in the top 25% based on prior value-added estimates.

Table 4. The Interquartile Range of Simulated Learning Changes

	Additional Students	Math	Reading	Science
Grade 5	3	[-0.0075 - 0.0122]	[-0.0056 - 0.0110]	[-0.0072 - 0.0125]
	6	[-0.0089 - 0.0151]	[-0.0069 - 0.0133]	[-0.0092 - 0.0154]
	12	[-0.0099 - 0.0159]	[-0.0074 - 0.0141]	[-0.0101 - 0.0164]
Grade 8	3	[-0.0065 - 0.0260]	[-0.0088 - 0.0126]	[-0.0029 - 0.0234]
	6	[-0.0068 - 0.0326]	[-0.0102 - 0.0164]	[-0.0061 - 0.0310]
	12	[-0.0066 - 0.0405]	[-0.0153 - 0.0273]	[-0.0071 - 0.0455]

Note: Values in brackets represent the interquartile range (25th and 75th percentiles) of simulated changes in student learning averaged at the school-grade-subject level.

Figure 1. More Students in Classes of Effective Teachers May Produce Learning Gains

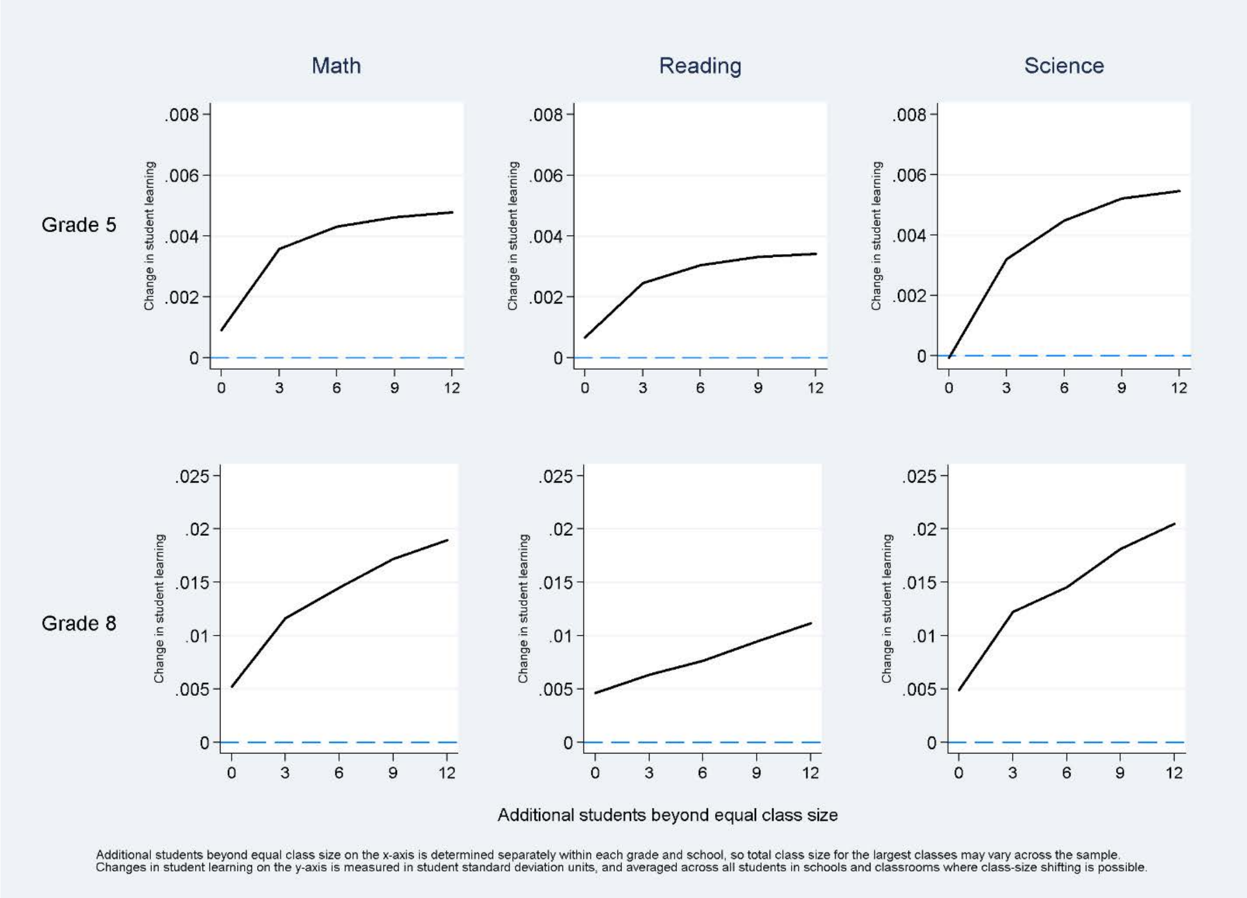
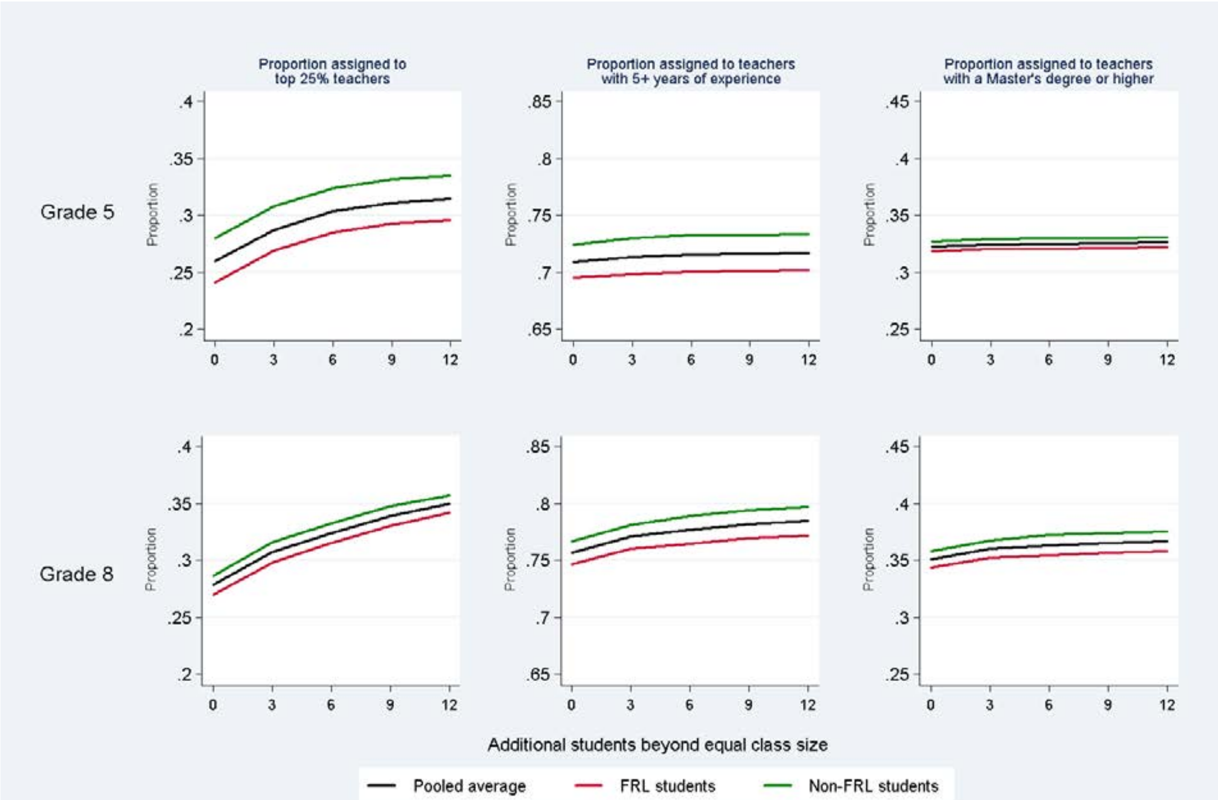
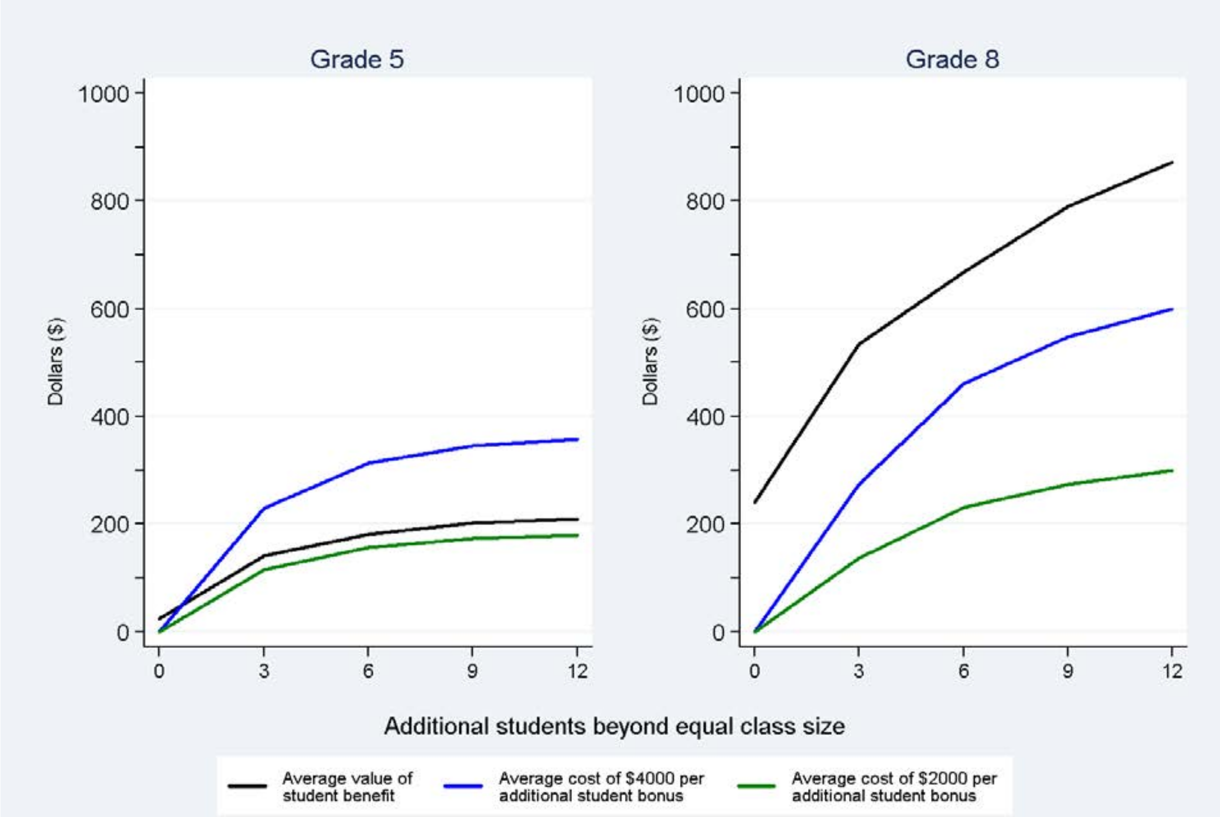


Figure 2. Shifting Strategies Increase Access to Effective Teachers, Though a Gap Persists



Additional students beyond equal class size on the x-axis is determined separately within each grade and school, so total class size for the largest classes may vary across the sample. The proportion measures describe the proportion of students assigned to teachers with the given characteristics. These proportions are calculated across all students, and on FRL and non-FRL student subgroups.

Figure 3. Benefits to Students and Costs for Teacher Bonuses



Additional students beyond equal class size on the x-axis is determined separately within each grade and school, so total class size for the largest classes may vary across the sample. The benefit to student line represents the average per-student net present value of future wages at age 28 as a result of the class-size shifting policy. The two cost lines represent the average per-student cost associated with compensating high-performing teachers for each additional student they teach beyond equal class size.