

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

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

by Michael Hansen

CALDER at the American Institutes for Research

Foreword by Michael J. Petrilli and Amber M. Northern

November 2013

Contents

Foreword	2
Acknowledgments	6
Executive Summary	7
Introduction	11
Sidebar: A Balancing Act	14
Data and Methods	16
Results	18
Cost Consideration	26
Sidebar: Feasibility Issues	26
Conclusion	31
Technical Appendix	34
Optimization Problem and Optimality Conditions	34
Detailed Method	36
Endnotes	45

The Thomas B. Fordham Institute is the nation's leader in advancing educational excellence for every child through quality research, analysis, and commentary, as well as on-the-ground action and advocacy in Ohio. It is affiliated with the Thomas B. Fordham Foundation, and this publication is a joint project of the Foundation and the Institute. For further information, please visit our website at www.edexcellence.net or write to the Institute at 1016 16th St. NW, 8th Floor, Washington, D.C. 20036. The Institute is neither connected with nor sponsored by Fordham University.

Foreword

by Michael J. Petrilli and Amber M. Northern

Last month, *USA Today* reported that officials in the Brevard County Schools had broken Florida state law—on purpose. Their offense? Placing more kids in classrooms than Florida’s Class Size Reduction statute allows. Officials had done the math and decided that complying with state policy would cost more than the penalty they’d pay for adding a handful of students to each classroom. The estimated fines totaled roughly \$170,000, which paled in comparison to the cost of the teachers that the district would have to hire to comply with the size-limiting mandate.

Yet it’s unclear how Brevard chose to allocate these additional students. Did administrators give *every* teacher more students in equal shares? Did they apportion shares to seasoned veterans or, more likely, to seniority-deprived new teachers? Maybe they drew straws?

But what if Brevard officials had chosen another option? What if they had assigned the “extra” students to their *most effective* teachers, leaving fewer pupils in classrooms presided over by weaker instructors? What would be the impact of such a practice on student achievement?

That’s the scenario that this empirical paper models. The idea is straightforward: Give the better teachers more kids and the weaker teachers fewer—then see what happens. It’s a common-sense option with many supporters. We know, for instance, that parents say they would opt for larger classes taught by excellent teachers, rather than smaller classes with instructors of unknown ability. In a study last year for the Fordham Institute, the FDR Group found that a whopping 73 percent of parents would choose a class with twenty-seven students—provided it is “taught by one of the district’s best performing teachers”—over a class of twenty-two students “taught by a randomly chosen teacher.” Further, given the choice between fewer students and more compensation, the teachers themselves choose the latter. In a well-done study of their own, Dan Goldhaber and colleagues found that 83 percent of educators in Washington State would prefer an additional \$5,000 in compensation versus having two fewer students in their classes.

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

Others in the academic and policy worlds have made similar arguments. Rick Hess has called for allowing “gold-star” teachers—those who post larger-than-average gains for two consecutive years—to teach up to 50 percent more students and be financially rewarded for doing so.

Likewise, analysts at Public Impact have offered up several strategies for “extending the reach” of effective teachers—one strategy is assigning more students to them. Marguerite Roza supports the idea, too; she found that districts could save significant sums by simply allowing class size to inch up by two pupils.

Yet, to our knowledge, no district assigns students to teachers based on their instructional effectiveness. Instead, pupils are divided roughly equally among teachers of the same grade in the same school, since parceling them out uniformly is viewed as fair to teachers.

But what if it’s not fair for kids? Or what if the costs fail to justify the benefits? We aimed to find out.

Given district aversion to assigning students in this way, we were forced to “simulate” such assignment using actual data from one state (North Carolina). To perform this statistical feat, we approached economist Michael Hansen, a senior researcher at the American Institutes for Research. Dr. Hansen, an expert in labor economics and the economics of education, has ample experience mining North Carolina data and conducting simulations of this genre. (He’s also an original member of Fordham and AEI’s [Emerging Education Policy Scholars](#), a group of highly able, young education-policy scholars.)

Hansen starts by examining the extent to which North Carolina already assigns students within schools based on teacher effectiveness. (He finds the state has a slight tendency to do so.) Then, he turns to the simulation, looking at fifth and eighth-grade test scores. He uses three years of data (2007–10) to generate past value-added measures. For the fourth year, he estimates how teachers actually performed, and then he simulates what the impact would have been if students instead had been allocated to teachers based on their prior performance, with an eye towards maximizing student gains. The allocation process results in larger classes for the most effective teachers and smaller for the least effective.

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

The key finding: Minor changes in assignment lead to improvements in student learning. The results were relatively modest for the fifth grade; there, even the most intense reallocation practice that Hansen simulated, assigning as many as twelve additional pupils to effective teachers, yielded gains equivalent to extending the school year by just two days.

At the eighth-grade level, however, the results were much more robust. Hansen found that assigning up to twelve more students than average to effective eighth-grade teachers can produce gains equivalent to adding two-and-a-half extra weeks of school. Yet adding fewer students pays dividends, too. In fact, 75 percent of the potential gain from allowing up to twelve students to be assigned to the best teachers' classes is already realized when allowing just six students to move. Specifically, adding up to six more than the school's average produces math and science gains akin to extending the school year by nearly two weeks. This impact is the equivalent of removing the lowest-performing 5 percent of teachers from the classroom.

That last point is worth reflection. Moving a handful of students to the most effective eighth-grade teachers is comparable to the gains we'd see by removing the lowest 5 percent of teachers. And that is *without actually removing them*. As Hansen explains, "Class-size shifting enables the lowest-performing teachers to become more effective than they may be otherwise." That's certainly a good thing. But does it mean that we should hang onto persistently ineffective teachers? Given the cost of keeping them on the payroll, probably not. At some point, giving ineffective teachers the luxury of small classes becomes an unsustainable financial burden. Or, to put it another way, we should shrink some teachers' classes down to zero students—and take the money saved thereby to bump up the compensation of effective teachers.

Last, Hansen examines whether this reallocation policy helps our neediest students gain more access to effective teachers. In a word: no. Gaps in access for economically disadvantaged students persist, primarily because the pool of available teachers in high-poverty schools remains unchanged under this strategy. Hence, this policy alone won't remedy achievement gaps. (Recall that the reassignment occurred *within* schools; if it had been carried out *across* schools, perhaps the results would differ.)

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

As for costs, Hansen shows that some class-size variation already exists within schools (a differential of three to five students); presumably these small differences are not compensated. Perhaps then, principals could choose to assign these extra students to their most effective teachers without costing taxpayers an extra penny. After all, that's the beauty of this strategy: It does not require a change in state policy or, in many cases, teacher contracts to make it happen.

Further, many contracts already allow extra pay for teaching more kids outside of the class-size caps. The twist here is simply to choose those teachers based on effectiveness. This small change in practice promotes performance-based pay without calling it that. This paper suggests some ways that existing funds could be repurposed to meet increased workloads for teachers. But Hansen shows that even if entirely new funds were needed to compensate teachers for participating, the strategy is still worth it, at least under certain assumptions. In short, the benefits of having effective teachers translate into higher future wages for students that outweigh the costs of paying teachers more in the near term.

In the end, one simple change—giving effective teachers a handful more students—could mean a big boost to student achievement. And it does not require that we let go of our least effective teachers, some of whom will improve their instruction with fewer pupils. For those who don't, we should be far less timid about letting them go, since doing so becomes a no-cost way of paying great teachers something closer to what they deserve.

Still, this “simple change” has not been tested in the real world. We've now simulated its impact using actual data from an actual state. But which district or state will be the first to try it out in real classrooms?

Acknowledgments

Generous support for this project was provided by the Searle Freedom Trust, as well as our sister organization, the Thomas B. Fordham Foundation.

We also thank the individuals who conducted and reviewed the research. First and foremost, thanks to Michael Hansen, senior researcher at American Institutes for Research (AIR), who took our simple idea for a research study and gave it sturdy empirical legs. He worked quickly, efficiently, and under a tight budget; we are grateful. Tiffany Chu, research assistant at AIR, provided excellent research support. Thanks also to these individuals for helpful comments on the study design and/or draft report: Nathan Barrett, senior research associate at the University of North Carolina at Chapel Hill; Bryan Hassel, codirector of Public Impact; and Dan Goldhaber, director of the Center for Education Data & Research at the University of Washington Bothell.

On the Fordham side, we extend thanks to Chester Finn, Jr. for his input on earlier drafts. Matt Richmond oversaw production, while Michelle Gininger and Joe Portnoy managed dissemination. Anne Himmelfarb served as copyeditor and Alton Creative, Inc. designed our layout.

Executive Summary

Public schooling in America suffers from a triple problem that a single policy solution might solve: 1) Our best teachers aren't paid enough, 2) not enough kids benefit from great teachers, and 3) too many are stuck with weak teachers. This paper describes—and demonstrates the value of—a change in policy that could address all three issues at once, and could be done at no additional cost to taxpayers. Following this route, however, means reversing position on a widely popular—but pricey and none too effective—approach to “educational improvement”: class size reduction. Instead of trying to keep classes small, we should be leveraging our existing teacher talent by *enlarging* the classes taught by our best instructors—and compensating these excellent teachers for the extra work involved.

Playing to our classroom strengths by redistributing students is not a new idea. Both Secretary of Education Arne Duncan and Bill Gates have encouraged schools to try it to soften the blow of recent budget cuts. The rationale is straightforward: We know from rigorous research that the impact on student achievement that comes from having a good versus average teacher clearly trumps the effect of smaller class sizes.¹ 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? No district appears to have purposefully allocated students in this manner. To answer this question in the absence of real-world implementation, this study used North Carolina data to simulate student outcomes under the assumption that schools strategically assign larger classes to the strongest teachers and smaller classes to the weakest.

The study poses a number of questions. Specifically:

1. To what extent do existing class-size assignments already reveal differences in teacher performance?

Using the North Carolina data, the study documents how students are actually assigned to teachers based on their prior performance across fifth- and eighth-grade classrooms. There is a

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

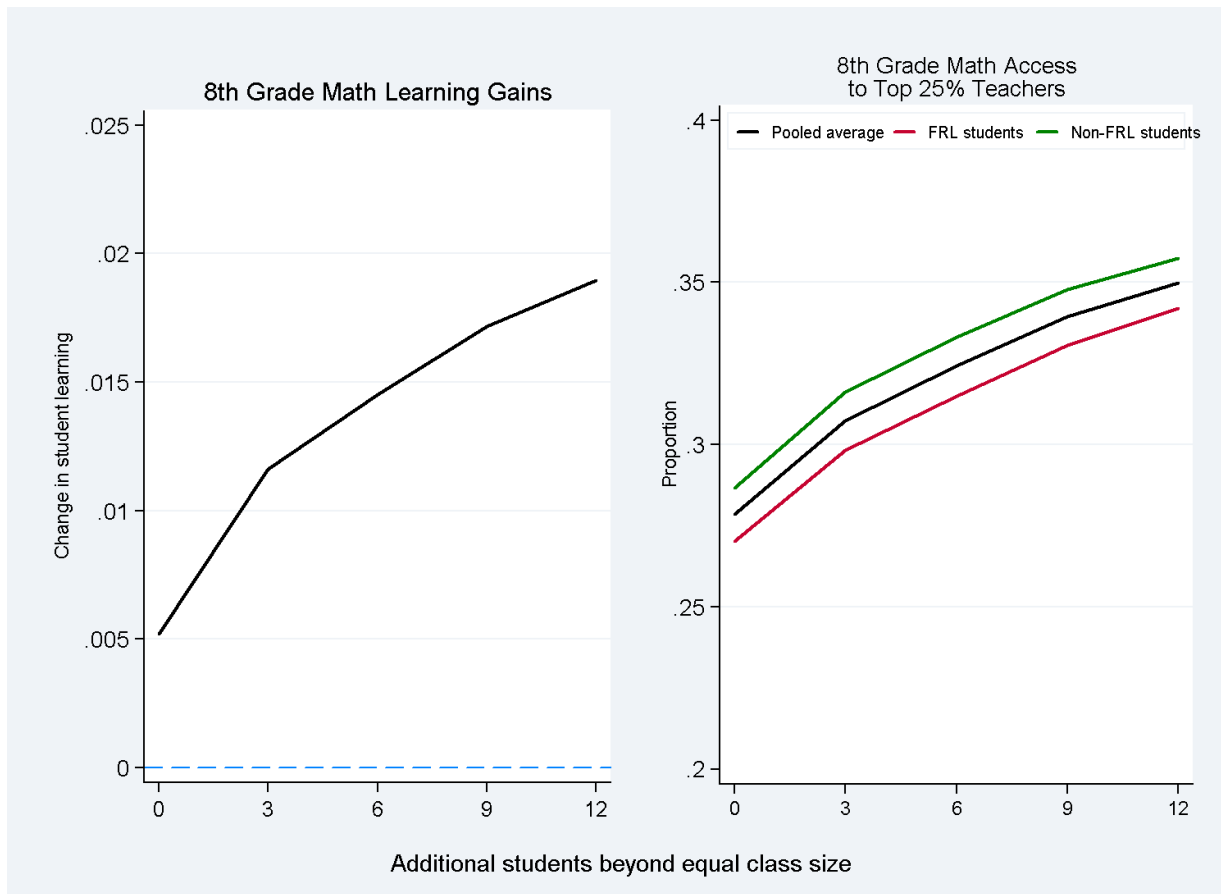
very slight tendency to place more students in the classes of effective teachers; but still only about 25 percent of students are taught by the top 25 percent of teachers. Thus North Carolina appears to differentiate class sizes based on teachers' performance only to a very limited extent.

2. To what extent does a purposeful policy of allocating students to teachers on the basis of teacher effectiveness boost achievement?

The results show that relatively minor changes in the way that students are assigned to teachers can lead to significant learning gains. But the size of these gains depends on grade level, with eighth grade more promising than fifth grade. 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 (see figure ES-1). 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. (And these class-size differences are, incidentally, similar to those actually observed in the North Carolina data set.)

The potential impacts on learning are more modest in fifth grade, where the large majority of teachers are in self-contained classrooms.² Even the most intensive shifts in fifth-grade class size—in which up to twelve additional students may be allocated to effective teachers—are equivalent to extending the school year by just two days.

Figure ES-1. Shifting Students to More Effective Teachers Can Increase Learning and Access to Effective Instruction



How to read this figure: The simulations incrementally allocate more and more students (above the average) into the classrooms of the most effective teachers in a school. These two graphs show potential learning gains in eighth-grade math arising from these changes in class size (on the left) and the proportion of eighth-grade students assigned to teachers in the top 25 percent of value-added performance (on the right).

3. *To what extent do these simulated classrooms affect students' access to effective teachers?*

Results show that the class-size-shifting 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—

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

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.

As for costs, these depend, of course, on what is done with weak teachers and how much is added to the salaries of strong teachers, but the changes *can* be feasibly made at no net cost to taxpayers (i.e., within current budgets). The study concludes by considering these issues, and offers rough calculations of long-run benefits.

Can student learning improve by simply reallocating the distribution of students among teachers? The results of this investigation say yes, though the extent of the achievement gain—and the price tag—depend on grade level and a suite of human-resource policy decisions. Perhaps most important, this analysis makes clear that universally shrinking class sizes may be counterproductive in terms of pupil achievement—and that purposefully expanding some of them can be more efficacious.

Introduction

Being assigned to the classroom of a great teacher can pay dividends for students fortunate enough to find themselves there, including greater academic achievement, improved odds of going to college, and higher future wages.³ So why do schools continue to assign the top 25 percent of instructors only about 25 percent of students?⁴ Why not enable more pupils to benefit from their effectiveness?

The answer is as straightforward as it is banal: American public education has long determined class size by dividing all students in a given grade in a given school approximately equally among all the teachers of that grade in that school. When the numbers grow uneven, students or teachers are moved in order to even them out again. Moreover, such “equality” is often enshrined in collective bargaining agreements and further shaped by district- or statewide class-size limits.

Unfortunately, this established practice doesn’t work very well for students. Given what we now know about the benefits students derive from having a high-quality teacher, uniform class sizes are not the most effective way to deploy our most valued instructors. Large differences in productivity across teachers imply that gains in student learning could be realized through simple, strategic reassignments of students across teachers, and by allowing—even intentionally causing—class sizes to vary.

In this study I dig into that possibility by investigating how student achievement would be affected if schools were to double down on their most effective teachers by shifting a few more students into their classrooms. It’s a simple idea that is gaining attention, in part because the Great Recession has forced states and school districts to reduce costs while attempting to hold student learning steady. High-profile advocates of the policy include U.S. Secretary of Education Arne Duncan and Bill Gates, who separately have urged strategic classroom assignments in which more students are assigned to the most effective teachers.⁵ Frederick Hess of the American Enterprise Institute and coauthor Olivia Meeks have argued for a policy identifying “gold-star” teachers and rewarding them with (among other things) the opportunity to teach larger classes for a larger salary.⁶ And the Opportunity Culture initiative, launched in 2011 by

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

the respected education research group Public Impact, has as its goal an increase in student exposure to the top 25 percent of teachers.⁷

At the same time, copious research on teacher quality over the last decade has overtaken the class-size debates of prior years. 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.⁸ Yet, the differences in teacher productivity across classrooms are large and important predictors of student learning and other outcomes. The magnitude of these differences *swamps* the expected effect of smaller classes,⁹ suggesting that a policy focus on teacher quality will result in a larger improvement in student performance in the nation's public schools than a focus on class size. At the same time, the two approaches need not be understood as in opposition to one another. This paper combines the two and posits that student outcomes can improve when class sizes are determined strategically for teachers based on their 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. To answer this question in the absence of real-world implementation, I use North Carolina data from fifth- and eighth-grade classrooms to simulate student outcomes under the assumption that schools strategically assign larger classes to the strongest teachers and smaller classes to the weakest. Specifically, I ask:

- To what extent do existing class-size assignments already reveal differences in teacher performance?
- To what extent does a purposeful policy of allocating students to teachers on the basis of teacher effectiveness boost achievement?

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

- To what extent does allocating students to teachers on the basis of teacher effectiveness influence students' access to quality teaching?

The cost of this policy is also an important consideration and will be discussed at the end.

Why Doubling Down on Teacher Performance Makes Sense

Research has shown that teachers' effectiveness varies widely, and that this variation accounts for a greater share of the differences in student learning than any other known education intervention.¹⁰ Studies have also shown that traditional measures of teacher preparation, such as licensure and education, are poor predictors of teachers' classroom performance and are thus ineffective strategies in boosting work-force quality.¹¹ State and district policies have begun to shift in response to these findings. For example, thirty-six states have revamped their teacher and principal evaluation systems since 2009, and most of these states have adopted value-added estimates—that is, teacher-productivity measures based on student gains on standardized tests—as a key component of these evaluation systems.¹²

Opinions vary on how these new teacher performance data can be most effectively used to improve student achievement. Proposals to use performance data directly—including pay-for-performance bonuses or selective retention of teachers for tenure—are abundant, but unpopular among teachers.¹³ However, using performance data *indirectly* for the same purpose may be more politically palatable. Strategically assigning larger classes to high-performing teachers is one such indirect mechanism. Yet it requires that we let go of our fascination with universally smaller class sizes. Yes, they're widely popular. And yes, there is some evidence that they boost student achievement. But 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 been demonstrated only in the earliest grades.¹⁴ For most students above the third grade, the evidence points to *at most* a small class-size effect, if any at all.¹⁵ (Using the North Carolina data, I likewise estimate small class-size effects in fifth and eighth grades.)¹⁶ Thus in effect, it would take an increase of at least ten to twenty additional students in a good teacher's class to dilute his productivity to that of an average teacher.¹⁷ Put another way, assigning a few extra students to

the class of an effective teacher can translate to big gains for these students, while making only 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 work-force quality in its implicit demand to increase the quantity of teachers.¹⁸ Bottom line: There's little reason to hold onto universally smaller class sizes when there are smarter ways to assign classes to teachers who vary in their classroom effectiveness.

The class-size-shifting strategy simulated here considers the assignment of students to classes as a "learning optimization" problem, asking how class sizes would be assigned to different teachers if total achievements across all students were the objective.¹⁹ In practical terms, 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 equal-sized classrooms. The "shifted" students benefit from being reassigned to a better teacher, and their gain exceeds the "penalty" imposed on other pupils already in that classroom who now have a slightly larger class (for more, see the "Balancing Act" sidebar). What's more, the remaining students in the less-effective teacher's class receive a "benefit" because their class becomes smaller. Depending on how many students are allowed to shift between classes, the resulting class sizes become more imbalanced, with the most effective teachers leading the largest classes and the weakest the smallest. In theory at least, this disproportionate alignment of students should translate to greater learning gains overall as more students receive high-quality instruction, and weaker teachers receive small classes that are more conducive to increasing their performance as well. How well this theory translates into practice will be explored in the simulation below.

[SIDEBAR]

A Balancing Act

Naturally, bigger classes for the best teachers are not always better, even if there were no constraints on how students were assigned to classrooms. Achieving an optimal class size requires the balancing of three variables: the distribution of teachers' performance in the school, the class-size "penalty" incurred among students whose classes get larger, and the class-size "benefit" enjoyed by students whose classes get smaller. Of these three components, it is the

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

distribution of teacher performance that plays the primary role in achieving the optimal class size; in most schools, the direct benefit to students shifted to a better teacher tends to be many times larger than class-size effects. The net class-size effects, moreover, are even smaller than one may think: Shifting students between classes implies some larger classes and some smaller classes, and if only a few students are shifted, the benefits to the small classes virtually cancel out the penalties to the larger classes.²⁰

[/SIDEBAR]

Data and Methods

The primary data source is longitudinal data from North Carolina, spanning grades 4 through 8 over a period of four school years (2007–08 through 2010–11). The first three years of the data are used to estimate teachers' effectiveness based on a value-added measure across all grades; the fourth is the target year used to simulate the changes from implementing the class-size-shifting strategy in grades 5 and 8. The administrative files from North Carolina include various student and teacher characteristics, which are used to predict student achievement and teacher performance in that target year. The standardized test scores are extracted from the criterion-referenced End-of-Grade (EOG) tests in math, reading, and science administered as part of the state's ABCs accountability system.²¹ Six distinct samples are analyzed, corresponding to the unique subject-grade combinations in the data for grades 5 and 8 (three subjects across two grades).²² Within these six samples, schools with multiple teachers instructing the same subject and grade—necessary conditions for carrying out this shifting strategy—are flagged, and these are the subjects of the simulation.²³

The simulation itself has two major components.²⁴ First, estimates are generated for teachers' past and present value-added performance, along with coefficient estimates of class size, classroom composition, and teacher experience—all of which will be utilized in the next step. The second component involves manipulating the classrooms of teachers observed in the 2010–11 school year by reallocating students based in part on teachers' prior value-added measures. In short, this simulated pupil-assignment process results in larger classes for the most effective teachers, smaller classes for the least effective. The main outcomes of the simulation are then calculated based on these new class assignments, using the teacher's value-added measure for the 2010–11 school year. In other words, students are shifted across teachers' classrooms based on teachers' *expected* performance²⁵ while the simulated outcomes are based on their *realized* performance.

This classroom manipulation is carried out at varying degrees of “intensity,” which represent the number of additional students reallocated into the highest-performing teacher's class (relative to the number that would be otherwise expected with equal-sized classes).²⁶ These levels range from zero additional students (in other words, class sizes are equal across teachers in a school) to

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

twelve additional students (e.g., the best teacher is assigned up to thirty-two students, if the equal class sizes were expected to be twenty). Keep in mind that the additional students are being shifted from other classes in the same school and grade—meaning that (again, given a normal class of twenty students) a weaker coworker may be teaching as few as eight students at the twelve-additional-student level. At each intensity level of class-size shifting, two types of outcomes are calculated: mean change in student learning (by subject) and the proportion of students assigned to teachers with various characteristics (teachers in the top 25 percent of prior value-added estimates, teachers with five or more years of experience, and teachers with a master’s degree or higher). For more details on the methods, see the appendix.

Results

Finding 1: Effective teachers in North Carolina already appear to have very slightly larger classes.

Before presenting the results of the simulation, let's look at the patterns of assignment actually observed in the North Carolina data, which will be a point of reference for the simulation results that follow. This is an important first step—if class sizes are already divvied up according to effectiveness, then there is little to gain from changing the status quo.

Table 1 illustrates the baseline condition of class-size assignments with respect to prior teacher performance in the 2010–11 school year. Three points are noteworthy. First, North Carolina 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 the school. For example, a value of 2.738 for fifth-grade math means that large classes in a school had about three more students than the mean, while smaller classes in the same school—with teachers instructing an equivalent course—had three fewer students. In other words, using the average fifth-grade math class size of about twenty-two students as a reference, a school may typically have one larger class with twenty-five students and another smaller class with nineteen students. I find that average class-size deviation is consistently larger in eighth grade (ranging from 3.8 to 5.7) than in fifth grade (1.7 to 3.0).²⁷

Second, the within-school relationship between expected teacher performance and class size (row 3) is slightly positive across all of the samples. (A value of zero would mean that, in a single school, there is no relationship between a teacher's effectiveness and her class size, whereas a positive value means a more effective teacher would be more likely to have a larger class.) The correlation coefficients range from 0.01 to 0.13.²⁸ This suggests that some strategic assignment of students to more effective teachers may already be occurring to a very limited extent. A helpful way of conceptualizing this possibility is to think of a single leftover student. That is, if classes are equally assigned to all teachers at the beginning of the year and then a single student arrives on the first day of school, who should he be assigned to? If schools want to maximize achievement, it's obvious that this singleton should go to the best teacher—someone

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

has to accept the larger class size, so the student may as well be placed where he is likely to be best-served. Row 4 of the table shows that in practice, fewer than half of these students are actually assigned to teachers in the top 25 percent. By way of comparison, row 5 shows only slightly smaller proportions going to teachers identified in the bottom 25 percent (eighth-grade math and reading students are slightly more likely to be assigned to these teachers). So, once again, while the data show a small tendency to assign larger classes to better teachers, it is indeed slight.

Table 1. Snapshot of Observed Class-Size Assignment in North Carolina, 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. Percentage of leftover students assigned to teachers in top 25 percent	24.3%	24.9%	25.9%	30.3%	28.8%	38.0%
5. Percentage of leftover students assigned to teachers in bottom 25 percent	22.8%	22.1%	24.6%	31.8%	29.6%	36.9%
6. Percentage of students assigned to teachers in top 25 percent in prior value added	25.8%	28.7%	23.7%	25.1%	24.4%	25.4%
7. Percentage of FRL students assigned to top 25 percent teachers	23.5%	26.0%	21.7%	23.2%	24.3%	22.6%

Note: Statistics derived from student-teacher assignments based on 2010–11 school year. FRL = free and reduced-price lunch.

How to read this table: Each column represents one of the six samples (corresponding to each grade and subject combination). Measures representing current class-size assignments (across the rows) are calculated in each of the samples. In row 6, for example, 25.8 percent of students in the fifth-grade math sample are assigned to teachers in the top 25 percent based on prior value-added estimates.

And finally, the percentage of students assigned to teachers in the top 25 percent of effectiveness ranges from about 24 percent to 29 percent (row 6).²⁹ However, the percentage of students who are eligible for federal free or reduced-price lunch (FRL) assigned to top-quartile teachers (in row 7) ranges from 22 percent to 26 percent, roughly two percentage points lower across most of the six samples,³⁰ and revealing a gap in access to effective teaching for economically

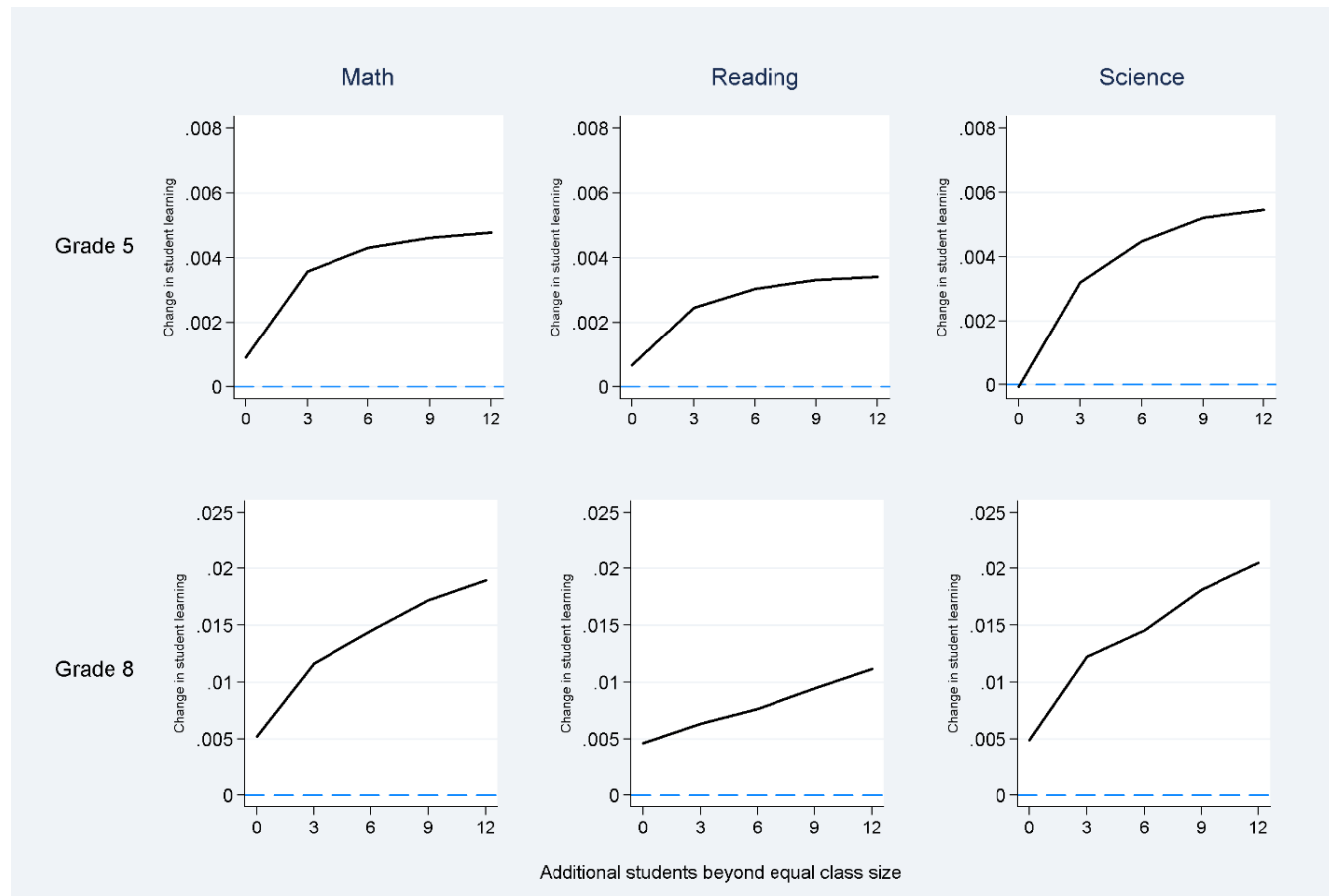
disadvantaged students. This gap means that disadvantaged students tend to get relatively weaker teachers, which reinforces preexisting achievement gaps.³¹ Note that these gaps in access may arise in two ways: 1) due to the uneven distribution of effective teachers across schools (high-need schools tend to have a harder time attracting the most effective teachers), and 2) due to disadvantaged students being disproportionately assigned to the weakest teachers in a school. The class-size-shifting strategy analyzed here addresses *only* problems of the latter type by directly increasing student access to the most effective teachers within a school.

Finding 2: Simulated student achievement improves overall with class-size shifting, with larger gains computed in eighth grade than in fifth.

The simulated gains in average student learning are presented in Figure 1, which includes six graphs—one for each tested subject (across columns) in both fifth and eighth 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 additional students allowed in an effective teacher’s classroom; recall the value of zero represents equal-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. This is an important result: *as the best teachers teach larger classes and the weakest teachers have progressively smaller ones, the net result is improved student learning*. Note, however, that the magnitudes of the student learning gains on the vertical axis differ across grades, with greater gains in the eighth-grade than in the fifth-grade estimates.

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



Note: The number of 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. Change in student learning on the y-axis is measured in student standard deviation units, and averaged across all students in schools and classrooms where class-size shifting is possible.

How to read this figure: Each of the six graphs represents the simulated learning gains (on the vertical axis) for students in a given grade (row) and subject (column). The simulations incrementally allow the most effective teachers to have additional students shifted into their classrooms (on the horizontal axis), relative to equal-sized classrooms. The patterns show students learn more overall as more students are allowed to shift into effective teachers' classrooms.

An interesting pattern is the shape of these learning gains—they are not linear but 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 stemming from (or associated with) the policy. Allowing even more students to be shifted across classes still results in net student learning gains, but these gains are smaller than those expected from the

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

initial set of shifted students. In fact, across all grades and subjects (with the exception of eighth-grade reading), over 75 percent of the potential gain from assigning up to twelve students to the best teachers' classes is already realized when just six students are shifted. Hence, even shifting small numbers of students may generate the bulk of the potential gain from this strategy. That's an important selling point for strategic allocation—both in terms of getting teachers to agree to take on more students and in compensating them for doing so (more on the latter below). However, districts that are interested in pursuing this strategy more aggressively could continue to add more students beyond six to achieve the highest total gains.

Still, casual observation of these figures may lead one to conclude that, because the gains appear small, they may not be worth the hassle. Not so! While magnitudes of simulated gains are indeed modest, these are *average* learning gains across all students, and don't apply just to those who are moved. In practice, the gains are accrued primarily to individual students who are shifted into stronger teachers' classrooms. Yet even apparently modest average gains can be consequential in the aggregate. In the eighth-grade estimates, the potential gains of 0.02 standard deviations in math and science, with as many as twelve additional students, amount to an additional 2.5 weeks of schooling.³² While reassigning twelve students to the best teacher is likely not always feasible, even six additional students can result in gains of 0.015 standard deviations, which is equivalent to almost two weeks of additional schooling in eighth-grade math and science. Similar levels of student shifting translate to nearly a week of additional schooling in eighth-grade reading. These are considerable increases obtained by simply reorganizing classes in ways that don't differ much from current practice (North Carolina classes already show disparities in class size near this level—principals just need to be more strategic about assigning students.) In the fifth-grade estimates, the potential gains from this strategy are more modest than in those for eighth grade. Allowing up to twelve additional fifth-grade students for the best teachers can produce math and science achievement gains of 0.005 standard deviations, which equates to roughly two additional days of schooling.

Why such a large difference between grades? I speculate that the difference is driven by the self-contained classrooms typical of elementary education, in which students are instructed by the same teacher in all subjects. This self-contained approach has two consequences. First, value-

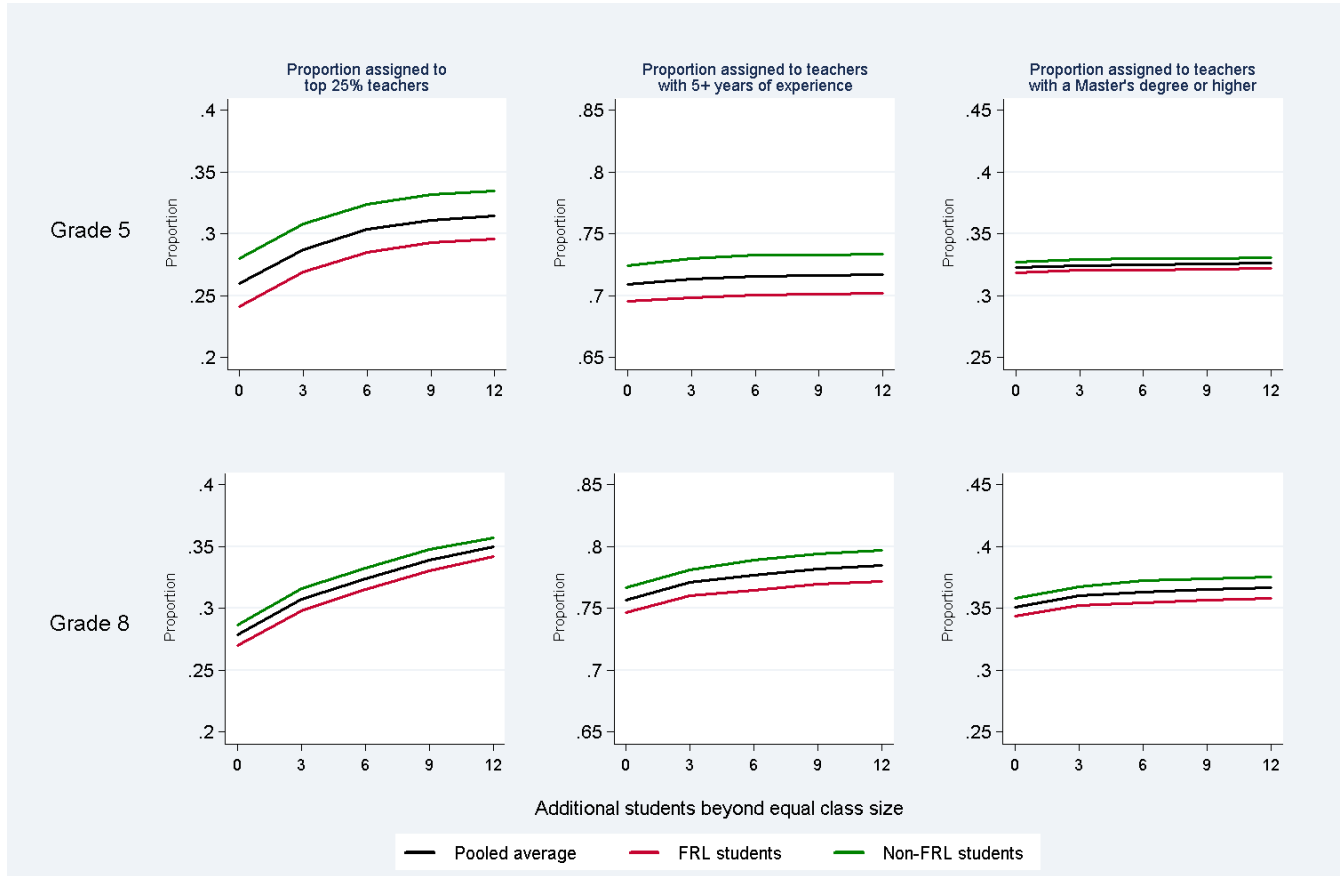
added estimates from a single class are less accurate than those from departmentalized grades (which are based on multiple classes), so a principal's prediction of an elementary teacher's performance (if it relies *only* on value-added measures) is going to be less reliable. Second, we know that teacher performance is positively correlated across subjects,³³ 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 and may therefore be able to realize greater gains than what are simulated here, but this is speculative.

Finding 3: Students gain more access to effective teachers as a result of class-size shifting, though a gap in access persists.

The net gains in student learning are a direct result of placing more students in effective teachers' classrooms. 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 percent of effectiveness (first column), teachers with five or more years of experience (second column), and teachers holding a master's degree or higher (third column).³⁴ Each graph includes three lines: one showing an overall level of exposure, one showing access by students who *are* eligible for free or reduced-price lunch (a common indicator of economic disadvantage), and one showing access by students who are *not* eligible.³⁵

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. Note that these metrics do not show the same rate of increase—the proportions climb most quickly for teachers in the top 25 percent, 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 the other two characteristics are only weakly correlated with it.

Figure 2. Shifting Students Increases Access to Effective Teachers, Though a Gap Persists



Note: The number of 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 percentage measures describe the share of students assigned to teachers with the given characteristics. These percentages are calculated across all students, and for FRL and non-FRL student subgroups.

How to read this figure: Each of the six graphs represents the proportion of students assigned to specific teachers (on the vertical axis) within a given grade (row) and having a specific characteristic (column). The simulations incrementally allow the most effective teachers to have additional students shifted into their classrooms (on the horizontal axis), relative to equal-sized classrooms. Three trend lines are shown, representing an average that pools all students together, and then separates students by FRL status. The patterns show that students increase their likelihood of being assigned to teachers with all three of these characteristics as more students are allowed to shift into effective teachers' classrooms. Yet a gap in access for FRL and non-FRL students persists regardless of how many students are allowed to shift between classes.

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

Yet look beyond the overall trend 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 three measures of teacher quality. Thus class-size shifting does not appear to mitigate the relative gap in access separating disadvantaged and non-disadvantaged students, although it does improve disadvantaged students' 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 must be due to the fact that effective teachers are not equally allocated across schools. Consequently, this *class-size-shifting strategy alone cannot reduce preexisting inequalities, and some other intervention would be necessary to remediate entirely gaps in students' access to the best teachers.*

Cost Consideration

The feasibility of this class-size-shifting strategy rests on a number of practical and policy-related issues (see the “Feasibility Issues” sidebar). Yet the key question is whether teachers are willing to participate, which brings us to compensation issues, which also brings us to the question of whether such a strategy would stretch already-strained budgets.

[SIDEBAR]

Feasibility Issues

There are several feasibility issues that need to be addressed before schools could begin implementing strategic class-size assignment.

First, there are practical constraints to implementation. Strategic shifting has two necessary conditions: 1) Schools must have multiple instructors teaching equivalent content to separate groups of students, and 2) principals must have relevant performance data for determining which teachers are better, and by how much. If these conditions are not met, implementation may be tricky or impossible. For example, some schools engage in tracking students by ability level, thus qualitatively differentiating otherwise similar classes; this practice complicates the shifting strategy because it makes it harder to move students from one class to another. Similarly, the strategy could be severely undermined by data problems, such as a lack of clear performance categories for teachers.

Second, there are state laws, district policies, and bargaining agreements that may stand in the way. Some of these already contain compensation clauses for teachers with class-size overages, which could encourage administrators to be more strategic in determining who gets the larger class. But they could also entirely prevent implementation of this strategy where class sizes are already at or near limits. In most other circumstances, though, schools could probably adopt an informal version of the strategy, shifting a few students across classrooms without any major policy change. More aggressive levels of shifting will require changes to policy, collective bargaining agreements, or both in some locales.

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

Finally, the gains portrayed here represent averages, and are not guaranteed for any given school. This uncertainty makes strategic assignment a risky proposition, because it will realize greater gains in some places than others. 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 it could be a dud—for example, the most effective teacher may have a bad year, 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 (compared to equal class sizes). However, on average, the overall gains for students are expected to be positive.

[/Sidebar]

Not all teachers would eagerly accept additional pupils, but many would be open to the prospect if it meant higher pay. For instance, a teacher compensation survey conducted in Washington State in 2006 found that 83 percent of educators would prefer an additional \$5,000 in compensation to the alternative of having two fewer students in their classes.³⁶ 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 (a three-student reduction). Forty-two percent of 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.³⁷ So depending on current class sizes, a proposal to add students for the most effective teachers could be met with reasonably widespread enthusiasm—or not.

Is extra compensation for teachers necessary if this strategy is to be implemented? Maybe not. Schools could implement limited class-size shifting, assigning only a few additional students to the highest-performing teachers, without having to offer additional pay. As shown in Table 1, some naturally occurring class-size variation already exists within schools—on the order of three additional students in fifth grade and five in eighth. Presumably most of these class-size

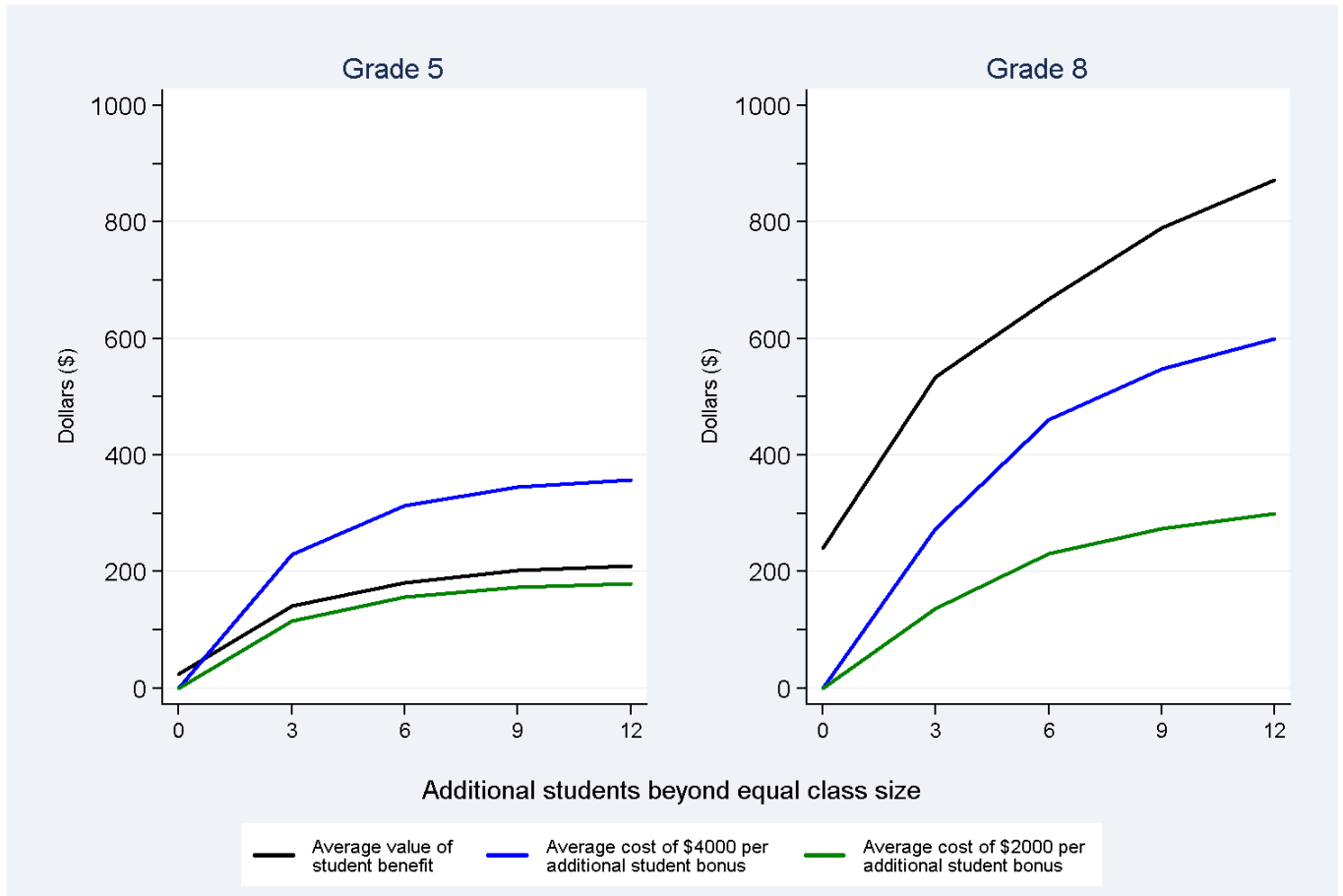
EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

differences are not compensated. So if principals operated a little more strategically and intentionally gave three to six more students to the best teachers, their schools could experience an even higher share of the potential gains under the policy—without additional costs. It is also possible to make a large class a more desirable assignment by providing nonmonetary rewards, such as recognizing teachers as “lead teachers” for the grade or subject, giving them first choice of nonteaching assignments, allowing increased time for lesson preparation, reducing adjunct duties, or providing them increased access to teacher aides.

In fact, though, part of the appeal of this strategy is that it is a way of paying outstanding teachers more—under the cover of giving them more students. It also suggests a path around the constraints of contracts and salary schedules since it is readily defended as “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 a policy like this, dollars could be repurposed 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 (sure to be unpopular!). Finally, money could be found in some schools if dwindling class sizes for the least effective teachers compelled principals simply to let them go and not replace them; average class sizes would then begin to creep up, but with strategically determined class sizes, this would not need to imply a loss in learning.³⁸

Even if entirely new funds were needed to compensate teachers for participating, the strategy might still be worth it—at least in eighth grade. A better teacher translates to better student outcomes in the long run, which could give the idea traction among parents and voters. 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).³⁹

Figure 3: Benefits to Students Could Outweigh Costs of Teacher Bonuses



Source: Author’s calculations based on Raj Chetty, John Friedman, and Jonah Rockoff, “The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood” (National Bureau of Economic Research Working Paper 17699, Cambridge, MA, 2011).

Note: The number of 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 “average value of student benefit” line represents the average per-student net present value of future wages at age twenty-eight 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 size.

How to read this figure: The two graphs represent the average costs and benefits associated with the class-size-shifting strategy in grade 5 (on the left) and grade 8 (on the right). The simulations incrementally allow the most effective teachers to have additional students shifted into their classrooms (on the horizontal axis), relative to equal-sized classrooms. Three trend lines are shown, representing 1) the average value of the student benefit of the class-size-shifting policy, 2) the average cost of a bonus that pays teachers \$4,000 for each additional student shifted into their classroom, and 3) the average cost of a bonus that pays teachers \$2,000 for each additional student shifted.

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

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.⁴⁰ The two bonus lines pictured here depict bonuses of \$2,000 (in green) and \$4,000 (in blue) for each additional student. For example, an effective teacher agreeing to teach twenty-five students, when an equal distribution of students would have resulted in a twenty-student class, is given a \$10,000 bonus under the \$2,000 level. 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.⁴¹ 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 fifth 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 eighth grade even when teachers are offered generous additional compensation.

Conclusion

Can student achievement improve by simply reallocating pupils among teachers? These results say yes. In both fifth and eighth 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 fifth grade, gains in eighth grade are large enough to result in a net-positive return even when teachers are generously compensated for accepting the additional student load. As a result, this strategic approach to class size shows promise as a single policy to reduce students' exposure to weak teachers, improve student access to effective instructors—and reward that effectiveness with cash.

These findings have several implications for state and local policymakers relative to equity and efficiency—some promising, some not. First, the not so promising. As shown in Table 1, gaps currently exist in students' access to high-quality teachers—whether quality 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.⁴²

More promising are the findings related to efficiency, which refers to the overall productivity of the work force 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 overall better outcomes for students in both larger and smaller classes, reflected in both cognitive and noncognitive measures.⁴³

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

The simulated learning gains in eighth grade are large enough to warrant special mention. Districts committed enough to aggressively pursue class-size shifting could realize the equivalent of an additional two and a half weeks of instruction in eighth-grade math and science. Even less-intensive strategies that would shift only six additional students into effective teachers' classes—creating disparities near those already seen 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 another proposed policy to promote teacher quality: removing the lowest-performing teachers from the classroom. The simulated gains in eighth-grade math and science achieved by shifting just six additional students to effective teachers are equivalent to the expected effect of removing the lowest 5 percent of teachers in these subjects—*and these gains can be achieved without actually removing them!*⁴⁴ Rather, class-size shifting enables the lowest-performing teachers to become more effective than they might be otherwise by allowing them to individualize instruction in smaller classes.

The sizable simulated gains in eighth grade contrast with the lesser gains estimated for fifth grade. Recall that these differences are likely due to averaging imprecise performance measures in self-contained classrooms. What if schools didn't use self-contained classrooms? We might have a very different result. Using a simulation, one study estimates 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 ten times the fifth-grade gain estimated from the class-size-shifting strategy presented here.⁴⁵ If elementary grades were departmentalized, as at least one study recommends,⁴⁶ pairing the gains from departmentalization with those from a class-size-shifting strategy would presumably increase the expected gains even more.

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.⁴⁷ Fortunately, pursuing this policy would likely require very small changes in practice. It could even be executed on a limited basis without any formal agreement or policy, requiring only minor changes to current practices in how students are allocated across teachers. On the other

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

hand, if schools were 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 would be well worth the modest effort required to make those changes.

Technical Appendix

Optimization Problem and Optimality Conditions

Consider the principal's assignment problem under variable teacher effectiveness, where her objective is to maximize total student learning (L) by distributing students across teachers with different levels of expected value-added performance. The main choice variable is 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 ($\tau_i(n_i)$).

$$(1) \quad \max L = \tau_j(n_j) \cdot n_j + \tau_k(n_k) \cdot n_k \text{ by choosing } n_j \\ \text{subject to: } n_j + n_k = N.$$

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:

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

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 condition⁴⁸:

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

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.

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

In short, an optimizing principal will allocate students across classrooms to the point where individual gains to a marginal student switching between classrooms (on the left-hand side) are offset by the net collective changes in learning among the remainder of 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 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. Under the prototypical case where teacher effectiveness varies across classrooms and there are small decreases in a teacher's realized effectiveness due to increasing class size, the optimizing principal will assign a larger class to the more effective teacher, and the smaller class to the less effective teacher. The difference in optimal class sizes between classrooms is a function of the difference in teacher effectiveness between classrooms. Note that with sufficiently large differences in teacher productivity across classrooms and/or sufficiently small marginal products associated with changes in class size, the optimal and unconstrained allocation may place all students in the most effective teacher's class.

In practice, when the assignment process is not otherwise constrained by class-size restrictions, it could look like this: The principal sorts teachers in descending order of expected productivity and 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 second-most 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 third-most effective teacher is greater than the losses to students in either of the other two classrooms, thereby introducing the third 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; 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).

Detailed Methods

Linking students with teachers

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 is provided in these files. 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.

Finally, payroll files are used to determine individual teacher experience, a key variable predicting a teacher's classroom performance. Teachers with missing experience values are recoded as 0 and flagged with a missing indicator variable.

Finding eligible schools

Some schools have only one unique teacher assigned to a particular grade (or grade-subject combination for middle schools) and therefore cannot be included in the strategic assignment process. Appendix Table 1 documents the percentage of unique classrooms and schools serving grade 5 (Panel A) or 8 (Panel B) in the 2010-11 school year where strategic sorting can occur since two or more unique teachers are assigned to classes of equivalent subjects and grades within a school. As shown, most classrooms can be strategically manipulated. For instance, 90 percent of all fifth-grade students are in self-contained classes in which the same teacher provides instruction across all tested subjects, and 92 percent of these classes are eligible for sorting due to the presence of another equivalent classroom taught by a different teacher. These strategic-eligible, self-contained classrooms are present in 96 percent of all schools that have self-contained classes serving fifth-grade students. The proportions of classrooms eligible for

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

strategic assignment are similar in eighth grade (ranging from 91 percent to 96 percent across subjects). The simulations below present results based on the sample of classrooms shown here to be eligible for strategic assignment.

Appendix Table 1. Classrooms and Schools in Which Class-Size Shifting is Possible

Panel A. Fifth-Grade Eligible Classrooms	
Number of student observations	113,046
Percent of students in self-contained/block classroom	90%
Total self-contained classes	5,611
Percent of unique self-contained classes eligible for class-size shifting	92%
Number of schools	1,410
Percent of schools with self-contained/block classroom	97%
Total unique self-contained schools	1,357
Percent of unique self-contained schools eligible for class-size shifting	96%
Panel B. Eighth-Grade Eligible Classrooms	
Number of unique math classes	2,653
Percent of unique math classes that are eligible for class-size shifting	95%
Number of unique reading classes	2,644
Percent of unique reading classes that are eligible for class-size shifting	96%
Number of unique science classes	1,453
Percent of unique science classes that are eligible for class-size shifting	91%
Total math/reading/science classes	6,378
Percent of unique math/reading/science classes that are eligible for class-size shifting	93%
Total schools in math/reading/science	791
Percent of unique math/reading/science schools that are shifting eligible in at least one subject	98%
Note: Counts and percentages based on 2010–11 school year using the North Carolina administrative data.	
Note: Strategically sorted classes are those in which two distinct teachers are observed with separate classes in the same school-grade-subject combination.	

Simulation details

The first three years of the North Carolina administrative data (2007–08 through 2009–10) are used to estimate a standard value-added model across years using teacher-averaged effects. These teacher-averaged effects from this model constitute their prior value-added estimates. This first model also includes explanatory variables on class size, classroom composition, and teacher experience; these coefficients are utilized later in the simulation. The class-size estimates and the variation of teacher value-added estimates are presented in Appendix 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 0.061 to 0.192). The prior value-added estimates and the coefficients on teacher experience are then combined to generate an “expected” performance value for all teachers that appear in the data during the 2010–11 school year. Finally, we estimate a teacher’s actual value-added performance in the 2010–11 school year, holding all teacher attributes and classroom composition constant. Obtaining both expected and realized teacher value-added estimates is critical for the simulation—the students are shifted across teachers’ classrooms based on *expected* performance while the simulated outcomes are based on *realized* performance.

Using the expected-performance estimates, all teachers in a school’s grade-subject combination are ordered from highest to lowest (teachers without prior value-added estimates are assumed average for their level of teaching experience). The simulation randomly sorts all students in a given school-grade, and then assigns them one at a time to available teachers in classrooms.⁴⁹ Students are assigned, one by one, to classrooms in which the difference between the expected gains for the student and expected losses for the classroom are greatest. The process continues until all students are assigned, with the last student assigned to the classroom where the difference between expected individual gains and classroom losses is the least. This simulation is performed three times, as the results vary slightly across each random iteration, and the outcomes are averaged across the three rounds to compute an expected value.⁵⁰

Estimating value-added, teacher experience, peer effects, and class size

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 five valid student observations in one school year are included in the analysis.⁵¹ I estimated using the following equation:

$$(5) \quad y_{i,c,j,t} = y_{i,t-1} \mathbf{I}_t \boldsymbol{\beta}_1 + \mathbf{X}_{i,t} \boldsymbol{\beta}_2 + \text{Size}_{c,j,t} \gamma + \mathbf{Class}_{c,j,t} \boldsymbol{\delta} + \mathbf{Exp}_{j,t} \boldsymbol{\zeta} + \mathbf{I}_j \boldsymbol{\tau} + \varepsilon_{i,c,j,t}$$

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 .⁵² 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 ($y_{i,t-1} \mathbf{I}_t$);⁵³ a vector of student characteristics ($\mathbf{X}_{i,t}$), 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 ($\text{Size}_{c,j,t}$); 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 percentile; a vector of categorical variables representing a teacher’s experience ($\mathbf{Exp}_{j,t}$);⁵⁴ 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 ($\hat{\boldsymbol{\delta}}$), within-teacher average returns to experience ($\hat{\boldsymbol{\zeta}}$), 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 Appendix Table 2. As expected, the estimates for 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).

Appendix 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*** (0.0005)	-0.0020*** (0.0005)	-0.0047*** (0.0005)	-0.0035*** (0.0002)	0.0000 (0.0003)	-0.0024*** (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.						

I do not interpret any of these estimates as causal, but will use them as parameters of interest for the simulation. 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 each 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.

I assume a principal forms his expectation of a teacher's future 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 ($\hat{\zeta}$), as estimated in Equation (5), which is associated with the teacher's level of experience for the coming school year. Because teachers' value-added estimates ($\hat{\tau}_j$) 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 $\hat{\lambda}$).⁵⁵ Thus, the principal's expectation of a teacher's performance is computed as:

$$(6) \quad \hat{\theta}_{j,t+1} = \text{Exp}_{j,t+1} \hat{\zeta} + \hat{\lambda} \mathbf{I}_j \hat{\tau}_j$$

The final element necessary is 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):

$$(7) \quad y_{i,c,j,2011}^* = y_{i,c,j,2011} - \text{Size}_{c,j,2011} \hat{\gamma} - \text{Class}_{c,j,2011} \hat{\delta} - \text{Exp}_{j,2011} \hat{\zeta}$$

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.⁵⁶ 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:

$$(8) \quad y_{i,c,j,2011}^* = y_{i,2010} \beta_1 + X_{i,2011} \beta_2 + \mathbf{I}_j \tau + \varepsilon_{i,j,2011}$$

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

Simulating educational outcomes under strategic class-size assignment

The estimation of the various elements discussed above leads to the simulation, where I strategically manipulate class sizes across teachers in relation to the expected difference in teacher productivity across 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, as previously documented in Appendix Table 1, in addition to those having more than five 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.⁵⁷

The unconstrained assignment process occurs as follows. Knowing each teachers' expected productivity functions ($\tau_i(n_i)$), 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 second-most 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 third-most effective teacher is greater than the losses to students in either of the other two classrooms. At this point the third teacher is introduced 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; 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).

Constrained optimization of total student learning follows a similar process. For instance, if there were a constraint limiting class-size for the highest-performing teacher to only three more students than what she expected under equal-sized classes, the optimal assignment process would proceed as described above; the only variation would be to introduce a new teacher when either the optimality condition held across classrooms (as above) or when the new three-student constraint was binding. Yet even under this scenario with additional constraints, the optimal class sizes will reflect expected differences in teachers' performance.

The calculated outcome measures are described in further detail below.

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

$$(9) \quad \Delta y_{i,c,j,2011} = \Delta I_j \hat{\tau}_{j,2011} + \Delta \text{Exp}_{j,2011} \hat{\zeta} + \Delta \text{Size}_{c,j,2011} \hat{\nu} + \Delta \text{Class}_{c,j,2011} \hat{\delta}$$

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.⁵⁸ Note that this expression assumes the other elements of student learning (namely, $y_{i,2010} \hat{\beta}_1 + X_{i,2011} \hat{\beta}_2 + \hat{\varepsilon}_{i,c,j,2011}$) 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 teachers in the top 25 percent** – 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 (both actual and simulated values of this measure are computed in the same way).
3. Proportion of FRL students **assigned to teachers in the top 25 percent**. 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 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.

Methodological limitations

The simulation results rest on several methodological assumptions, which limit the generalizability of the results. Namely, the analysis assumes that the difference in student outcomes is determined solely by the difference of estimated teacher productivity and classroom composition effects; and the simulation assumes a teacher’s performance is similarly insensitive

EMBARGOED UNTIL 12:01 AM ET, NOVEMBER 19, 2013

to changes in the classroom composition and size, beyond those that are accounted for with observable characteristics. Either of these assumptions could be violated in practice, which would result in some additional variation in outcomes that are not present in the calculated results. Additionally, the analysis estimates class-size effects that are assumed to be linear and constant across all teachers.⁵⁹ It is possible that the class-size effects estimated here may either under- or overestimate real class-size effects; this is not a serious threat to the results presented here, however, as either case implies the effects of implementing this policy are potentially larger.⁶⁰

A scholarly version of this study is forthcoming as a CALDER Working Paper, and will present further investigations and sensitivity checks on the results presented here.

Endnotes

¹ Several studies explicitly compare class-size effects with the variation in teacher effectiveness across the work force; see for example Eric A. Hanushek and Steven G. Rivkin, “Generalizations about Using Value-Added Measures of Teacher Quality,” *American Economic Review* 100, no. 2 (2010): 267–71; Barbara Nye, Spyros Konstantopoulos, and Larry V. Hedges, “How Large are Teacher Effects?,” *Educational Evaluation and Policy Analysis* 26, no. 3 (2004): 237–57; and Douglas Staiger and Jonah E. Rockoff, “Searching for Effective Teachers with Imperfect Information,” *Journal of Economic Perspectives* 24, no. 3 (2010): 97–118.

² Self-contained classrooms are those in which students are taught by the same teacher for all subject areas. For this study, the fifth-grade sample is limited specifically to self-contained classrooms.

³ For 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, see Raj Chetty, John Friedman, and Jonah Rockoff, “The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood” (National Bureau of Economic Research Working Paper 17699, Cambridge, MA, 2011).

⁴ These numbers are based on the observed data in North Carolina that are used for this study. See table 1 for statistics on current patterns of assignment.

⁵ Secretary Duncan includes this suggestion in a list of possible ways schools could constructively adapt to the “new normal” in education under budgetary pressures; Bill Gates promotes the idea in a *Washington Post* op-ed. See Arne Duncan, “The New Normal: Doing More with Less” (speech, American Enterprise Institute, Washington, D.C., November 17, 2010), <http://www.ed.gov/news/speeches/new-normal-doing-more-less-secretary-arne-duncans-remarks-american-enterprise-institut>; and Bill Gates, “How Teacher Development Could Revolutionize Our Schools,” *Washington Post*, February 28, 2011, <http://www.washingtonpost.com/wpdyn/content/article/2011/02/27/AR2011022702876.html>.

⁶ Frederick M. Hess and Olivia Meeks, “Sounding the Alarm: A Wake-Up Call with Directions,” in *Refocus Wisconsin* (Milwaukee, WI: Wisconsin Policy Research Institute, 2010).

⁷ This initiative provides ten different models (plus variations and combinations) of how schools can organize around their most effective teachers. Models include subject or role specialization, multiclassroom leadership, and time-technology swaps, among others. The strategy analyzed here, where principals strategically shift class sizes to give more students to the most effective teachers and fewer 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 optimal conditions for determining class size in the face of variable teacher productivity, as this paper does. For

more information on the models in Public Impact’s Opportunity Culture initiative, see <http://www.opportunityculture.org>.

⁸ Class-size policies are well liked by parents and teachers and are commonly believed to have positive effects on student outcomes; policies limiting class size have been adopted in thirty-six states, contributing to the steady decline of student-teacher ratios in public schools nationwide over the last few decades of the twentieth century. For further discussion, see June Ahn and Dominic J. Brewer, “What Do We Know About Reducing Class and School Size?,” in *Handbook of Education Policy Research*, ed. Gary Sykes, Barbara Schneider, and David N. Plank (New York: Routledge, 2009), pp. 426–37.

Yet the findings from the most rigorous studies on class size do not neatly conform to the popular perception. A recent review by Grover J. Whitehurst and Matthew M. Chingos provides an overview of the class-size research, focusing specifically on the few studies with robust research designs; see *Class Size: What Research Says and What It Means for State Policy* (Washington, D.C.: Brookings Institution, 2011). This appraisal suggests *at most* a moderate class-size effect from ongoing exposure to small classes. (For more on their review, see the appendix). Based on the magnitudes of class-size estimates compared with the effects due to teacher productivity, having a high-performing teacher (at the 85th percentile of value-added estimates) rather than an average one is equivalent to a ten-to-twenty-student reduction in class size. See Eric A. Hanushek and Steven G. Rivkin, “Generalizations about Using Value-Added Measures of Teacher Quality,” *American Economic Review* 100, no. 2 (2010): 267–71. Barbara Nye, Spyros Konstantopoulos, and Larry V. Hedges analyze teacher effectiveness using data from the Tennessee STAR class-size experiment and estimate that a ten-to-twenty-student difference in class size would be required to compensate for the learning difference between an average and highly effective teacher. See “How Large are Teacher Effects?,” *Educational Evaluation and Policy Analysis* 26, no. 3 (2004): 237–57.

⁹ 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, “Generalizations.”

¹⁰ See for example Daniel Aaronson, Lisa Barrow, and William Sander, “Teachers and Student Achievement in the Chicago Public High Schools,” *Journal of Labor Economics* 25 (2007): 95–135; and Hanushek and Rivkin, “Generalizations.”

¹¹ See Dan Goldhaber and Michael Hansen, “Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance,” *Economica* 80, no. 319 (2013): 589–612; and Douglas Staiger and Jonah E. Rockoff, “Searching for Effective Teachers with Imperfect Information,” *Journal of Economic Perspectives* 24, no. 3 (2010): 97–118.

¹² National Council on Teacher Quality, *State of the States 2012: Teacher Effectiveness Policies* (Washington, D.C.: National Council on Teacher Quality, 2012).

¹³ See Scholastic and the Bill and Melinda Gates Foundation, *Primary Sources: 2012—America’s Teachers on the Teaching Profession* (New York: Scholastic, 2012), http://www.scholastic.com/primarysources/pdfs/Gates2012_full.pdf.

¹⁴The largest credible estimated effect of class size reported in Whitehurst and Chingos, *Class Size*, is a 1999 analysis of the Tennessee STAR experiment in grades K–3 by Alan B. Krueger, “Experimental Estimates of Education Production Functions,” *Quarterly Journal of Economics* 115, no. 2 (1999): 497–532. The analysis found that students in small classes, averaging seven to eight fewer students than comparison classes, increased their test scores 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 improvement 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 considerably smaller in magnitude. Thus this estimate could be considered an upper-bound class-size effect, and its magnitude has not been duplicated in other rigorous studies.

¹⁵ Whitehurst and Chingos, *Class Size*, identify four studies showing credible, statistically significant, positive effects associated with smaller class sizes; another three studies showing mixed results (where estimates of the class-size reduction are significantly positive in some cases or for some student groups, but the overall class-size effect is not statistically significant); and another two finding no significant class-size effect. The authors interpret the 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).

¹⁶ The analysis presented here uses teacher fixed effects to estimate class-size effects in grades 5 and 8 with magnitudes (ranging from 0 to 0.0052 standard deviations of student achievement for each decrease in class size by one student) that are much smaller than Krueger’s estimates based on grades K–3; see Krueger, “Experimental Estimates.” These smaller returns for class-size reduction in the higher grades appear qualitatively consistent with the broader research. See full results in the appendix.

¹⁷ Within-school standard deviations of teacher productivity range from 0.08 to 0.26 standard deviations of student achievement in reading and 0.11 to 0.36 standard deviations in math; see Hanushek and Rivkin, “Generalizations.” In their analysis of teacher effectiveness using data from the Tennessee STAR class-size experiment, Nye, Konstantopoulos, and Hedges estimate that a ten-to-twenty-student difference in class size would be required to compensate for the learning difference between an average and a highly effective teacher; see Nye, Konstantopoulos, and Hedges,

¹⁸ For an account of how the increase in demand for teachers—a function of class-size reduction policies adopted in the late 1990s—inadvertently undermined the qualifications of the teacher work force in California, see Christopher Jepsen and Steven G. Rivkin, “Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size,” *Journal of Human Resources* 44, no. 1 (2009): 223–50.

¹⁹ See the appendix for a thorough development of this optimization problem and the optimality conditions.

²⁰ From a starting point of equal class sizes, the benefits to the weaker teachers' students exactly cancel out the penalties to the stronger teachers' students, but as more students are shifted, the net effect becomes slightly more negative. In addition, because the estimated magnitude of this class-size effect is small in the North Carolina data used here, the net class-size penalty makes only a small difference, even in scenarios where many students are shifted across classrooms.

²¹ End-of-grade math and reading tests are administered in grades 3–8, while science is administered only in grades 5 and 8. North Carolina's standardized tests changed in 2012–13 (the school year following the time span of the data here) to begin testing to the Common Core State Standards.

²² Additional details on the construction of the data are presented in the appendix.

²³ For the fifth-grade samples, I focus specifically on self-contained classrooms (where the same instructor teaches at least both reading and math, or all three subjects); I remove departmentalized elementary schools from the samples. Over 90 percent of schools meet the necessary conditions to implement this strategy; further details are presented in the appendix. It is important to note that the class-size-shifting strategy discussed here may not be readily applied in all schools. The contexts in which implementing this strategy is more difficult, if not impossible, are discussed in the "Feasibility Issues" sidebar on p. 26.

²⁴ See the appendix for full methodological details.

²⁵ Some may presume, since prior performance is known, that there is no uncertainty about who is an effective teacher. Though we do have data about which teachers are expected to perform better, we cannot predict future performance with certainty. Some teachers predicted to perform well actually do not, while others may improve; hence, I use the term "expected performance" deliberately.

²⁶ Note that all legal class-size limits are ignored for these particular simulations. According to the National Council on Teacher Quality's Tr3 Database, twenty-eight states have explicit limits on class size in at least one grade; in states that do not directly limit class size, district-level collective bargaining agreements commonly include such provisions. Only ten states include explicit restrictions (several other states have recommendations, but not limits) about some element of the distribution of class sizes—for example, requiring that the average class size in a grade not exceed a certain limit. 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 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.

²⁷ Note that the average class-size and average class-size deviation values vary across subjects in fifth grade, though these are self-contained classrooms. This variation 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.

²⁸ The weak positive association between these variables is consistent with the evidence presented in Nathan Barrett and Eugenia F. Toma's analysis using data from ten districts in Kentucky, again suggesting that some strategic class-size determination may already be occurring to a limited extent; see "Reward or Punishment? Class Size and Teacher Quality," *Economics of Education Review*, forthcoming.

²⁹ Designations for teachers in the top 25 percent 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.

³⁰ Recall that six distinct samples are analyzed, corresponding to the unique subject-grade combinations in the data for grades 5 and 8 (three subjects across two grades).

³¹ Though not reported in table 1, a gap in access also arises when using more traditional teacher credentials. For instance, roughly 71 percent of fifth-grade students are assigned to teachers with five or more years of teaching experience, but 69 percent of FRL students have such assignments. The corresponding numbers for eighth grade are 72 percent and 70 percent, respectively. The proportion of students assigned to teachers with a master's degree or higher is 33 percent overall and 32 percent among FRL students in fifth grade, and 34 percent overall and 32 percent among FRL students in eighth grade.

³² A gain of 0.02 standard deviations of student achievement is approximately 7 percent of the average annual gain for the grade 7 to 8 transition, based on nationally normed tests; see Carolyn Hill, Howard Bloom, Alison Black, and Mark Lipsey, "Empirical Benchmarks for Interpreting Effect Sizes in Research," *Child Development Perspectives* 2, no. 3 (2008): 172–77. 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.

³³ Dan Goldhaber, James Cowan, and Joe Walch, "Is a Good Elementary Teacher Always Good? Assessing Teacher Performance Estimates across Subjects," *Economics of Education Review*, forthcoming.

³⁴ Prior teaching experience and holding a graduate degree are not equally predictive of future performance (see Goldhaber and Hansen, "Is It Just a Bad Class"); however, they both may signal greater commitment to the teaching profession and are more common metrics of student access to credentialed teachers.

³⁵ In eighth 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 eighth-grade measures represented in figure 2 are calculated from the eighth-grade math sample. Teacher assignments in fifth grade are not subject specific.

³⁶ Dan Goldhaber, Michael M. DeArmond, and Scott DeBurgomaster, “Teacher Attitudes about Compensation Reform: Implications for Reform Implementation,” *Industrial Labor Relations Review* 46, no. 3 (2011): 441–63.

³⁷ 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>.

³⁸ 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.

³⁹ The analysis is by Chetty, Friedman, and Rockoff, “Long-Term Impacts.” The authors estimate the marginal impact of one standard deviation increase in teacher quality on the net present value of students’ future earnings at age twenty-eight at 0.9 percent, or roughly \$4,600 per grade (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.

⁴⁰ Note that bonuses are assumed to be paid for *extra* students, but the cost is averaged across *all* students.

⁴¹ It may seem counterintuitive to directly compare future private benefits with current public costs, but this is an inherent property of public education, which is an intergenerational 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.

⁴² A recent study 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. Some of Public Impact’s Reach Extension models seek to bridge the teacher-quality divide across schools by utilizing technology to broadcast effective teachers across multiple campuses. See Jason Grissom, Susanna Loeb, and Nathaniel Nakashima, “Strategic Involuntary Teacher Transfers and Teacher Performance: Examining Equity and Efficiency” (National Bureau of Economic Research Working Paper 19108, Cambridge, MA, 2013).

⁴³ Students assigned to high-quality teachers benefit on several dimensions, as shown by Chetty, Friedman, and Rockoff, “Long-Term Impacts.” Those in smaller classes also benefit; the cognitive evidence was discussed above, and Thomas S. Dee and Martin R. West estimate positive returns on noncognitive dimensions, such as school engagement, due to smaller class sizes in eighth grade. See Thomas S. Dee and Martin R. West, “The Non-Cognitive Returns to Class Size,” *Education Evaluation and Policy Analysis* 33, no. 1 (2011): 23–46. The class-size strategy laid out here increases the overall variation in class sizes because it increases 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.

⁴⁴ The calculation is based on the expected mean of the teacher distribution after removing the bottom 5 percent, and is taken from William H. Green, *Econometric Analysis* (Upper Saddle River, NJ: Prentice-Hall, 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 eighth-grade math and science in appendix table 2. Eric Hanushek estimates that the cumulative effect of removing the lowest-performing 6 percent to 10 percent of teachers from the profession would be sufficient to bring overall achievement levels in the United States to those in Canada. See “Teacher Deselection,” in *Creating a New Teaching Profession*, ed. Dan Goldhaber and Jane Hannaway (Washington, D.C.: Urban Institute Press, 2009), pp. 165–80.

⁴⁵ See Goldhaber, Cowan, and Walch, “Is a Good Elementary Teacher Always Good.”

⁴⁶ See Brian A. Jacob and Jonah E. Rockoff, “Organizing Schools to Improve Student Achievement: Start Times, Grade Configurations, and Teacher Assignments” (Hamilton Project Discussion Paper 2011-08, Brookings Institution, Washington, D.C., 2011).

⁴⁷ Of course, the specific class-size-shifting strategy analyzed here is one of many potential ways that more students could be exposed to high-quality teaching. Other strategies include digital learning to supplement high-quality classroom instruction without lowering class sizes, or leadership roles for effective teachers to allow them to manage and supervise other teachers’ efforts.

⁴⁸ Nathan Barrett and Eugenia F. Toma, “Reward or Punishment? Class Size and Teacher Quality,” *Economics of Education Review*, forthcoming, 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. Barrett and Toma assume that 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 consider only 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).

⁴⁹ Given the expected-performance estimates for all teachers, the distribution of class sizes across teachers is predetermined, but the composition of the class—the students assigned to them—is not. Hence, the random ordering of students and successive assignment are necessary to observe the variation in outcomes across different classes of students assigned to teachers.

⁵⁰ 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, twenty 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 precise result.

⁵¹ Note that I use all classrooms at this point to generate the estimates of interest, with the exception of dropping classes with five or fewer students to avoid the inclusion of specialty classes. I will be limiting the sample in the simulation to those classrooms that are flagged for strategic class-size manipulation.

⁵² 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."

⁵³ Students with missing values in their same-subject prior test score are dropped from the analysis sample. For those who are 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, I base imputation on other observed variables in the data.

⁵⁴ The omitted category is zero years of experience, and separate indicator variables are established at the following intervals: one to two years of experience, three to four years, five to nine years, ten to fourteen years, fifteen to nineteen years, and twenty or more years.

⁵⁵ I obtain these scalar weights by using a version of equation (5) that produces two-year value-added estimates based on the 2007–08 and 2008–09 school years (instead of producing three-year estimates). These two-year 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 multiyear value-added estimates. For elementary grades, I use the average value of this scalar across subjects to create the expected-performance measures.

⁵⁶ 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 in the interest of treating all teachers uniformly, regardless of the number of classes they taught in 2010–11.

⁵⁷ 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 eighth-grade math teachers in a school, and in

the actual data one teaches three classes while the other teaches one, either teacher could potentially be assigned up to three classes in the simulations. I additionally impose a constraint that limits the number of unique classes within a school-grade-subject to be less than or equal to the original number of classes observed; this constraint ensures that learning gains do not come from artificially lower class sizes, achieved by assigning the same number of students to more simulated classes.

⁵⁸ 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.

⁵⁹ One study of class-size effects speculates about an interaction between class-size effects and teacher productivity where class-size effects are larger for less effective teachers but near zero for the most productive teachers. See Ludger Woessmann and Martin West, “Class-Size Effects in School Systems around the World: Evidence from Between-Grade Variation in TIMSS,” *European Economic Review* 50, no. 3 (2006): 695–736. I do not consider this potential interaction in the simulations. Such an interaction would imply both larger and smaller classes could be more productive than what the simulations estimate, and the results presented here may be considered lower-bound estimates on probable outcomes from this strategy.

⁶⁰ If class-size effects were smaller than estimated (i.e., zero), teacher assignment itself would be the only key component in calculating student gains, where class-size effects currently contribute a small net negative effect on this calculation. If class-size effects were larger, say as large as those presented in Alan B. Krueger’s 1999 study, the implication is that the variation in teacher quality is much larger (recall the slightly positive sorting between class size and teacher quality; if class-size effects are larger, true teacher effects must show greater variability than what is estimated here). 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. The simulation was conducted with larger imposed class-size effects to investigate this. See Alan B. Krueger, “Experimental Estimates of Education Production Functions,” *Quarterly Journal of Economics* 115, no. 2 (1999): 497–532,