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

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

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

Reaching Further and Learning More? Evaluating Public Impact's *Opportunity Culture* Initiative

Ben Backes
Michael Hansen

WORKING PAPER 181 • January 2018

Reaching Further and Learning more? Evaluating Public Impact's *Opportunity Culture* Initiative

Ben Backes
American Institutes for Research

Michael Hansen
The Brookings Institution

Contents

Acknowledgements

We gratefully acknowledge funding from an anonymous funder. Public Impact is a cograntee on the larger project, which included funding for Public Impact's data gathering to help improve service delivery and promote its work, but was not directly involved in the external evaluation presented here. Public Impact cooperated with the evaluation by participating in informational phone calls and sharing program data, but the analysis was conducted and the report was written independently from Public Impact. Public Impact reviewed and offered clarifying comments on an early draft (which was also circulated to the funder and partner districts), though only Backes and Hansen have editorial privilege over the paper. We also thank the three partner districts for their helpfulness in providing data. Amanda Oliveira and Gerhard Ottehenning provided excellent research assistance. Finally, we thank seminar participants at the University of Maryland Baltimore County and Center for American Progress for helpful comments.

This research was supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER) funded through Grant R305A060018 from the Institute of Education Sciences, U.S. Department of Education.

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

Reaching Further and Learning more? Evaluating Public Impact's *Opportunity Culture* Initiative

Ben Backes, Michael Hansen

CALDER Working Paper No. 181
January 2018

Abstract

Public Impact's *Opportunity Culture* (OC) initiative provides a suite of models aimed at extending the reach of highly effective teachers and has partnered with school districts to implement these interventions in schools. Using administrative data from three partner school districts that collectively include 44 OC schools, we estimate the relationship between OC staffing models and student achievement in math and reading. We find that the bulk of exposed students received treatment under OC's multiclassroom leadership model, in which a master teacher with demonstrated effectiveness intensively leads and coaches a team of teachers, and that these students scored higher in math in all specifications. In reading, while most specifications find positive and significant learning gains for students taught by team teachers, the specification that performs best in our placebo tests – one that includes school-by-year fixed effects to account for overall improvement in treatment schools – does not find any impact. Results for other specifications are more mixed and are imprecisely estimated.

I. Introduction

Teachers matter more to student achievement than any other schooling input. An excellent teacher in the classroom makes sizeable and significant impacts on a myriad of student outcomes, both in the short and long term (Chetty et al., 2014). Yet, teachers' effects on student learning vary greatly across the teacher workforce (Hanushek & Rivkin, 2010), often leaving historically disadvantaged students with lower access to effective teachers (Isenberg et al., 2013; Goldhaber et al., 2015).

Many current reforms to improve public education in the United States now seek to implicitly manipulate teacher talent to increase the overall productivity of the teacher workforce and thus improve student outcomes broadly. Examples of these reform efforts focused on teacher quality include improving teacher evaluations and the formative support teachers receive, more selective retention, and efforts to draw more talent into the profession.[1]

A less prominent approach to improving the productivity of the workforce as a whole is to make better use of the existing supply of highly effective teachers by deploying them differently in schools (Hansen, 2013, 2016). This approach reworks traditional teacher roles or classroom staffing practices, or uses technology in ways that seek to amplify the influence of effective teachers across more students. This alternative strategy contrasts with current practice, which deploys all teachers in approximately equivalent ways even though it is known that their productivity varies greatly. The idea of extending the reach of more effective teachers is in line with labor specialization that occurs in other industries (Hess, 2009) as a means of productivity gains. This specialization approach could be particularly beneficial in disadvantaged schools given evidence that much of the variation in

[1] See for example DC IMPACT, in which teachers were given rewards or dismissal threats based on performance (Dee & Wycoff, 2015), and reforms to teacher tenure decisions (Loeb et al., 2015).

teacher quality is within schools, rather than across schools; hence, even the typical disadvantaged school will have a pool of high-quality teachers from which to draw (e.g., Sass et al., 2012).

Public Impact, a national education organization based in Chapel Hill, North Carolina, designed and launched the *Opportunity Culture* (OC) initiative in 2009, and has been a leading advocate of these "reach extending" strategies to strategically utilize effective teachers already in schools. OC provides a suite of models and roles aimed at extending the reach of highly effective teachers within the constraints of schools' normal operating budgets, and to date has partnered with 19 school districts to implement these interventions in schools.[2]

This paper is an evaluation of a subset of Public Impact's partner districts that have adopted these OC models. According to Public Impact's records, 52 schools across four states have implemented OC models as of the 2015–16 school year (North Carolina, New York, Tennessee, and Texas). We collected administrative data from three partner school districts that collectively include 44 OC schools, although because not all of OC partner schools contain students in tested grades and subjects, not all of them appear in our analysis sample.[3] In total, approximately 15,000 unique students were exposed to an OC intervention model for one or more years in our analysis sample. We estimate the relationship between implementing these OC staffing models and student achievement based on standardized tests in reading and math. In addition, we explore the selection of teachers for prominent roles in OC models. To our knowledge, this is the first empirical study of these alternative staffing models in authentic learning environments.

In summary, we find that OC models reached thousands of students in these three school districts, with more than 50 percent of students in treatment schools being exposed to an OC model in the final year of the study, 2015-16. The bulk of exposed students received treatment under OC's

[2] A list of partner districts may be found at http://opportunityculture.org/dashboard/sites/.
[3] As described in further detail below, because we do not have unique school identifiers in one of our three partner districts, we cannot calculate the exact number of treatment schools that appear in the analysis sample. This district represents a small share of the analysis sample.

multiclassroom leadership model, in which a master teacher with demonstrated effectiveness intensively leads and coaches a team of teachers. Treated students in these models were in classrooms led by a team teacher, and we find these students scored 11% of a standard deviation higher in math in our preferred specification. This is an educationally significant estimate, roughly equivalent to the improvement associated with replacing an average teacher in the classroom with one in the top quartile of expected value-added performance in our study data. In reading, while many models find positive and significant effects, the model that performs best in our placebo tests – one that includes school-by-year fixed effects to account for overall improvement in treatment schools – does not find any impact. Results for other models are more mixed.

II. Theory of Action and Prior Research

The underlying theory of action behind all OC models is that teachers who have demonstrated excellence in the classroom apply for and are given roles or assignments that enable them to influence more students' learning. Prior research over the last decade has shown teachers to be the most critical school-based factor influencing student achievement gains, though teachers vary significantly in their classroom effectiveness in ways that are difficult to predict (see Aaronson et al., 2007; Goldhaber, Brewer, & Anderson, 1999; Hanushek & Rivkin, 2010; Nye et al., 2004). Exposure to a high-quality teacher shows an arguably causal impact on myriad student outcomes in the short and long term (Backes & Hansen, 2018; Chetty et al., 2014a, 2014b; Staiger & Rockoff, 2010). OC's models that strategically assign effective teachers to influential roles are hypothesized to benefit students who now have greater access to instruction from effective teachers, rather than the presumably less effective instruction they would have otherwise received under traditional staffing practices. As we discuss below, however, the strength of the prior evidence supporting the theories motivating each of these OC models varies.

As suggested by the name, Public Impact designed the OC models with the intent to provide an opportunity for teachers to be selected and rewarded for excellence in the classroom via extending their reach to more students, either directly or by leading and supporting a team of teachers. To be sustainable as a viable long-term strategy for school improvement, the models assume that providing such opportunities to teachers will reward and retain highly effective teachers, and incentivize new and less effective teachers to strive for excellence and provide them with better support. These opportunities are believed to create a culture of upward mobility based on excellence in an occupation with career trajectories that are generally flat and resistant to differentiating by performance. Based on recruitment numbers from Public Impact, participating schools receive multiple applications for each OC position they attempt to fill (ranging from 2 to 74 applicants per position); thus, teachers appear to value this opportunity.[4]

OC shares similar objectives of identifying and rewarding excellent teachers and/or enhancing opportunities for upward mobility with typical performance-based career ladder, merit pay, or retention programs. Importantly, however, OC differs in that it indirectly rewards teachers for excellence. Differential pay in OC is predicated upon taking on a different teaching role, with performance only determining eligibility for that role. Because of this indirect link between performance and rewards, OC models should presumably be more politically palatable for a profession that has largely resisted differential pay for the same work as they differentiate compensation based on different work roles.[5] OC pay is also substantial, with supplements averaging $12,000 in 2017-17 that are typically paid through the school year, not as a one-time bonus.[6] Differential roles and pay appear to be particularly

[4] See http://opportunityculture.org/dashboard/recruitment/.
[5] For a lengthier discussion of differentiating teacher compensation and teacher attitudes toward these proposals, see Goldhaber et al. (2011).
[6] See http://opportunityculture.org/dashboard/extra-pay/

critical in retaining effective teachers in disadvantaged schools (see Simon & Johnson, 2013). Furthermore, students in disadvantaged schools tend to suffer from relatively lower levels of access (relative to students in more affluent schools) to teacher quality based on a variety of teacher performance metrics (e.g., Isenberg et al., 2013; Goldhaber et al., 2015),[7] hence increasing the urgency about attempting to provide access to quality teaching in different ways.[8]

The following subheadings discuss prior research that is relevant to the two OC models that we observe in the study's data.

Multiclassroom Leadership

Under the multiclassroom leadership strategy, highly effective teachers take on a leadership role for a team of teachers and become accountable for the learning of all students in the classrooms of team teachers. The multiclassroom leader (MCL) becomes a mentor and instructional resource for all on the team, and leadership responsibilities include supervising instruction, evaluating and developing teachers' skills, and facilitating team collaboration and planning. To focus efforts on leading and developing other teachers, the leaders themselves often do not have any students assigned alone to them (or they may have a light teaching load).[9] Rather, they are expected to be frequently involved with, supporting, and even in front of the classrooms of team teachers.[10] In the estimates presented in this paper, exposure to "MCL" refers to students taught by the MCLs directly, while exposure to "team

[7] See Isenberg et al. (2016) for a dissenting view.
[8] Other OC principles include respect for the autonomy of teachers (eligible teachers are given the choice to participate, as such roles may entail taking on a larger workload), matching authority and autonomy to each person's responsibilities, and staffing changes and rewards made without stretching school budgets. To be financially sustainable, Public Impact expects to fund rewards for OC teachers out of resources that would otherwise be used for hiring separate learning specialists or make similar staffing tradeoffs. Public Impact maintains that schools are able to sustain these models in cost-neutral ways, aside from the upfront costs of adoption.
[9] MCLs who do not have students assigned to them do not show up in our data as linked to a student. Our estimates below for MCLs themselves are only for MCLs linked to students. However, we are able to identify students who are taught by team teachers serving under an MCL, which we refer to as a team teacher effect.
[10] Note that this role is distinct from the lead teacher role that is common in schools throughout the United States, which is typically a first-among-equals arrangement in which a teacher may facilitate small-group meetings with other teachers in the same subject or grade, but still teach their own full class load and rarely exert any direct influence over teachers or students in other classrooms.

teacher" refers to students taught by a team teacher serving under an MCL. It is the latter who constitute the bulk of students exposed to OC models in the study data.

We do not know of any research directly evaluating the efficacy of the MCL and team teachers as OC describes in its model. However, many prior studies have explored the impact of mentors and instructional coaches among teachers. Kraft et al. (2016) present a meta-analysis of 44 causal studies (often experiments) of individualized, intensive, sustained, focused, and context-specific teacher coaching programs. Overall, they find large positive impacts on measures of teachers' instructional practice, but more muted effects (roughly one third in magnitude of the instructional practice effect) on student achievement, with the largest impacts in reading. There is some evidence to suggest the quality of the coach matters: Papay et al. (2016) report on an experiment in Tennessee in which teachers with low evaluation scores were strategically paired with strong teachers on the same dimensions where the partner teachers were weakest, and found statistically and educationally significant improvements in student learning among teachers who were coached. Similarly, Rockoff (2008) finds that metrics of high-quality mentoring were positively correlated with coached teachers' self-reported improvements in classroom performance. Mentoring programs have long been perceived as effective strategies to develop and retain new teachers, particularly in high-need schools, and generally (Ingersoll & Strong, 2011) though not always (Rockoff, 2008) these hypotheses are borne out in practice. Blazar and Kraft (2015) explore the mechanisms of effective teacher coaching using data from New Orleans and find teacher-to-coach ratios, the duration and intensity of the coaching relationship, and turnover among coaches as important factors predictive of differential outcomes among teachers who were coached. In summary, the research evidence supports the notion that intensive and frequent instructional coaching is beneficial for teachers' practices and student achievement, and having an effective coach, as the MCL model uses, is an important dimension of coaching success.

Direct Reach Extension

OC has several models through which highly effective teachers may directly reach more students, though only one—the time swap model—is observed in the study data.[11] For brevity, we focus only on the underlying theory behind this particular model. The time swap model utilizes different learning stations in a classroom or school to enable effective teachers to lead instruction for more students in critical subjects. The other stations typically include either computer-based learning exercises or occasionally small-group or independent learning time facilitated by paraprofessionals. In our sample, time swap teachers are labeled as either "Blended Learning" or "Expanded Impact". These models are similar but the use of technology is more of an emphasis with blended learning, while expanded impact does not necessarily incorporate the use of technology. The potential gain from these strategies comes from unbundling the typical instructional tasks that occur in classrooms (and are all performed by the same instructor), allowing the most effective teacher to lead the critical lecture-based instruction while less critical tasks are supervised by staff.

Rocketship Education, a charter school network, uses this station-based approach prominently in its classrooms and has been highlighted for its lower cost structures than neighboring traditional public schools and its aggressive growth (Kowal & Brinson, 2011; Herold, 2014). Taylor (2015) investigates the interaction of teacher quality and computer-aided instruction in classrooms and finds that less effective teachers gain the most by supplementing their instruction with computers, while the most effective teachers were penalized by using computers. The OC blended learning time swap model is loosely consistent with an optimal staffing strategy based on these findings: that is, the weakest

[11] Elements of two additional strategies for changing the utilization of teachers are also present to some extent even though they do not show up in the teacher job titles in our data. First, MCLs and Direct-Reach Teachers can specialize in a specific subject and will be labeled as "MCL" or "Direct-Reach Teacher" rather than subject specialist. And second, while class size increase is not a named role in the data, both BLT and Expanded Impact are essentially class size increases with a rotation and support of a paraprofessional.

teachers (or, in OC's case, teaching paraprofessionals) use blended learning while the strongest teachers focus on live instruction.

III. Data and Setting

The data for this study come from educational administrative records in three public school districts: Charlotte-Mecklenburg Schools (North Carolina), Cabarrus County Schools (North Carolina), and Syracuse City School District (New York). All districts have partnered with Public Impact to implement OC models in at least three schools for at least two academic years. At a minimum, all districts identify teachers participating in an OC model using teacher rosters provided by Public Impact, provide information about which model is being implemented, and report student achievement on the state's standardized tests in Grades 3 through 8 in elementary and middle school and end of course exams in high school.[12] Importantly, these test results are used as key outcomes for this evaluation. As the extent of the Public Impact partnership and the amount of data provided by each district varies, we describe the setting and data available in each district below.

Charlotte-Mecklenburg Schools (North Carolina)

Charlotte-Mecklenburg Schools (CMS) has been one of Public Impact's closest partner districts for the OC initiative, and boasts both the longest length of implementation (OC schools have been operating since the 2013–14 school year) and the greatest number of OC schools (in our sample, 18 schools implement at least one OC model), resulting in nearly 90% of students exposed to OC models in our sample coming from CMS. OC models were initially targeted to CMS schools that were operating in

[12] The end of course exams are Algebra I in all districts as well as NC Math I and English II in the two North Carolina districts. All test scores are standardized within test type, grade, and year to have mean zero and standard deviation one. About 85 percent of OC student-teacher links in math and 95 percent in reading are at the elementary and middle school levels, with the remainder in grades 9 and above.

Project L.I.F.T., the district's effort to turn around low-performing schools. However, OC models are now made available to other schools in the district and have been widely adopted by many schools.

The CMS data for this evaluation are the most comprehensive data of the three districts. Participating schools are identified separately through an anonymized school identifier, and teacher evaluation scores are available for all teachers in the district and can be linked to the student-teacher assignment data. In addition, CMS is the only district for which we observe some student demographic information (such as race). Due to these differences in data availability, any table or result that requires teacher evaluation scores will only include CMS data, as this is the only district providing this information.

Cabarrus County Schools (North Carolina)

Cabarrus County Schools (CCS) borders CMS and has been implementing OC models since the 2014–15 school year. Despite being less widespread as in CMS, there have still been more than 500 students reached as of the 2015–16 school year in our analysis sample.

The data that CCS provided for the evaluation are limited in comparison to the CMS data in two key ways. First, for privacy concerns, CCS declined to provide a school identifier; therefore, we cannot observe OC schools separately from non-OC schools or otherwise evaluate models making use of a school identifier (e.g., a school fixed- effects model). We can link students to teachers in CCS and we observe different OC models across teachers; hence, our exposure to intervention comes from direct exposure only, not through the school level. This is a disadvantage in that we are unable to control for differential conditions in schools that implemented OC relative to schools that did not, which our results from CMS suggest is an important consideration.

Syracuse City School District (New York)

Syracuse City School District (SCSD), like CCS, has been implementing OC models in schools since 2014–15. However, most SCSD OC schools are high schools and therefore are not included in tested grades/subjects and cannot be included in the analysis. Only two OC schools with many students assigned to OC models appear in the data, and only four schools in total appear in our data as implementing OC models.

SCSD was able to provide anonymized school identifiers for OC schools and therefore can be included in models with school fixed effects and school explanatory variables. However, SCSD, like CCS, declined to provide teacher evaluation data and therefore cannot be included in analyses that require such data.

Identifying OC Models

We display the process through which we identify students linked to OC teachers in Table 1. Because each district reports OC roles in slightly different ways, we use some discretion in creating categories of OC roles based on conversations with Public Impact. For example, we categorize as MCLs teachers in CMS with the job title "Teacher, Multi Classroom 1" and teachers in SCSD with staff role "MCL." Based on the data provided by the district, we create four possible OC roles: MCL, team teacher, blended learning teacher (BLT), and expanded impact teacher. One weakness of the data is that we only have direct links between students and the teachers they are assigned to. This means that for students exposed to the MCL model, we know which team teacher they are exposed to but we do not know which MCL they are indirectly exposed to. It is only for MCLs who are directly in charge of students that are linked to students in our data.[13]

[13] According to data collected by Public Impact, there are about six team teachers for each MCL across the three study districts.

Table 2 displays the number of students reached by each model in each district, along with the number of classrooms and schools. The bulk of students exposed to an OC model are those linked directly to team teachers (i.e., these students are indirectly exposed to MCL); recall that CMS contributes by far more observations than any other district. Our analytic sample includes 805,976 student-teacher observations (186,559 unique students and 7,028 unique teachers). Figure 1 displays the share of students in OC partner schools who are exposed to at least one of the OC models in a given year. There is a slight uptick in the 2013–14 school year with the early adopters in CMS first implementing OC models, leading to a large increase in 2014–15 as all districts implement. Figure 1 also shows that a substantial share of students in chosen schools are exposed to OC models.

Figure 2 displays math and reading test scores over time in the schools that implemented OC relative to the rest of the district. The y-axis represents standard deviations from test scores that have been normed to a mean of zero within each district, year, and tested grade and subject. Two main patterns are evident. First, OC schools are largely chosen from a subset of disadvantaged, low-performing schools, with mean test scores being well under test scores in other district schools. And, second, there is an uptick in math performance (and to a lesser extent reading) that occurs at the same time of the widespread implementation of OC models in the 2014–15 school year. As we describe below, models that do not account for this rise in school performance in treatment schools fully attribute this rise to the implementation of OC, which may not present an accurate picture of the program's impact.

We now provide a descriptive overview of the types of teachers and students who are in OC-exposed classrooms. First, we display sample means for teachers by district and OC role in Table 3. Looking at both leave-one-out value added forecasts of teacher performance (Chetty et al., 2014, described below in Section IV) and evaluation ratings, teachers were clearly systematically chosen for OC roles, especially for MCLs. For example, value added, leadership ratings, and facilitating learning ratings were all substantially higher than district average in the districts where each is available. BLTs and Expanded Impact teachers in CMS also were above average in these three categories. Although we do not have teacher evaluation data in other districts, MCLs also were substantially more effective as measured by value added in SCSD and CCS. Team teachers, on the other hand, were closer to district averages in our evaluation measures, which is expected as they are teachers intended to be coached.

Table 4 gives a sense of the types of students exposed to OC roles relative to other students in the district. As noted above when discussing Figure 2, the strongest pattern emerging is that students in schools that were selected to partner with OC came in with relatively weaker prior achievement, meaning that it is important to control for school setting in our models. In addition, even within schools, there appears to be differential selection in terms of OC exposure. In CMS, average prior student achievement in non-OC schools is 0.03 standard deviations above the overall district mean compared with 0.28 standard deviations below the mean for students in OC schools who did not themselves receive OC treatment and 0.44 standard deviations below the mean for students who did receive OC treatment. Although we cannot break out OC schools from non-OC schools in CCS due to lack of school identifiers, prior achievement of OC-exposed students also is substantially lower in that district.

IV. Methods

The main outcome of interest is student achievement on standardized tests in math and reading. In an ideal setting, schools would be randomly assigned to treatment and control groups. Under randomization, the treatment effect would simply be the difference in average outcomes between the treatment and control schools. However, schools were targeted for potential participation by their districts and then made a decision about whether to participate, meaning that treatment schools do not represent a random sample of schools. For example, at the four CMS 2013–14 pilot schools, the share of students eligible for free or reduced-price lunch (FRPL) in tested grades and subjects was 86% compared with 55% at the remaining 142 CMS schools. In addition, the share of Black students was substantially higher than that of the rest of the district—76% to 42%. Finally, math achievement was more than one third of a standard deviation lower in the pilot treatment schools. In addition, not all students within a school are exposed to OC, leading to potential nonrandom assignment within schools. To estimate OC impacts in the face of nonrandom selection among treatment schools, we estimate models using several different approaches, utilizing differences across classrooms, across schools, and over time as described in the following section.

Estimating Effect of OC Exposure

Our baseline analysis measures the difference in achievement of students in classrooms exposed to OC models and comparison students who are in other classrooms. Our approach follows similar studies of measuring the classroom achievement of students exposed to certain types of teachers, such as Teach For America and the New York City Teaching Fellows program (Boyd, Lankford, Loeb, & Wyckoff, 2006; Hansen, Backes, Brady, & Xu, 2014; Kane, Rockoff, & Staiger, 2008). We estimate the following equation:

$$y_{ist} = \beta_0 + \beta_1 y_{ist-1} + \beta_2 X_i + \beta_3 OC_i + \varepsilon_{ist} \text{ , (1)}$$

where y_{ist} indicates the score on a math or reading exam (with separate regressions for each) for student *i* in school *s* in year *t*, y_{ist-1} is a vector of cubic functions of prior-year test scores in math and reading, OC_i is an indicator for whether student *i* was taught by a teacher in an OC role in that subject, and X_i contains a vector of student *i*'s characteristics (only available in CMS), including race, gender, and eligibility for FRPL. In addition, ε_{ist} represents a randomly distributed error term. In all analyses where we observe school-level indicators (i.e., when not including CCS), standard errors are clustered at the school-cohort level to allow for arbitrary within-school clustering of the error terms (Chetty et al., 2014a).

The coefficient of interest, β_3, represents the average change in achievement for a student exposed to an OC classroom relative to a student who is not. Both experimental work and nonexperimental tests suggest that controlling for prior test scores as in Equation (1) is sufficient for estimating teacher effects with little bias (Bacher-Hicks, Kane, & Staiger, 2014; Chetty et al., 2014a; Kane, McCaffrey, Miller, & Staiger, 2013; Kane & Staiger, 2008). We include models in which we do and do not include controls for classroom average prior achievement. Because there are several distinct OC roles, we use dummies for each OC role instead of the combined OC coefficient displayed in Equation (1).

The counterfactual we wish to observe is how a given OC-exposed classroom would have performed if it had not been exposed to OC. Although we cannot directly observe this counterfactual, we perform a series of comparisons in which we expand on the basic set of controls given in Equation (1), each with its own strengths and weaknesses.

If a school elevates highly effective teachers to lead teams and reduces their direct teaching, it may need to hire replacements who are less effective. Those replacements may teach non-treated classrooms. Thus, it is possible that the performance of the control classroom is also affected by the intervention, whether because an effective teacher was removed from the classroom and replaced by a less effective teacher (in which case, OC estimates would be biased towards finding something positive for OC due some students being exposed to a less effective teacher than they would have otherwise) or because improvements by the team teachers spillover onto the other teachers (e.g., Jackson and Bruegmann, 2009). A potential additional confounding factor is the possibility of other simultaneous changes in overall school performance that could raise or lower the performance of both OC and non-OC classrooms within the school.

Thus, we make several different comparisons. First, we include average school-level prior performance. The comparison being made here is the achievement of students in OC classrooms relative to students in other schools with similar prior achievement, whether those schools are implementing OC or not. This alleviates the potential spillover problem by including classrooms in non-treated schools but does not overcome potential non-random selection of schools for treatment. For example, if performance in OC schools is uniformly lower than other schools, even after controlling for prior test scores (for example, due to systematic selection of low-performing schools for treatment), OC estimates would be biased towards finding negative effects.

Second, we use a school fixed-effects model. This model compares students in OC classrooms with other students in the same school who are not exposed to OC. These nonexposed students include both students prior to OC implementation and students present in the school after OC implementation. The advantage of this model is that it accounts for the selection of schools for treatment with school fixed effects and is not as susceptible to contamination from spillover because it includes classrooms from prior to treatment. On the other hand, as we discuss further below, a drawback of the school fixed-effects model is that any simultaneous change in school performance would be attributed to OC even if it were not caused by OC.

Third, we use a school-by-year fixed-effects model. This compares students in OC classrooms to other students in the same school in the same year. The limitation of this model is that if, for example, there are spillovers from OC classrooms into non-OC classrooms, one would expect the school-by-year fixed-effects model to understate the true OC impact because all students in the comparison group of the same school in the same year could potentially receive these spillovers. Because these estimates net out the potential impact of spillovers, we think of these as conservative estimates of the true impact of OC; however, previous work has found that the scope for spillovers across teachers is modest (e.g., Jackson and Bruegmann, 2009. On the other hand, if there were a concurrent rise in school performance in OC schools due to factors other than OC, school-by-year fixed effects would avoid attributing this rise to OC.

Finally, we include a student fixed-effects model, in which we compare a student's achievement in years in which he or she is exposed to OC with years in which he or she is not.[14] As with the school fixed effects model, if the collective achievement of students within a school rises for reasons unrelated to OC, this model would overstate the impact of OC.

Estimating Teacher Effectiveness

In some models, we control for the estimated effectiveness of a teacher in order to measure the extent to which OC effects are driven by exposure to effective teaching. In order to forecast teacher effectiveness, we use the leave-one-out method for forecasting value added developed by Chetty et al. (2014a). Thus, a teacher's forecast for a given year uses all years of data other than that year. We make a modification to Chetty et al. (2014a) in that we do not use any year that a teacher participates in OC in the forecast. For example, if a teacher is flagged as a team teacher in 2013–14 and 2014–15 and we are forecasting teacher effectiveness in 2014–15, we would only use years prior to 2013–14 to construct the forecast. Leaving out all OC years avoids conflating underlying teacher quality and possible changes in student achievement that are driven by the OC models.

V. Results

Main Results

[14] We also experiment with restricting the set of control schools based on propensity score matching for CMS schools. In results not shown but available from authors, we first use a school's prior math and reading scores, as well as the percentage of its students who are Black and Hispanic, to select a closest match for each treated school. We then estimate OC effects analogous to the results shown in Tables 5 and 6, and find similar results with the restricted set of comparison schools. As described in further detail below, we also conduct a placebo test in which we randomly generate "placebo" treated teachers.

We present our main results for math in Table 5 and for reading in Table 6. Each table has four panels of results, one for all districts pooled together and three additional panels, one for each district. Because CMS is by far the largest district, the pooled results are almost always driven entirely by CMS. However, we show results from each district for completeness in Tables 5 and 6.

Each column of Tables 5 and 6 displays results from a different set of controls. Taking CMS in Table 5, for example, the first column shows results when only controlling for student prior test scores and whether a school was ever selected for OC or Project L.I.F.T. The first panel displays the result from a regression that pools together all OC models, which is significant and positive in all specifications. In the second panel, which disaggregates by OC role, coefficients for all roles in all specifications are positive in math, although the Expanded Impact and BLT coefficients are generally not statistically significant due to smaller sample sizes. Column 2 shows results when controlling for student-level demographics, which are only available in CMS, and results are broadly similar to Column 1. Column 3 controls for classroom-level prior test scores, which leads to decreases in magnitude for the OC model coefficients.[15] Column 4 adds school-level average prior scores in CMS and SCSD. Because we do not have school identifiers in CCS, Column 4 is left blank for CCS and the "All" panel is a pooled result for CMS and SCSD. This also is true for Column 5, where school fixed effects are included instead of school prior test scores. When using school fixed effects, only the team teacher coefficient remains statistically significant.

Column 6 adds school-by-year fixed effects. As we show below, this is the model that performs best in our specification checks, making it our preferred specification. In this specification, students in expanded impact or team teacher classrooms have significant and positive results, although note that the sample size for expanded impact is extremely small. For team teachers, the effect size is 0.11

[15] It is possible that controlling for classroom average prior test scores absorbs some of the true difference in teacher effectiveness (e.g., Goldhaber et al., 2014).

standard deviations in Column 6. In Column 7 we use a student fixed-effects model, controlling for only student fixed effects and class average scores in reading and math. Across all models for CMS and the pooled sample, students taught by team teachers have statistically significant learning gains in math. Results are broadly similar whether using student fixed effects (Column 7), a type of school fixed effects (Columns 5 and 6), or no fixed effects (Columns 1–4). This is notable because the average effectiveness of team teachers (prior to participation in OC) is identical to non-OC teachers in the same school in the largest district, CMS, and similar in the other two districts.

Turning to reading in Table 6, we see that the patterns for students taught by team teachers are largely similar, although smaller in magnitude. However, in contrast to each of the other specifications, our preferred specification in Column 6 finds results for reading to be slightly negative and not statistically significant. In addition, students exposed to blended learning teachers in reading now have large and negative impacts in some models.[16] This may not be surprising given that students have been found to have worse learning outcomes in other settings where face-to-face learning was replaced with online learning (e.g., Xu & Jaggars, 2013).[17]

Results by Estimated Teacher Effectiveness

In this section, we set out to answer two questions. First, whether the positive results for students directly exposed to MCLs can be fully explained by MCLs being more effective on average. Because MCLs are generally selected from more effective teachers (see, for example, Table 3), we expect them to be strong classroom performers, though we are not sure whether the MCL role will have

[16] In reading, all BLT in our sample are in elementary grades.
[17] The OC rollout in CMS was somewhat different across Project L.I.F.T. and non-L.I.F.T. schools in that L.I.F.T.'s OC positions have somewhat different job descriptions and stipend levels. When breaking out the results of Tables 5 and 6 into L.I.F.T. and non-L.I.F.T. schools, we find that the achievement gains associated with team teacher exposure are higher for non-L.I.F.T. schools in math and higher for L.I.F.T. schools in reading in most specifications. However, in our preferred specification (column 6), point estimates for team teachers are identical in L.I.F.T and non-L.I.F.T in math and not statistically significant in reading.

a systematic difference in performance associated with it independent of MCLs' prior effectiveness. Second, we investigate the extent to which the stronger or weaker team teachers (as measured by value added: for details on how we forecast value added, see the methods section above) disproportionately benefit from participating in the MCL model. For example, relatively weak teachers could greatly benefit from having an MCL's coaching and mentorship, while those who were already fairly competent may see little to no benefit from that relationship. To investigate these relationships, we estimated the same models as above while including indicator variables representing a teacher's forecasted value-added estimate put him or her in the top or bottom quartile of value-added performance, and then interact these estimates with the team teacher flag.

Results are shown in Table 7 (math) and Table 8 (reading). From here onward, we omit specification (2) from the tables because it can only be performed in one district. As expected, the point estimates for being in either the top or bottom quartile of forecast value-added are statistically significant in the expected directions across all models. In the basic model that controls for only student and classroom prior test scores (column 3), students taught by a math teacher in the top quartile of the forecasted value-added distribution score 0.16 standard deviations higher than comparison students in similar contexts, and students taught by bottom-quartile teachers score 0.13 standard deviations lower. The range for reading is somewhat smaller, which is consistent with the variance of teacher effects being larger in math than reading (e.g., Chetty et al., 2014). Our interest in these results, however, is not in these forecast value-added estimates themselves, but in whether the variables on the various OC roles are significant different even in the inclusion of these variables.

In regards to our first question on the independent effect of MCLs, controlling for teachers' expected performance does not fully explain the advantage associated with learning in a classroom led by an MCL, as almost all of the MCL point estimates across specifications are still positive and large in both Tables 7 and 8. This would suggest that serving as an MCL itself may improve these teachers' performance above and beyond that which they've demonstrated in the past. Yet, it is hard to know how much to read into these estimates given the limited sample of MCLs in charge of their own classrooms while serving as MCLs.[18]

For the second question on which team teachers benefit from the MCL's coaching, we look to the interactions of these top and bottom quartile indicators with team teacher.[19] In math, the benefits associated with being in a team teacher-led classroom are larger for team teachers who were already fairly strong, although this is not statistically significant due to large standard errors. Using the results from Column 6 of Table 7, a student taught by a teacher in the top quartile of the value-added distribution would be expected to score 0.11 standard deviations higher, with an additional 0.19 standard deviations if that teacher is a team teacher (combining both the 0.08 main team teacher estimate and the 0.11 interacted top-quartile estimate). Team teachers from the bottom-quartile, on the other hand, would see a smaller bump in performance of 0.02 standard deviations (combining the 0.08 team teacher estimate with the -0.06 interacted bottom-quartile estimate). These results are suggestive of teachers being under an MCL accentuating their skills or weaknesses in math, although we again stress that these are imprecisely estimated. In reading, on the other hand, in most specifications there appears to be no link between prior teacher effectiveness among team teachers and student outcomes. One potential explanation for this diverging pattern in math and reading, if any, is that because the range of teacher effects is wider in math than reading (as evidenced by the variance of teacher effects being larger in math), MCLs are better able to raise the productivity of high-ability team teachers in math.

---

[18] Results are similar when using a continuous measure of leave-one-out value added rather than indicators for the top and bottom quartiles.

[19] We do not display interactions between teacher quality and the other OC models for two reasons that make models very noisy at best or impossible to estimate at worst. First, very few teachers in OC models other than team teachers have leave-one-out value-added estimates. Second, there is very little variation in value added among OC teachers who were not team teachers as they were selected among a pool of effective teachers. For example, every single MCL with a valid leave-out-out value-added estimate landed in the top quartile of the value-added distribution. Although our estimate of value added is not what is used by the districts and (presumably) was not used to select teachers for OC roles, it does appear to be highly correlated with what went into the selection process.

Placebo Tests

In a section above, we discussed how the general rise in test scores in OC treatment schools could lead to an upward bias in OC estimates if this rise is not fully caused by OC. To test the extent of this potential bias, we randomly select non-OC teachers from within treated schools in 2014-15 and 2015-16 and call them "placebo" OC teachers.[20] We then re-run our models from Tables 5 and 6 with an additional role of "placebo" OC teacher. Because these teachers are not real OC teachers and represent no genuine treatment, we expect the point estimates on these placebo OC teachers to be statistically indistinguishable from zero.

Results are shown in Table 9. For both math and reading, the point estimates for placebo OC teachers are strikingly similar to the point estimates for team teachers, with the exception of the school-by-year fixed effects model in Column 6.[21] In math, Column 6's point estimate for the placebo teacher's is -0.02 and not statistically significant, while for reading it is 0.05 and statistically significant. Because the Column 6 specification passes the placebo test in math and fails it in reading to a lesser degree than the other specifications, we take these as our preferred estimates.

Put together, the results from Table 9 suggest that many of the positive OC point estimates are due to schoolwide improvements that are not necessarily due to OC, though as we discuss further below, it is also possible to interpret these as positive spillovers from the implementation of OC models. Additional support for general schoolwide improvement is provided in Figure 3, where we plot the average test scores of treated versus non-treated students within OC schools. In both math and reading,

[20] Specifically, we assign every non-OC teacher within a treatment school a random number with the uniform() command in Stata and then take teachers whose random value lies below a threshold chosen so that there are about as many "placebo" teachers as there are team teachers within the school.

[21] The point estimates for the various OC models in Table 9 are slightly different than in Tables 5 and 6 because the set of comparison teachers within OC schools has changed due to the removal of some OC teachers from the comparison teacher pool in order to count as placebo teachers.

we see evidence of schoolwide improvement among non-treated students that pre-dated OC, especially in reading.

To test this further, we re-estimate the models in Tables 5 and 6 but only include the final two years of data, 2014-15 and 2015-16. While the team teacher results for math are still positive and significant in both models with school fixed effects (and positive in all models), the team teacher results for reading are now very small and not statistically significant across the board, which is consistent with the reading results in Table 6 being driven by schoolwide performance being higher in these latter two years than earlier in the sample.[22]

After demonstrating that some of the positive findings in Table 6 are due to overall school improvement in treatment schools, the question is the extent to which it is possible that these improvements could be driven in part by OC implementation in the school (even for classrooms that were not exposed). If improvement in non-OC classrooms had nothing to do with OC, then we should prefer the Column 6 estimates and conclude that OC models had a positive effect in math and no effect in reading (and a negative effect in the case of blended learning in reading). However, if some of this general schoolwide improvement in treatment schools was because of a school's involvement with OC, then we should place more weight on the other specifications, which find students taught by team teachers scoring consistently higher in both subjects. While one can point to examples of spillovers between teachers (e.g., Jackson & Bruegmann, 2009) as evidence that these spillovers are possible, it is difficult to know in practice the extent to which this occurred in treatment schools. In short, while we can say that implementation of OC at a partner school was associated with an increase in performance in math and reading, we cannot confidently say that the increase was caused by OC in reading.

[22] In results available from the authors, we also experimented with collapsing the data to the school-grade-year-subject level and regressing test scores on the share of each OC model in the cell. Results are broadly similar to the school fixed effects model in Tables 5 and 6 but standard errors are large.

Mechanisms

One weakness of the data in all districts is that we do not have links between team teachers and MCLs. We know which students are taught by which team teachers; however, all team teachers are advised by an MCL, and we do not know who those MCLs are. In addition, even if we did know which MCLs were in charge of which team teachers, it would be hard to investigate, for example, the extent to which MCL quality drives team teacher success for two reasons. First, because many MCLs do not teach students in many years of our data coverage, we do not have value-added style estimates for them. And, second, because most, if not all, MCLs are high in quality, there is no variation in quality to exploit in order to investigate the association between MCL quality and team teacher performance.

Why do students of team teachers perform so well in math? They are not being directly exposed to peer teachers who are any more effective, on average. And this question is important because it is through team teachers that the large majority of students are being reached. In addition to the estimated learning gains, the observation ratings of team teachers appear to improve under coaching as well: In years in which teachers who were ever team teachers were not coached, 39% of teacher-year observations received a "distinguished" or "accomplished" rating in facilitating learning, compared with 45% of teacher-year observations in the years in which teachers were coached.[23] However, as discussed in Section II above, it is impossible to separate the extent to which this improvement is because mentor teachers are chosen based on their effectiveness versus whether we are observing a mentorship effect (e.g., Kraft et al., 2016). Because there is very little variation in the effectiveness of mentor teachers, we cannot make a credible attempt to distinguish between the two in our data. In addition, part of the improvement may also be due to the team's structure enabling more continuous instruction when teachers are absent. Based on conversations with Public Impact, MCLs can directly step in when one of the team teachers is absent or support the instruction of a substitute. With empirical evidence

[23] As shown in Table 3, the state average for non-OC teachers is 48%.

suggesting typical substitutes offer low-quality instruction (Herrmann and Rockoff, 2012), this may be an additional way through which MCLs may affect a team teacher's classroom.

VII. Conclusion

This paper presents an evaluation of Public Impact's Opportunity Culture initiative in three pilot school districts, exposing more than 15,000 students to at least one of the program's models. To our knowledge, this is the first empirical study of an implemented treatment explicitly designed to manipulate teacher roles to give effective teachers a broader reach across students—either directly or indirectly through coaching peer teachers.

This program is significant in that it represents a new and qualitatively different approach at promoting teacher quality in the workforce, distinct from efforts to directly incentivize strong performance or punish weak performance. Rather, by attaching past performance to opportunities to work in different roles and then offering different compensation accordingly, the OC models offer both the promise of upward career mobility to teachers without having to leave teaching and less political resistance generally associated with most merit-pay programs. Beyond the benefits to highly effective teachers, the MCL model offers professional support for team teachers, which may be viewed as a benefit of working in an OC school and thus may enhance the school's climate and teacher retention. Though questions of teacher perceptions, school climate, and teacher retention are beyond the scope of this paper, the evidence of the program's outcomes we present here suggests both reason for optimism and skepticism for these alternate staffing strategies.

We find that OC schools both elevated their most effective teachers to OC roles (as intended) and significantly improved exposed students' performance in math. The evidence points to the multiclassroom leader model as the key model in helping to produce those gains. The theory of action behind the multiclassroom model relies on instructional coaching from an effective classroom teacher to help average or weak teachers improve their practice, and the evidence we find in math is generally consistent with prior studies demonstrating the effectiveness of intensive, personalized coaching (e.g., Kraft et al., 2017). By elevating the status of some of the most effective teachers in these schools, coached math teachers demonstrate significant improvements in the classroom, achieving a net overall improvement in teacher quality.

Less optimistically, however, we also find null results for the multiclassroom leader model for reading in our preferred specification and significant and negative results associated with OC's blended learning model in some specifications. The blended learning model helps effective teachers find time to reach more kids through the use of learning stations and online instruction overseen by paraprofessionals. Though this strategy had a lower uptake in the three pilot districts we studied, the estimates we report here are discouraging because a known effective teacher was utilized in a different role that turned out to be less effective for all students than what we would have otherwise expected. Modifying this blended learning role in the future may be able to offer different, more promising outcomes for either teachers or students, but the evidence based on its implementation here should give this model pause.

Beyond the two models implemented in the districts we studied, OC promotes the use of three others in its initiative. These other models continue to offer promise in theory (e.g., Goldhaber et al., 2013; Hansen, 2013), though to our knowledge they have not been demonstrated in practice in isolation. Future research should look to evaluate these models in real-life settings, if and when they are implemented.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of labor Economics*, *25*(1), 95-135.

Bacher-Hicks, A., Kane, T.J., & Staiger, D.O. (2014). Validating teacher effect estimates using changes in teacher assignments in Los Angeles. *NBER working paper No. 20657.*

Backes, B., & Hansen, M. (2018). The Impact of Teach for America on Non-Test Academic Outcomes. Forthcoming, *Journal of Education Finance and Policy.*

Blazar, D., & Kraft, M. A. (2015). Exploring mechanisms of effective teacher coaching: A tale of two cohorts from a randomized experiment. *Educational Evaluation and Policy Analysis*, *37*(4), 542-566.

Boyd, Donald J., Hamilton Lankford, Susanna Loeb, and James Wyckoff, (2011). "Teacher Layoffs: An Empirical Illustration of Seniority versus Measures of Effectiveness," *Education Finance and Policy* 6 (3): 439-454.

Chetty, Raj, John Friedman, and Jonah Rockoff. (2014a). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104(9): 2593–2632

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, *104*(9), 2633-2679.

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, 34(2), 267-297.

Goldhaber, Dan, Dominic J. Brewer, and Deborah J. Anderson. (1999). "A Three-Way Error Components Analysis of Educational Productivity," *Education Economics,* 7(3): 199-208.

Goldhaber, Dan, James Cowan, and Joe Walch. (2013). "Is a Good Elementary Teacher Always Good? Assessing Teacher Performance Estimates Across Subjects," *Economics of Education Review*, 36: 216-228.

Goldhaber, D., DeArmond, M., & DeBurgomaster, S. (2011). Teacher attitudes about compensation reform: Implications for reform implementation. *ILR Review*, *64*(3), 441-463.

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, *44*(5), 293-307.

Goldhaber, D., Walch, J., & Gabele, B. (2014). Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. *Statistics and Public Policy*, 1(1), 28-39.

Hansen, Michael. (2013). "Right-Sizing the Classroom: Making the Most of Great Teachers," *Thomas B. Fordhom Institute*. http://edex.s3-us-west-2.amazonaws.com/publication/pdfs/20131119-Right-Sizing-the-Classroom-Making-the-Most-of-Great-Teachers-FINAL.pdf

Hansen, M. (2016). "The Future of Student Growth Models in Managing the Teacher Workforce: Less Direct, More Indirect" in *Student Growth Measures in Policy and Practice*, Kimberly Kappler Hewitt and Audrey Amrein-Beardsley (eds.). New York, NY: Palgrave Macmillan.

Hansen, M., Backes, B., Brady, V., & Xu, Z. (2014). Examining spillover effects from Teach For America corps members in Miami-Dade County public schools. *National Center for Analysis of Longitudinal Data in Education Research Working Paper*, *113*.

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

Harold, Benjamin. (2014). Growing Pains for Rocketship's Blended-Learning Juggernaut. *Education Week*. https://www.edweek.org/ew/articles/2014/01/21/19el-rotation.h33.html

Herrmann, M. A., & Rockoff, J. E. (2012). Worker absence and productivity: Evidence from teaching. *Journal of Labor Economics*, *30*(4), 749-782.

Hess, Frederick. (2009). "How to Get the Teachers We Want," *Education Next*, 9(3).

Ingersoll, R. M., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of educational research*, *81*(2), 201-233.

Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., Hock, H., & Hansen, M. (2013). Access to Effective Teaching for Disadvantaged Students. NCEE 2014-4001. *Institute of Education Sciences*.

Isenberg, E., Max, J., Gleason, P., Johnson, M., Deutsch, J., & Hansen, M. (2016). Do low-income students have equal access to effective teachers?

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1(4), 85-108.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education review*, *27*(6), 615-631.

Kane, T. J., & Staiger, D. O. (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation (No. w14607). *National Bureau of Economic Research*.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. *Seattle, WA: Bill and Melinda Gates Foundation*.

Kowal, J., & Brinson, D. (2011). Beyond Classroom Walls: Developing Innovative Work Roles for Teachers. *Center for American progress*.

Kraft, M. A., Blazar, D., & Hogan, D. (2016). The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence.

Loeb, S., Miller, L. C., & Wyckoff, J. (2015). Performance screens for school improvement: The case of teacher tenure reform in New York City. *Educational researcher*, 44(4), 199-212.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational evaluation and policy analysis*, *26*(3), 237-257.

Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. (2016). *Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data* (No. w21986). National Bureau of Economic Research.

Public Impact. (2012). Redesigning schools to reach every student with excellent teachers: Financial planning summary. Chapel Hill, NC. Retrieved from http://opportunityculture.org/wp-content/uploads/2012/05/Financial_Planning_Summary-Public_Impact.pdf

Rockoff, J. E. (2008). *Does mentoring reduce turnover and improve skills of new employees? Evidence from teachers in New York City* (No. w13868). National Bureau of Economic Research.

Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, *72*(2), 104-122.

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

Taylor, E. (2015). New technology and teacher productivity. *Unpublished manuscript. Cambridge, MA: Harvard Graduate School of Education*.

Xu, D., & Jaggars, S. S. (2013). The impact of online learning on students' course outcomes: Evidence from a large community and technical college system. *Economics of Education Review*, *37*, 46-57.

Figure 1: **Treated Students in OC Partner Schools by Year**

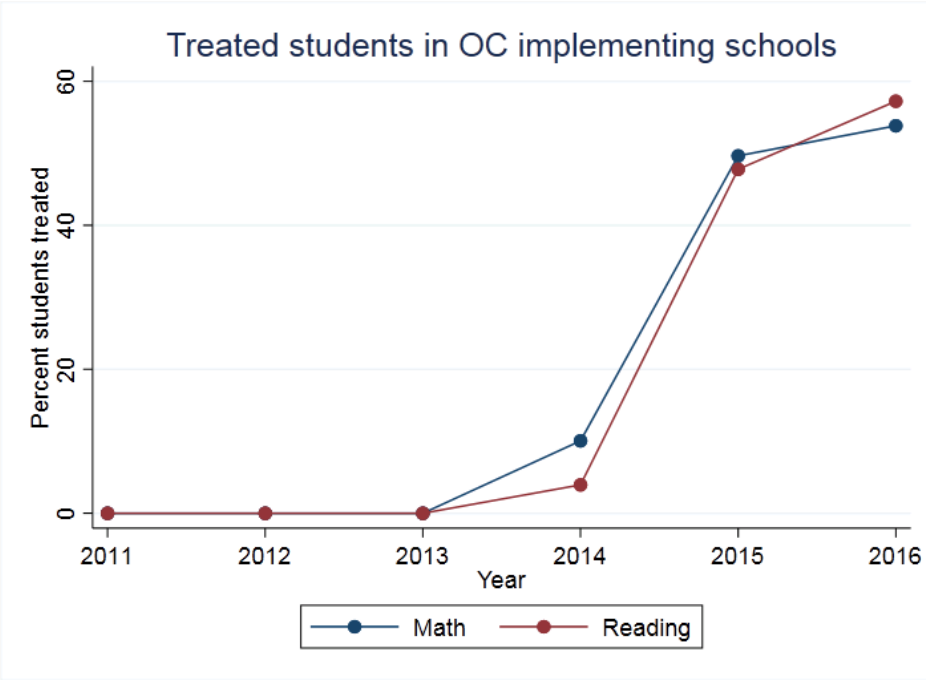

Figure 2: **Average Standardized Test Scores in OC Partner Schools and Other District Schools by Year**

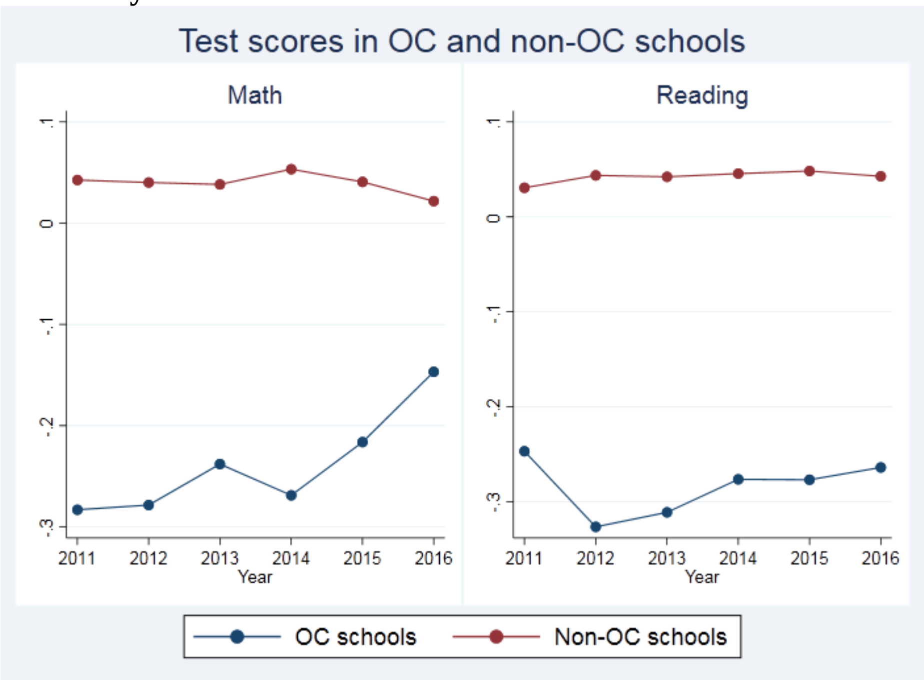

Figure 3: Average Standardized Test Scores in OC Partner Schools by Year – Treated Students versus Non-Treated Students

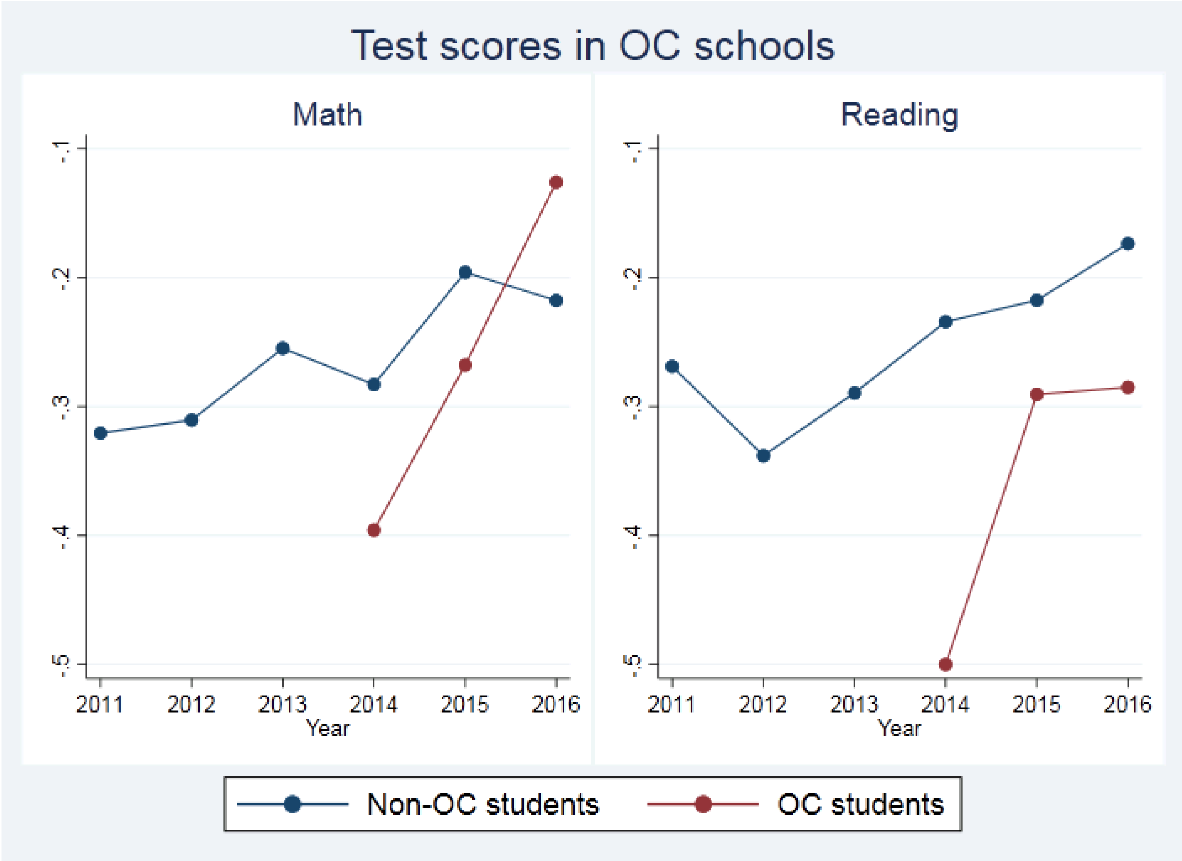

Table 1: OC Role Classification Based on Administrative Data

District	Information	Our Categorization
CMS	Job title: "Tchr, Reach Multi Classroom 1"	Expanded Impact
CMS	Job title: "Tchr, Reach Multi Classroom 2"	Expanded Impact
CMS	Job title: "Tchr, Sr. Reach"	Expanded Impact
CMS	Job title: "Tchr, Master Reach"	Expanded Impact
CMS	Job title: "Tchr, Reach Team"	Team Teacher
CMS	Identified as served under MCL	Team Teacher
CMS (LIFT)	Job title: "Teacher, Multi Classroom 1"	MCL
CMS (LIFT)	Job title: "Teacher, Multi Classroom 2"	MCL
CMS (LIFT)	Job title: "Teacher, Blended Learning - PL"	BLT
CMS (LIFT)	Job title: "Teacher, Expanded Impact - PL"	Expanded Impact
CMS (LIFT)	Flagged as OC LIFT and no other role found	Team Teacher
SCSD	Staff role: "Mcl"	MCL
SCSD	Staff role: "Mcl2"	MCL
SCSD	Staff role: "Team Teacher"	Team Teacher
CCS	Teacher role: "MCL"	MCL
CCS	Teacher role: "BLT"	BLT
CCS	Teacher role: "Team Teacher"	Team Teacher

Table 2: OC Exposure by Role and District

	MCL	Team Teachers	Extended Impact	BLT
Charlotte				
Number of students	160	9601	2609	1217
Number of classrooms	8	387	128	48
Number of teacher-years	6	233	43	19
Number of school-years	4	38	18	10
Cabarrus				
Number of students	2	530	0	16
Number of classrooms	1	24	0	6
Number of teacher-years	1	22	0	5
Number of school-years	.	.	.	.
Syracuse				
Number of students	303	537	0	0
Number of classrooms	22	41	0	0
Number of teacher-years	11	24	0	0
Number of school-years	5	3	0	0

Note: Cabarrus did not provide school-level identifiers so we cannot report the number of unique schools. Our analytic sample includes 805,976 student-teacher observations (186,559 unique students, 7,028 unique teachers).

Table 3: Teacher Characteristics

	CMS						CCS				SCSD				
	Non-OC	Non-OC in OC School	MCL	BLT	EI	TT	Non-OC	MCL	BLT	TT	Non-OC	Non-OC in OC School	MCL	BLT	TT
Leave-one-out VA	0.02	0.00	.	0.07	0.21	0.00	0.02	0.09	.	-0.01	-0.02	-0.03	0.34	.	0.01
Leadership: top (%)	51	45	100	100	84	43	.	.	.	.	.	.	.	.	.
Diversity: top (%)	30	29	0	50	27	24	.	.	.	.	.	.	.	.	.
Content: top (%)	24	23	0	50	23	18	.	.	.	.	.	.	.	.	.
Facilitating learning: top (%)	48	41	100	73	78	45	.	.	.	.	.	.	.	.	.
Reflection: top (%)	25	27	0	63	32	22	.	.	.	.	.	.	.	.	.
Black (%)	27	37	17	26	37	49	.	.	.	.	.	.	.	.	.
Hispanic (%)	2	2	0	0	0	1	.	.	.	.	.	.	.	.	.
Female (%)	84	84	50	89	91	79	.	.	.	.	.	.	.	.	.
Total unique teachers	5511	700	3	8	29	193	792	1	3	22	453	170	7	0	21
Total teacher-year observations	13388	1506	6	19	43	233	2153	1	5	22	1000	326	11	0	24

Notes: Top = accomplished or distinguished evaluation rating (top two categories). Missing means no information or insufficient information. Columns show averages for non-OC teachers and for Opportunity Culture models multi-classroom leader (MCL), blended learning teacher (BLT), extended impact, and team teacher. “Leave-one-out VA” refers to forecasted value added; see text.

Table 4: Student Characteristics

	Charlotte			Cabarrus		Syracuse		
		OC schools					OC schools	
	Non-OC schools	Not exposed	Exposed	Non-OC students	OC students	Non-OC schools	Not exposed	Exposed
Prior Math Achievement	0.03	-0.28	-0.44	0.03	-0.21	0.04	-0.1	0.04
	(0.98)	(0.95)	(0.91)	(0.84)	(0.95)	(0.93)	(0.95)	(0.95)
Prior Reading Achievement	0.07	-0.27	-0.4	0.03	-0.11	0.05	-0.08	0.04
	(0.99)	(0.91)	(0.88)	(0.99)	(0.92)	(0.97)	(0.95)	(0.94)
Student-yr observations	340815	49076	10483	54997	546	14668	5210	838
Students	112938	24147	8284	31649	546	5818	2515	667
School-yr observations	867	72	36	.	.	107	13	5
Schools	162	0	18	.	.	25	0	3
Black Students (%)	40	53	65	.	.	.	.	.
Hispanic Students (%)	19	16	17	.	.	.	.	.
Female Students (%)	49	51	49	.	.	.	.	.

Notes: for OC schools, "not exposed" columns represent students who were in OC schools but not exposed to OC models, while "exposed" represent students who were exposed to OC models. For Cabarrus, because we do not have school identifiers, we can only report statistics for exposed students versus non-exposed students.

Table 5: Main Results for Math

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			All Districts: OC models pooled				
All OC models (pooled)	0.14***		0.11***	0.12***	0.12***	0.09**	0.17***
	(0.04)		(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
			All Districts				
MCL (direct)	0.37***		0.35***	0.34***	0.27***	0.28**	0.25***
	(0.08)		(0.07)	(0.07)	(0.09)	(0.14)	(0.10)
Team Teacher on MCL-led team	0.14***		0.11**	0.12***	0.14***	0.11**	0.17***
	(0.05)		(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
BLT	0.1		0.03	0.04	0.01	0.06	0.07
	(0.08)		(0.08)	(0.09)	(0.05)	(0.06)	(0.12)
Expanded Impact	0.17*		0.12	0.13	0.07	0.02	0.21*
	(0.10)		(0.09)	(0.09)	(0.05)	(0.06)	(0.12)
			Charlotte-Mecklenburg Schools				
Team Teacher on MCL-led team	0.17***	0.17***	0.13***	0.14***	0.16***	0.13**	0.18***
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	-0.05
BLT	0.11	0.12	0.04	0.05	0.02	0.06	0.07
	(0.08)	(0.08)	(0.08)	(0.08)	(0.05)	(0.06)	-0.12
Expanded Impact	0.17*	0.15*	0.13	0.13	0.07	0.02	0.22*
	(0.10)	(0.08)	(0.09)	(0.09)	(0.06)	(0.06)	-0.12
			Cabarrus County Schools				
Team Teacher on MCL-led team	-0.06*		0.10***				0.13***
	(0.04)		(0.03)				(0.04)
			Syracuse County School District				
Team Teacher on MCL-led team	-0.18		-0.23	-0.23*	-0.24*	-0.22	-0.2
	(0.14)		(0.14)	(0.12)	(0.12)	(0.16)	(0.20)
Prior test scores	x	x	x	x	x	x	
Race		x					
Classroom prior tests			x	x	x	x	x
School prior tests				x			
School FE					x		
School-year FE						x	
Student FE							x

Notes: estimates show coefficients of each Opportunity Culture model: multi-classroom leader (MCL), blended learning teacher (BLT), extended impact, and team teacher. The dependent variable is standardized test scores in math. “All districts” estimates in columns 4 - 6 do not include Cabarrus County Schools because of the unavailability of school identifiers. Standard errors clustered at the school-cohort level. MCL coefficients for individual districts suppressed due to very small sample sizes.

Table 6: Main Results for Reading

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All Districts: OC models pooled							
All OC models (pooled)	0.06***		0.04***	0.04***	0.05***	-0.02	0.08***
	(0.02)		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
All Districts							
MCL (direct)	0.20***		0.17**	0.17**	0.17**	0.13*	0.08
	(0.06)		-0.08	(0.08)	(0.07)	(0.07)	(0.06)
Team Teacher on MCL-led team	0.06**		0.05***	0.05**	0.05***	-0.03	0.07***
	(0.02)		-0.02	(0.02)	(0.02)	(0.02)	(0.02)
BLT	-0.07		-0.15**	-0.14**	-0.08	-0.05	-0.08***
	(0.05)		-0.06	(0.06)	(0.06)	(0.03)	(0.03)
Expanded Impact	0.05		0.03	0.04	0.05*	-0.02	0.11***
	(0.03)		-0.03	(0.03)	(0.03)	(0.03)	(0.02)
Charlotte-Mecklenburg Schools							
Team Teacher on MCL-led team	0.08***	0.08***	0.06***	0.06***	0.07***	-0.02	0.09***
	(0.02)	-0.02	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
BLT	-0.07	-0.05	-0.15**	-0.13**	-0.08	-0.04	-0.08***
	(0.05)	-0.05	(0.06)	(0.06)	(0.06)	(0.03)	(0.03)
Expanded Impact	0.05	0.05*	0.04	0.04	0.05*	-0.02	0.11***
	(0.03)	-0.03	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Cabarrus County Schools							
Team Teacher on MCL-led team	-0.06*		0.02				0.04
	(0.04)		(0.03)				(0.04)
Syracuse County School District							
Team Teacher on MCL-led team	-0.05		-0.15	-0.14	-0.17	-0.17**	-0.20*
	(0.10)		(0.09)	(0.10)	(0.11)	(0.07)	(0.11)
Prior test scores	x	x	x	x	x	x	
Race		x					
Classroom prior tests			x	x	x	x	x
School prior tests				x			
School FE					x		
School-year FE						x	
Student FE							x

Notes: see notes from Table 5.

Table 7: Interactions With Teacher Effectiveness, Math

	(1)	(3)	(4)	(5)	(6)	(7)
MCL (direct)	0.42***	0.39***	0.38***	0.32***	0.34**	0.28**
	(0.09)	(0.08)	(0.08)	(0.10)	(0.14)	(0.12)
Team Teacher on MCL-led team	0.12	0.10*	0.12**	0.16***	0.08	0.15***
	(0.07)	(0.06)	(0.06)	(0.06)	(0.07)	(0.05)
BLT	0.17**	0.09	0.1	0.05	0.1	0.14
	(0.08)	(0.08)	(0.08)	(0.05)	(0.06)	(0.12)
Expanded Impact	0.14**	0.10*	0.10*	0.09	0.05	0.16**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
Top Quartile VA	0.20***	0.16***	0.15***	0.10***	0.11***	0.13***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Bottom Quartile VA	-0.18***	-0.13***	-0.12***	-0.10***	-0.11***	-0.10***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
No VA estimate	-0.10***	-0.08***	-0.08***	-0.08***	-0.09***	-0.05***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Team Teacher # Top Q	0.07	0.12	0.11	0.12	0.11	0.16
	(0.08)	(0.10)	(0.10)	(0.11)	(0.09)	(0.13)
Team Teacher # Bot Q	0.02	-0.02	-0.04	-0.09	-0.06	-0.01
	(0.13)	(0.12)	(0.11)	(0.10)	(0.09)	(0.12)
Team Teacher # Missing VA	0.05	0.03	0.01	-0.01	0.06	0.02
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)
Prior test scores	x	x	x	x	x	
Race						
Classroom prior tests		x	x	x	x	x
School prior tests			x			
School FE				x		
School-year FE					x	
Student FE						x

Notes: value added (VA) forecasted constructed using leave-one-out estimate (Chetty et al., 2014), excluding all years for which a teacher participated in Opportunity Culture. We do not estimate interactions with models other than team teachers because of the lack of variation in value added among teachers participating in those models. Standard errors clustered at the school-cohort level. We omit model (2) from this table as it pools together multiple districts and we only observe race in one district.

Table 8: Interactions With Teacher Effectiveness, Reading

	(1)	(3)	(4)	(5)	(6)	(7)
MCL (direct)	0.24***	0.20***	0.19**	0.20***	0.16**	0.09
	(0.06)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
Team Teacher on MCL-led team	0.05	0.04	0.03	0.04	-0.03	0.07***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
BLT	-0.06	-0.13**	-0.11*	-0.06	-0.02	-0.06*
	(0.05)	(0.06)	(0.06)	(0.06)	(0.03)	(0.03)
Expanded Impact	0.08**	0.07**	0.07***	0.08***	0.01	0.13***
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Top Quartile VA	0.15***	0.09***	0.08***	0.07***	0.08***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Bottom Quartile VA	-0.19***	-0.08***	-0.07***	-0.06***	-0.07***	-0.02***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
No VA estimate	-0.06***	-0.04***	-0.04***	-0.04***	-0.04***	-0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Team Teacher # Top Q	-0.15**	-0.03	0.00	-0.02	-0.06	-0.01
	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)
Team Teacher # Bot Q	0.10	0.03	0.04	0.02	0.02	-0.05
	(0.07)	(0.04)	(0.04)	(0.04)	(0.06)	(0.05)
Team Teacher # Missing VA	0.04	0.03	0.04	0.03	0.02	0.03
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Prior test scores	x	x	x	x	x	
Demographics						
Classroom prior tests		x	x	x	x	x
School prior tests			x			
School FE				x		
School-year FE					x	
Student FE						x

Notes: see notes from Table 7.

Table 9: Placebo Treated Teachers

	(1)	(3)	(4)	(5)	(6)	(7)
		Math				
BLT	0.11	0.04	0.05	0.02	0.05	0.07
	(0.08)	(0.08)	(0.09)	(0.05)	(0.07)	(0.13)
Expanded Impact	0.18*	0.13	0.14	0.08	0.01	0.22*
	(0.10)	(0.09)	(0.09)	(0.06)	(0.06)	(0.12)
MCL (direct)	0.38***	0.36***	0.34***	0.28***	0.27**	0.27***
	(0.08)	(0.07)	(0.07)	(0.09)	(0.14)	(0.09)
Team Teacher on MCL-led team	0.14***	0.11***	0.13***	0.14***	0.10*	0.18***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.05)
“Placebo” team teacher	0.13**	0.11*	0.10*	0.09**	-0.02	0.18**
	(0.06)	(0.05)	(0.05)	(0.04)	(0.03)	(0.08)
		Reading				
BLT	-0.07	-0.14**	-0.13**	-0.07	-0.02	-0.07**
	(0.05)	(0.06)	(0.06)	(0.06)	(0.03)	(0.03)
Expanded Impact	0.06*	0.04	0.05*	0.06*	0	0.11***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
MCL (direct)	0.22***	0.18**	0.18**	0.18**	0.15**	0.09
	(0.06)	(0.08)	(0.08)	(0.07)	(0.07)	(0.06)
Team Teacher on MCL-led team	0.07***	0.05***	0.06***	0.06***	-0.01	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
“Placebo” team teacher	0.09**	0.09***	0.09***	0.09***	0.05**	0.10***
	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
Prior test scores	x	x	x	x	x	
Race						
Classroom prior tests		x	x	x	x	x
School prior tests			x			
School FE				x		
School-year FE					x	
Student FE						x

Notes: excludes CCS. Placebo teachers randomly generated from non-OC teachers within treated schools (see text). We omit model (2) from this table as it pools together multiple districts and we only observe race in one district.

Table 10: Sample restricted to 2014-15 and 2015-16

	(1)	(3)	(4)	(5)	(6)	(7)
		Math				
MCL (direct)	0.32***	0.37***	0.34***	0.28***	0.23	0.20***
	(0.10)	(0.10)	(0.09)	(0.11)	(0.15)	(0.07)
Team Teacher on MCL-led team	0.03	0.04	0.06	0.12**	0.13**	0.06
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.08)
BLT	0.06	0.07	0.07	0.07	0.07	0.06
	(0.10)	(0.09)	(0.11)	(0.09)	(0.10)	(0.13)
Expanded Impact	0.06	0.07	0.08	0.03	0.05	0.12
	(0.11)	(0.10)	(0.10)	(0.06)	(0.06)	(0.22)
		Reading				
MCL (direct)	0.16**	0.18**	0.18**	0.15**	0.15**	0.08
	(0.06)	(0.08)	(0.08)	(0.07)	(0.07)	(0.10)
Team Teacher on MCL-led team	-0.01	0.01	0	-0.02	-0.02	0
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
BLT	-0.21***	-0.24***	-0.22***	-0.05	-0.06	-0.13
	(0.08)	(0.08)	(0.08)	(0.09)	(0.05)	(0.11)
Expanded Impact	-0.02	-0.01	-0.01	-0.01	-0.02	0.04
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Prior test scores	x	x	x	x	x	
Race						
Classroom prior tests		x	x	x	x	x
School prior tests			x			
School FE				x		
School-year FE					x	
Student FE						x

Notes: excludes CCS. Placebo teachers randomly generated from non-OC teachers within treated schools. We omit model (2) from this table as it pools together multiple districts and we only observe race in one district.