

NBER WORKING PAPER SERIES

SCHOOL DISTRICT REFORM IN NEWARK:
WITHIN- AND BETWEEN-SCHOOL CHANGES IN ACHIEVEMENT GROWTH

Mark J. Chin
Thomas J. Kane
Whitney Kozakowski
Beth E. Schueler
Douglas O. Staiger

Working Paper 23922
<http://www.nber.org/papers/w23922>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2017

This project was funded by a grant from the Startup:Education Foundation (now the Chan-Zuckerberg Initiative.) The foundation also provided grant funding to support the Newark reform effort which we evaluate in the paper. We thank Bari Erlichson, Kendra Carter, Kate Fletcher, Jesse Margolis, Sarah McLean, James Riddlesperger and Peter Shulman for their help explaining the context in Newark and accessing the New Jersey data. We thank staff at the Newark Public Schools for reviewing earlier drafts of the report for errors. Kirsten Slungaard Mumma helped analyze the data and Jon Fullerton managed the project. Although unrelated to the current project, we disclose the following interests: Kane served as an expert witness for Gibson, Dunn, and Crutcher LLP to testify in Vergara v. California. Staiger is co-founder, has an equity interest in, and consults for ArborMetrix, a company that sells performance measurement systems and consulting services to healthcare providers. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Mark J. Chin, Thomas J. Kane, Whitney Kozakowski, Beth E. Schueler, and Douglas O. Staiger. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

School District Reform in Newark: Within- and Between-School Changes in Achievement Growth

Mark J. Chin, Thomas J. Kane, Whitney Kozakowski, Beth E. Schueler, and Douglas O. Staiger
NBER Working Paper No. 23922

October 2017

JEL No. H41,H43,I21,I24,I28

ABSTRACT

In 2011-12, Newark launched a set of educational reforms supported by a gift from Facebook CEO Mark Zuckerberg and Priscilla Chan. Using data from 2009 through 2016, we evaluate the change in Newark students' achievement growth relative to similar students and schools elsewhere in New Jersey. We measure achievement growth using a "value-added" model, controlling for prior achievement, demographics and peer characteristics. By the fifth year of reform, Newark saw statistically significant gains in English and no significant change in math achievement growth. Perhaps due to the disruptive nature of the reforms, growth declined initially before rebounding in recent years. Aided by the closure of low value-added schools, much of the improvement was due to shifting enrollment from lower-to higher-growth district and charter schools. Shifting enrollment accounted for 62 percent of the improvement in English. In math, such shifts offset what would have been a decline in achievement growth.

Mark J. Chin
Harvard Graduate School of Education
Center for Education Policy Research
50 Church St., 4th Floor
Cambridge, MA 02138
mark_chin@g.harvard.edu

Thomas J. Kane
Harvard Graduate School of Education
Center for Education Policy Research
50 Church St., 4th Floor
Cambridge, MA 02138
and NBER
kaneto@gse.harvard.edu

Whitney Kozakowski
Harvard Graduate School of Education
Cambridge, MA 02138
wkozakowski@g.harvard.edu

Beth E. Schueler
Harvard Kennedy School of Government
Cambridge, MA 02138
bes674@mail.harvard.edu

Douglas O. Staiger
Dartmouth College
Department of Economics
HB6106, 301 Rockefeller Hall
Hanover, NH 03755-3514
and NBER
douglas.staiger@dartmouth.edu

In the Fall of 2010, Facebook CEO Mark Zuckerberg, New Jersey Governor Chris Christie, and Newark Mayor Corey Booker announced a school improvement effort in Newark, to be aided by a \$100 million gift from Zuckerberg and Pricilla Chan. This \$100M gift – combined with \$100 million in matched funds from other donors – represented roughly 4% of Newark’s school budget over the five years of the grant. Yet, it provided the city and the district with the flexibility to implement an ambitious slate of reforms. With the appointment of a new school superintendent, Cami Anderson, the reforms were launched in the fall of the 2011-12 school year. On this fifth anniversary of the reforms, we provide a summary of the results so far.

“Reform” in Newark was not a single intervention, but a package of measures including a new teacher contract, new school leaders, a new curriculum aligned to the Common Core standards, school turnaround efforts (known as “renew schools”), charter school expansion, school closure, and, eventually, a universal choice plan (which allowed parents to submit a single application to attend a district or local charter school). Rather than attempt to parse the effect of each reform component (likely an impossibility), we divide the reforms into two broad categories: “within-school” reforms (such as personnel changes, Common Core implementation, and school turnaround efforts aimed at improving the results of existing schools) and “between-school” reforms (such as school closures, charter expansion, and universal choice, aimed at re-allocating students toward more effective schools). Using a value-added model to compare the achievement growth of students in Newark to similar students attending similar schools elsewhere in New Jersey, we decompose the change in average annual achievement growth into its “within-school” and “between-school” components.

A similar framework has been used to measure the sources of productivity growth in different countries (e.g. Bartelsman, Haltiwanger and Scarpetta (2013)) and industries (e.g.

Chandra, Finkelstein, Sacarny and Syverson (2016)). In those studies, productivity growth was often driven by gains in market share by more efficient firms-- as opposed to productivity improvements within existing firms. In K-12 education, the contributions from changing market share are typically blunted when school assignment is based on residence. However, by closing some schools, allowing students to move to charters and instituting a universal choice plan, the Newark reform strategy allowed parents to switch schools without switching residences. We assess the degree to which this opportunity improved the system's overall productivity.

Compared to the two academic years preceding the reforms (those ending in the spring of 2010 and 2011), we find that achievement growth in Newark improved significantly in English by 2016 and showed no significant change in math. However, overall progress was not linear. In the initial years of the reform, annual average achievement growth declined in math and English, in both the charter and district schools, only to increase in the most recent two years. After bouncing back in 2014-15, achievement growth in the average Newark school is now significantly higher than for similar students in the rest of the state in English and higher, though not significantly so, in math.

A distinguishing characteristic of the Newark reforms has been the shift in enrollment toward higher value-added schools. The relationship between achievement growth and shifting enrollment has been much stronger in Newark than elsewhere in New Jersey. The shifts occurred for several reasons, including the closure of schools with below-average achievement growth, expansion of charter schools with higher growth, and the introduction of universal choice (allowing parents to apply for district schools, magnet schools and charter schools in a single application). Indeed, 61 percent of Newark's improvement in English achievement growth between the baseline years and 2015-16 was due to shifting enrollment. In math, the sole source

of improvement was between-school movement, as math achievement growth would have declined in Newark relative to the rest of the state if not for enrollment shifts toward schools with faster achievement growth in math.

Our paper adds to a growing literature on school and district turnarounds (e.g., Abdulkadiroglu, Angrist, Hull & Pathak, 2016; Dee, 2012; Fryer, 2014; Gill, Zimmer, Christman & Blanc, 2007; Harris and Larsen, 2016; Heissel & Ladd, 2016; Papay & Hannon, 2015; Schueler, Goodman & Deming, 2017; Strunk, Marsh, Hashim, Bush & Weinstein, 2016; Young et al., 2009; Zimmer, Kho, Henry & Viano, 2015). As other research has begun to show, turnaround efforts can produce both positive and negative externalities for student achievement growth. For instance, school choice and closures may increase pressure on schools to improve. On the other hand, large movements of students and teachers can be disruptive both for students who move and their new classmates. By examining district-wide productivity change, our study incorporates both types of spillover effects.

Furthermore, existing studies tend to focus on the impact of reform on students in schools while they are undergoing turnaround. However, less attention has been paid to understanding how disruptions in one set of schools could affect other schools and eventually alter system-wide productivity for future cohorts. For example, several studies examine the effect of school closure on displaced students (Brummet, 2014; Carlson & Lavertu, 2015; de la Torre & Gwynne, 2009; Engberg, Gill, Zamarro & Zimmer, 2012; Kemple, 2015), but few consider the impact of closure on students who avoid attending a low-performing school because it had been closed prior to their entry into the district. In contrast, we apply a framework that examines the relative contributions of various turnaround mechanisms—within-school improvements and between-school reallocation of students—to overall district-wide productivity change over time.

Newark Reforms

The Newark Public School district (NPS) has long been the focus of reform efforts. In its 1985 *Abbott v. Burke* ruling, the New Jersey State Supreme Court declared the state's reliance on local property taxes for financing K-12 schools unconstitutional, and classified Newark as one of 28 districts needing additional state assistance (Hess, 2007).¹ These “*Abbott* districts” are characterized by poor academic performance and high concentrations of economically disadvantaged students. NPS was placed under state control in 1995 after the New Jersey Department of Education determined Newark was not meeting district certification standards. The district remains under partial state control today.²

In 2010, then Mayor Cory Booker and Governor Chris Christie began collaborating on an effort to reform the Newark school system. They found a willing investor in Mark Zuckerberg, chief executive of the social networking website, Facebook. Zuckerberg and Priscilla Chan (co-founder of the Chan-Zuckerberg Initiative) committed \$100 million to NPS, matched by an additional \$100 million, primarily from foundations and private donors (Fulbeck et al., 2016). Zuckerberg, Booker, and Christie announced the donation on the Oprah Winfrey show in September 2010. To put the magnitude of the donation in perspective, the \$200 million (which was to be spent over five years) was roughly four percent of the district's five-year \$5 billion budget. To carry out the reforms, Christie appointed Cami Anderson as superintendent in May

¹ There are now 31 such districts.

² In New Jersey, the state can take control of up to five areas of a district's activities: governance, fiscal management, personnel, operations, and instruction. In August 2016, the state returned local control of personnel, on top of operations and fiscal management, which had already been returned in 2007 and 2014, respectively (see details at <http://www.state.nj.us/education/news/2016/0803new.htm>). The state is moving toward fully restoring local control by 2017-18 (see details at http://www.nj.com/essex/index.ssf/2016/08/newark_likely_to_regain_local_control_of_district.html).

2011. Anderson had led New York City's alternative high school program under Chancellor Joel Klein. We provide a timeline of the Newark reforms in Figure 1.

The reforms can be grouped into two broad categories: those that aimed to improve existing schools ("within-school" reforms) and those that aimed to reallocate students toward more effective schools ("between-school" reforms).

Within-School Reforms. Over her first three years as superintendent, Anderson replaced more than half of the district's principals, including one-third in her first year (Russakoff, 2015). She reorganized the central office, cutting 120 positions and bringing in new leadership (Meyer, 2013). The new team implemented district-wide reforms such as monthly training sessions with principals, an online platform facilitating mutual consent hiring, an early warning system to prevent drop out, changes to data and accountability systems (e.g., school data dashboards), a revamped student registration system, and extended learning time in a subset of schools (Russakoff, 2015). Newark also piloted "blended learning models" incorporating educational software in classrooms (NPS, 2013).

In November 2012, the district ratified a new contract with the Newark Teachers Union (NTU). The four major contract provisions included: (1) a new teacher evaluation system (combining classroom observations, student growth, and other indicators such as lesson plans and teacher attendance), (2) differentiated teacher compensation, including incentives for high performers to stay in low-performing schools, (3) extended learning time in 28 schools, and (4) greater school-based decision making (including the ability for teachers to vote to overturn portions of the collective bargaining agreement). In return, NTU members received a \$31 million one-time payment to resolve outstanding wage demands from prior years and \$20 million in stipends during the first year of implementation (Fulbeck et al., 2016). A majority of teachers

perceived the new evaluation system to be fair and accurate, and reported that the extended learning time was helpful to students (American Institutes for Research, 2016). Although teachers had mixed support for differentiated compensation, Newark was able to retain teachers with higher ratings at higher rates under the new contract (Fulbeck, Citkowicz, Hester, Manzeske, Yisak & Eisner, 2016).

Newark also began emphasizing the Common Core state standards before many other New Jersey districts, adopting Common Core-aligned math and literacy instructional materials in grades K-8 starting in 2013-14. For example, nearly 90 percent of schools serving grades 3-8 adopted the Expeditionary Learning (EL) curriculum in English. These resources have been highly rated by EdReports for Common Core alignment. The change in curriculum may have contributed to temporary declines in performance on the NJASK test, as the materials were less aligned with NJASK and teachers and students likely needed time to learn how to use these new resources effectively.

The final major within-school reform was the attempted turnaround of underperforming schools. Eight K-8 schools were chosen for turnaround during the 2012-13 academic year, while an additional eight were chosen for 2014-15.³ The schools—labeled “renew schools”—were the focus of a multi-pronged effort. First, principals were required to reapply for their jobs; in the first round, half of the principals were replaced (Russakoff, 2015). Once hired, principals were able to rehire or replace teachers (McGlone, 2013; Calefati, 2012). About half of the teachers were replaced in the first round.⁴ Beyond staffing changes, students were given extended learning time and teachers received professional development. Renew schools also conducted

³ A small number of high schools were renewed in 2013-14 as well, but our analysis focuses on K-8 renew schools.

⁴ Teachers who were not rehired were provided other employment options within the district.

extra outreach to families and offered students better access to nurses, social workers, and community-based mentoring (Fulbeck et al., 2016; Kamenetz, 2013).

Between-School Reforms. While working to improve the district schools, Newark leaders also sought to shift students into higher quality schools. School closures were one component of that strategy. Between 2011-12 and 2014-15, eleven traditional district schools serving grades K-8 were closed based on low enrollment and poor performance.⁵ Some students from closed schools transferred to charter schools; others were re-assigned to nearby district schools, some of which were renew schools. Four of the eight schools in the first round of renewal received students from one of the four schools which closed that year. In addition to district school closures, three charter schools were closed between 2011-12 and 2014-15 for several reasons including poor academic performance and financial and legal concerns (Mooney, 2013). Students impacted by the 2013-14 school closures were offered transportation from several shuttle hub locations to their new schools. In addition to the K-8 closures, the district opened seven new schools serving students in grades six and above during this period, including two single sex schools.

Meanwhile, enrollment in Newark's charter schools grew rapidly. In spring 2011, 14 percent of all NPS students were enrolled in charters. By 2015-16, the percentage had doubled to 28 percent and was even higher among K-8 students (32 percent). The charter sector included 20 different operators by 2015-16. The largest were TEAM Academy, affiliated with the national Knowledge is Power Program (KIPP) network, and North Star Academy, affiliated with the national Uncommon Schools network. These two operators accounted for 46 percent of

⁵ We define closed schools as those whose school codes were no longer in use and whose facilities were no longer being used by the school that used to reside there.

Newark's charter enrollment in 2015-16 and 56 percent of the growth in charter enrollment between 2010-11 and 2015-16.

In the winter and spring of 2014, Newark instituted a universal choice system (now known as “Newark Enrolls”). This system allowed families to rank their preferred schools on a single application, choosing from among traditional district schools, magnet schools, and nearly all of Newark's charter schools. Students were then matched to schools.⁶ Previously, families were assigned to a district school based on their residence, and if they wanted to attend a charter school or magnet school, they were required to submit a separate application to each school. Beyond the shuttles mentioned above, NPS directly provided transportation only to students with an Individualized Education Program (IEP) and provided public bus tickets to students living two or more miles from their school (NPS, 2016). TEAM Academy (KIPP) was the only charter network that provided busing (Russakoff, 2015).

Despite the national attention on Newark, the impact of the reforms on student achievement growth has never been evaluated.⁷ We focus on annual achievement growth—rather than levels of achievement—given the changes in the composition of students attending district and charter schools in Newark. Below we describe our data, empirical strategies, and results.

⁶ In addition to accounting for families' rankings, the algorithm prioritized keeping siblings together, allowing students to attend a school in their residential neighborhood, and increasing the representation of students with an IEP or FRPL eligibility in schools where these students are underrepresented in the applicant pool relative to the citywide average. For magnet schools, the matching also accounts for school rankings of applicants.

⁷ Baker and Weber (2015) used publicly available school-level data from grades 6-8 to compare trends in average scale scores (not growth) from 2009 to 2014 in Newark to the rest of New Jersey, and found no difference.

Data

Our data cover all New Jersey public school students in grades four through eight for seven academic years (2009-10 through 2015-16). The data include student demographic and program participation indicators including age, gender, race or ethnicity, limited English proficiency status (LEP), eligibility for free- or reduced-price lunch (FRPL), and disability classification, as well as current- and prior-year test performance on state mathematics and English language arts (ELA) standardized tests.

For the academic years ending in 2009 through 2014, our student outcomes are NJASK state test scores.⁸ In spring of 2015 and 2016, New Jersey administered the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment, designed to reflect the Common Core State Standards. For both tests, we standardized scores by grade and year. We limited our analysis to students with non-missing outcomes and covariates. We also dropped those in schools and grades with fewer than five students in a year.

In Table 1, we report mean characteristics for four subgroups: students attending Newark district schools, Newark charter schools, other *Abbott* district schools and the remainder of New Jersey schools. Like the students in the other *Abbott* districts, students in Newark district schools were more likely to be African American or Latino and more likely to be eligible for subsidized lunches than students in the rest of New Jersey. They were also lower achieving, with mean scores in the prior year between one-half and two-thirds of a standard deviation below the state average in math and English. Compared to Newark district schools, students in Newark charters

⁸ The distribution of NJASK scale scores across grades and years revealed slight ceiling effects for math. As such, we calculated rank-based standardized scores (also known as van der Waerden scores (Conover, 1999)) for math, standardized by grade and year. Because we observed no apparent ceiling effect in the NJASK ELA scale scores or in the PARCC math and ELA scores, we standardized by grade and year using the usual method (subtracting the mean and dividing by the standard deviation).

were more likely to be African American, but less likely to be Latino, special education, or LEP students. Newark charter school students had somewhat higher achievement in the prior year than the Newark district students, scoring 0.16 standard deviations (SD) below the statewide average, rather than two-thirds of a SD below.

Methods

To assess the impact of the Newark reforms, we measure the rate of annual achievement growth for all Newark students (including those attending charter schools) relative to other students in New Jersey with similar baseline achievement and demographics, attending schools with similar mean baseline achievement and demographics.

Specifically, we estimate the following value-added model:

$$A_{ijt} = \beta_0 + X_{ijt}\beta_1 + \delta_{jt} + \varepsilon_{ijt} \quad (1)$$

Where A_{ijt} represents the score on the mathematics or English test for student i in district j at time t . X_{ijt} represents our vector of covariates, including student demographic characteristics, prior test performance, the average characteristics of peers in the school and grade, grade-by-year fixed effects and (especially given the change to the new test in 2015) grade-by-year interactions with the student and peer covariates. δ_{jt} are district-by-year fixed effects. Although New Jersey treats charter schools as districts of their own, we categorize charter schools in Newark and the *Abbott* districts with their neighboring district when estimating equation (1).

We include district-by-year fixed effects to avoid conflating the influences of student and school-level characteristics (for which we seek to adjust) and the effect of district-level policy changes (which we seek to measure). In other words, we estimate the effect of student and

school characteristics based only on within-district-year variation in those covariates, which may be correlated with district-level policy changes. (For instance, other low-income districts may have been pursuing the same policies as Newark). Accordingly, our estimates adjust Newark's performance relative to other schools based solely on the measured effect of those school characteristics within districts in each year.

We then estimate the difference in adjusted achievement growth between Newark and other New Jersey districts with the following second step equation:

$$A'_{ijt} = \gamma_0 + \gamma_{2010-2011} Newark_j + \sum_{t=2012}^{2016} \gamma_t Newark_j * Year_t + \tau_t + u_{ijt} \quad (2)$$

Where A'_{ijt} is student i's achievement, adjusted by subtracting off the estimated effect of prior achievement, demographics and mean characteristics ($X_{ijt}\widehat{\beta}_1$) from equation (1), leaving the district-by-year effects and the residuals ($A'_{ijt} = \widehat{\delta}_{jt} + \widehat{\varepsilon}_{ijt}$). This 2-step method is analogous to the approach used in Chetty, Friedman and Rockoff (2014), in which the authors used only within-teacher variation to estimate the effects of covariates (Equation 1), and then analyzed the between-teacher variation in a second stage (Equation 2).

The coefficient in Equation 2 on the Newark indicator, $\gamma_{2010-2011}$, captures the adjusted achievement gain in Newark relative to similar students and schools elsewhere in New Jersey in the pre-reform years (2010 and 2011). The remaining coefficients, $\gamma_{2012} \dots \gamma_{2016}$, represent the differential change in adjusted achievement growth in Newark from the pre-reform years, relative to the change in other New Jersey districts. To capture changes in achievement growth outside Newark, we include year fixed effects (τ_t). For both equations we calculate standard

errors by clustering at the school level. Additionally, we estimate similar equations including separate time trends for Newark charters, renew, and other district schools.

Within-school Changes in Achievement Growth. We also explore the within-school changes in achievement growth in Newark by estimating a model identical to the one depicted in Equation (2) except that we (a) include school fixed effects, μ_s and (b) omit the main effect of enrollment in Newark (i.e., the baseline value-added of Newark in Equation (1), $NEWARK_j$) due to collinearity of the Newark indicator with the school fixed effects. Thus, our coefficients of interest— γ_t —represent the average change in productivity within Newark schools over time. We also explore within-school productivity trends of different school types (charters, non-renew traditional district, and renew traditional district schools).

Enrollment Growth and Achievement Growth. Given Newark’s efforts to allow children to move between schools, we estimate the relationship between baseline measures of school achievement growth and subsequent school enrollment changes. Specifically, we follow Chandra et al. (2016) and estimate the following model at the school-level, by year and subject, separately for Newark and for comparison districts:

$$\Delta Enrollment_{sj,t+1} = \lambda_0 + \lambda_1 \hat{\mu}_{sjt} + \alpha_j + \mu_{sjt} \quad (3)$$

Where $\Delta Enrollment_{sj,t+1}$ is the difference between school s ’s enrollment in time $t+1$ and the school’s enrollment in time t . $\hat{\mu}_{sjt}$ is an estimate of school s ’s value-added in year t and α_j is a

district fixed effect.⁹ The coefficient of interest, λ_1 , captures the within-district relationship between school value-added in time t and the subsequent change in enrollment between time t and $t+1$. A positive value of λ_1 would mean higher value-added schools gained market share. We first present results using enrollment in grades from which school value-added is constructed (four through eight). However, given Newark has many K-8 schools, the earliest grades may be most sensitive to increased availability of schooling options. To capture this, we present results using enrollment in grades one through three.¹⁰ We estimate the model using three samples—Newark, other *Abbott* districts, and the rest of New Jersey—to compare the relationship between school productivity and enrollment growth observed in Newark to other districts in the state.

Finally, we decompose the change in district-wide average school value-added to assess the extent to which productivity changes in Newark were due to within-school change versus between-school movement. We follow Chandra et al. (2016) in their adaptation of Foster, Haltiwanger, and Krizan (2001), Foster, Haltiwanger, and Syverson (2008), and Baily et al. (1992). Our decomposition takes the following form:

⁹ We estimate school value-added in year t first by estimating equation (1) across all students in New Jersey:

$$A_{ijt} = \beta_0 + X_{ijt}\beta_1 + \delta_{jt} + \varepsilon_{ijt} \quad (1)$$

Then, we regress A'_{ijt} on school random effects, estimating each model one year at a time. Since we use value-added as a predictor in equation (3), we estimate empirical Bayes shrunk random effects for schools to account for differences in the reliability of estimates across schools due to enrollment differences (Kane & Staiger, 2008).

¹⁰ We exclude kindergarteners due to concerns that kindergarten enrollments may be inaccurate in some years in the data. Additionally, attendance at schools exclusively serving kindergarten and pre-K in Newark appears to have increased over this period. These schools lack value-added data because they do not serve tested grades. As a result, we cannot evaluate the relationship between value-added and enrollment shifts for kindergarteners. To avoid confounding the effects of this shift with other shifts related to enrollment in grades one through eight over this time, we drop kindergarten students from the analysis.

$$\begin{aligned}
\Delta \bar{q}_t = & \underbrace{\sum_{s \in C_t} \theta_{s,t-1} \Delta q_{s,t}}_{\text{within}} + \underbrace{\sum_{s \in C_t} (q_{s,t-1} - \bar{q}_{t-1}) \Delta \theta_{s,t}}_{\text{between}} + \underbrace{\sum_{s \in C_t} \Delta q_{s,t} \Delta \theta_{s,t}}_{\text{cross}} + \underbrace{\sum_{s \in M_t} \theta_{s,t} (q_{s,t} - \bar{q}_{t-1})}_{\text{entry}} \\
& + \underbrace{\sum_{s \in X_t} \theta_{s,t-1} (\bar{q}_{t-1} - q_{s,t-1})}_{\text{exit}} \quad (4)
\end{aligned}$$

In Equation (4), \bar{q}_t is the district's average value-added in a given subject in school year t and Δ represents change.¹¹ Therefore, the outcome is the difference in average value-added between two periods ($t - 1$ and t). On the right-hand side, $q_{s,t}$ is value-added for school s in year t , and $\theta_{s,t}$ is the share of students in the district enrolled in school s in year t . C_t is the set of schools that were open in both $t - 1$ and t , M_t is the set of schools that opened in year t , and X_t are schools that closed between $t - 1$ and t .

The above model decomposes average change in a district's productivity into five terms. The "within" term reflects changes in average value-added in the district due to value-added improvements among schools holding their share of the district's enrollment constant. The remaining terms reflect the various ways students re-allocate across schools. The "between" term represents the extent to which productivity improvements were due to students moving from lower value-added schools to schools that were already higher value-added in year $t - 1$.

The "cross" term represents enrollment growth at schools with value-added that improved between the two periods. If schools that were growing in enrollment were also seeing

¹¹ To recover district and school average value-added within a year for the decomposition analysis, we first estimate equations (1) and (2) for both subjects across all New Jersey students. Then, we take A'_{ijt} from equation (2) and subtract the year effects ($\hat{\tau}_t$) estimated in equation (2) to get student-level value-added (to remove remaining year effects in the district-year fixed effects estimated in equation (1)). We average this to the school- and district-level for each year.

declines in the average achievement growth, this term will be negative. We categorize the “cross” term as a “between school” movement, since it captures the degree to which enrollment shifted toward improving schools. (Some authors categorize this term neither as “within” nor “between”, but as it’s own category).

The “entry” term captures improvements in value-added due to new schools opening with higher value-added than the previous district average. The “exit” term captures productivity gains due to lower than average value-added schools closing. Each of these terms could be negative if changes in value-added, enrollment shares, or the district’s portfolio of schools contributed to reductions in average district-wide value-added.

We decompose the change in value-added into a “long difference” where $\Delta\bar{q}_t$ represents the change in average value-added between the two baseline years (2010 and 2011) and our final year, 2016, in a given district. We also decompose the change in value-added between each consecutive year in our time series to explore whether the within and between contributions varied over time. We calculate these terms for Newark and the other *Abbott* districts, treating the *Abbotts* as a single district for the decomposition. As a result, the “between” term for the *Abbott* districts captures student movement across both schools and districts.

Results

Figure 2 portrays the trend in district-by-year effects ($\widehat{\delta}_{j,t}$) for Newark and the average of the other *Abbott* districts relative to the rest of the state. The first year, ending in spring 2010, occurred before the \$100 million donation was announced in September 2010. The second year occurred after the gift was announced but before the appointment of the new superintendent and

before major reforms began in Newark. Thus, we consider 2010 and 2011 as the two baseline years, and 2012 through 2016 as the five years post-reform.

Relative to the rest of the state and the *Abbott* districts, Figure 2 illustrates that value-added in Newark declined (or, at the very least, showed no obvious improvement) in the first years of the reform (2012-2014) in both math and ELA. However, value-added increased sharply for Newark in 2015 in both subjects. In that single year, Newark's achievement growth increased by 0.09 SD in math (from 0.03 to 0.12) and 0.11 SD in English (from -0.05 to 0.06). Compared to the largest districts in NJ over this time, Newark's change was larger than 1 SD of the change that the average large district experienced.¹²

Newark was not the only low-income district to see achievement growth increase with the administration of the new Common Core-aligned PARCC assessments in 2015. During the same year, growth in the *Abbott* districts also increased (although less than Newark) by 0.01 and 0.06 SD in math and English respectively. In data published by the state, both Newark and the average *Abbott* district also saw large increases in their 2015 Student Growth Percentiles in English and math.¹³ In 2016, achievement growth in Newark remained well above the state and the *Abbott* districts in English, but declined in math (remaining above the *Abbott* districts and the state, but no longer significantly so).

Table 2 reports estimates from Equation 2 showing the achievement growth trends,

¹² We defined the largest districts as those which enroll at least 3,000 tested students in grades 4-8. This yields a total of 31 districts which meet this threshold.

¹³ A student growth percentile (SGP) in year t is the percentage of students at the same grade level statewide with similar performance on the $t-1$ test who the student outperformed in year t . Therefore, a median school SGP greater than 50 indicates that the median student outperformed more than half of her peers statewide with the same baseline score the previous spring. The weighted average of median SGP's in ELA in the Newark district schools jumped from 40 in 2014 to 49 in 2015 to 54 in 2016. However, it could be that a portion of the jump in SGP scores in Newark—especially in ELA—was an artifact of a transition to the PARCC. Other districts serving low-income students also witnessed a jump in SGP's in 2015. While we adjust for the mean characteristics of a student's schoolmates (including the percentage receiving FRPL) using within-district variation and allow those adjustments to vary by year when estimating value-added, SGP models only adjust for a student's baseline achievement.

overall and separately for traditional and charter schools in Newark.¹⁴ The first row of the table reports the difference between Newark and the rest of the state in value-added in the two years prior to the reforms. As reported in the first column, the combined achievement growth of Newark district and charter schools in math was 0.068 SD higher than similar students elsewhere in New Jersey pre-reform. In column four, the 0.017 SD difference in English growth was not statistically different from zero. However, as reported in the third and sixth columns, the Newark advantage in achievement growth before the reforms was primarily due to the Newark charter schools, where annual achievement growth was 0.319 SD higher in math and 0.215 SD higher in ELA. To put the magnitude of these estimates in context, one standard deviation in value-added across the largest NJ districts was .08 in math and .05 in ELA in the baseline years. In other words, Newark charters were more than one SD above the highest district value-added in math and almost two SDs above the highest district in ELA. In contrast, the Newark district schools had comparable achievement growth to similar schools elsewhere in New Jersey in both subjects in the two baseline years.

The Newark charters' advantage in achievement growth on the eve of the Newark reforms was quite large—two to three times as large as the difference in achievement growth associated with having a novice versus an experienced teacher. As other research suggests, Newark is home to one of the most effective charter sectors in the nation in terms of student growth on standardized exams (CREDO, 2015). The only other city known to have a similarly effective charter sector is Boston, Massachusetts (Angrist, Dynarski, Kane, Pathak & Walters, 2012; Abdulkadiroglu, Angrist, Dynarski, Kane & Pathak, 2011; Angrist, Dynarski, Kane,

¹⁴ We estimate the effect of district and charter schools in one pooled model for math (columns 2-3) and one pooled model for ELA (columns 5-6). The model includes a main effect for Newark district schools and Newark charter schools, Newark district school-by-year indicators, Newark charter school-by-year indicators, and year fixed effects.

Pathak, Walters, 2010). As we discuss below, the close proximity of an unusually effective charter sector played a major role in the impact of the Newark reforms.

The remaining rows of Table 2 report the change in Newark's value-added during the reform years relative to the rest of New Jersey. In the first three years of reform, Newark's achievement growth declined in math, in both district and charter schools. By 2013, achievement growth in math was 0.092 SD lower in district schools and 0.171 SD lower in charter schools than in the pre-reform years.

By 2015, across all Newark schools, achievement growth recovered to be a bit above pre-reform levels, although the difference was only marginally significant for English. Achievement growth by 2015 was slightly higher than in the baseline years for district schools and slightly lower for charter schools.

The improvements in English achievement growth persisted and were significantly above baseline by 2016, although there was some backsliding in math, with achievement growth significantly below the baseline years in both the charter and district sector. Yet, despite the decline within both sectors, when the charter and district schools are combined in column one, the Newark achievement growth advantage in math in 2016 was not statistically different from the baseline (2010/2011). The reason for the apparent contradiction was the sizeable shift in enrollment toward the charter sector, which we discuss below.

In terms of ELA achievement growth, the Newark district schools were ahead in 2016 of where they were in 2010/2011, relative to comparable schools in New Jersey. Instead of having achievement growth 0.014 SD below average, Newark district schools had achievement growth 0.044 (-0.014+0.058) SD above average in English. Meanwhile, Newark charter schools largely maintained their advantage in English, generating 0.168 SD more growth than schools working

with similar students in New Jersey (0.215-0.047).

Understanding the 2015 Increase. The timing of the sharp rise in Newark's achievement growth in 2015 corresponded with two other events, which could have affected Newark's measured performance: the transition from the NJASK to the PARCC assessment and an increase in the proportion of students missing scores on the state tests. We investigated both.

The PARCC differed from the previous NJASK assessment in both subjects. In math, this meant more emphasis on mathematical reasoning over procedural knowledge, while in English, this entailed more emphasis on written responses employing critical thinking skills to analyze reading passages (NJDOE, 2016). Moreover, the PARCC test in New Jersey was almost exclusively administered by computer (in contrast to the paper-based NJASK format). One possibility is that PARCC assessed different content than NJASK, and Newark students may have performed better on this content even in the years prior to 2015, had they been tested on it. Although we have no way to test it, this seems plausible given that Newark schools began emphasizing the Common Core State Standards earlier than other districts in New Jersey.

At the student level, the relationship between a student's achievement score one year to the next did not change with the transition to the new test. A student's performance on the 2015 PARCC test was only slightly less correlated with his or her performance on the prior year test as it had been in 2014 (0.82 versus 0.84 in math and 0.80 versus 0.83 in English). In other words, a student's prior performance on NJASK was equally predictive of his or her PARCC achievement as it was of her subsequent performance on NJASK in prior years. Kane et al. (2016) reported similar findings regarding the year-to-year correlations in test scores in four states administering Common Core aligned assessments in 2015 (Massachusetts, New Mexico, Maryland and

Delaware).

However, a student's performance on any test reflects a combination of factors: those that are stable (such as their family background, test-taking behaviors, prior learning) as well as recent additions to their knowledge (their improved understanding of the new standards). Student-level correlations could be due to family background and prior learning—which we would expect to be stable—even if the school-based additions to their knowledge changed with the change in assessment.

As a result, we estimated district-level differences in achievement growth within New Jersey in each of the years (which reflect additions to knowledge, after differencing out the effect of students' background and prior learning). During the NJASK years (2010-2014), the year-to-year correlation between district-level value-added averaged 0.64 for math and 0.66 for ELA.¹⁵ However, even after excluding Newark, the correlation between district value-added in 2015 (with PARCC) and district value-added in 2014 (with NJASK) fell to 0.47 in math and 0.29 in ELA. Thus, the introduction of the PARCC test reshuffled district rankings in value-added more than in pre-PARCC years. Assuming both tests have similar levels of measurement error, this implies that the PARCC and NJASK were assessing different sets of skills and the districts that excelled in preparing students for PARCC were not necessarily the same as the districts that excelled at preparing students for NJASK. Thus, what appears to be a single-year gain in performance may have been present before 2015, but was simply undetected by earlier NJASK tests.

We also investigated whether the 2015 increase in Newark value-added was related to the increase in the proportion of students missing assessment results in the state data. Figure 3

¹⁵ Correlations were similar among the largest quarter of New Jersey districts.

reports the proportion of New Jersey students with missing math scores.¹⁶ Parent and teacher groups opposed to standardized testing (e.g., United Opt Out) were especially active in New York and New Jersey in spring 2015. Across New Jersey, the proportion of students missing scores spiked in 2015. Although we cannot distinguish between those actively opting out and those missing scores for other reasons, it is likely that the sharp increase was due to students and parents “opting-out” of the PARCC test at greater rates (Ujifusa, 2015). The 2015 rise in the proportion missing scores was particularly large in Newark.

However, the increase in the proportion of students missing test scores does not seem to have caused the increase in Newark’s 2015 student achievement. Figure 4 plots the change in school-level value-added against the increase in the proportion of students with missing scores between 2014 and 2015, separately for Newark and the remaining New Jersey districts. If the opt-out movement explained Newark’s 2015 value-added increase, we would expect to see the schools with greater increases in the proportion missing scores to experience greater value-added increases. However, the opposite seems to have been true. The increase in the proportion of students missing scores was negatively related to the increase in a school’s value-added, indicating that the rise in the proportion missing test scores does not seem to have driven the increase in value-added in Newark.

Within-School Changes. Table 3 investigates the change in value-added within existing Newark schools, both overall and separately for three categories of schools: charter schools, the eight district schools renewed in 2013, and the remaining district schools. The first and fifth columns represent the overall estimates with the inclusion of school fixed effects to isolate within-school changes. In both math and ELA, there were continuing declines in achievement

¹⁶ The ELA plot was very similar.

growth within schools through 2014. Charter schools had the largest within-school declines in achievement growth in both subjects in 2012. There were especially large declines in renew schools in 2013 (the first year of renewal), when achievement growth declined by 0.17 SD in math and 0.13 SD in English relative to the baseline years. These declines were followed by a rebound in 2015 in both subjects. For ELA, this rebound persisted into 2016 but within-school math value-added declined below pre-reform levels again in 2016. On a within-school basis, the average charter school achievement gain in math was still 0.19 SD below the pre-reform years in 2016. In the district schools that were neither renewed nor closed, achievement growth was 0.081 SD below pre-reform levels. However, by 2016, annual achievement growth in the renew schools had recovered to the point that they were not statistically different from pre-reform levels in either subject.

One possible explanation for the within-school declines in achievement growth was the disruption that occurred as new curricula were introduced, schools closed, their students were absorbed into other schools, and renew schools hired new teachers. In 2013, there was a spike in new students entering schools in Newark, particularly in the renew schools that absorbed the lion's share of students from closed schools. This surge of new students could have contributed to the decline in value-added in both Newark's district and charter schools. However, in supplemental analyses not reported here, controlling for the proportion of students in non-entry grades new to a school does not explain the within-school decline in value-added.

In addition to student turnover, Newark's schools were also experiencing significant teacher turnover during the early reform years. There was a surge in new teachers entering renew schools in 2013, many of whom were in their first year of teaching in the state. Specifically, nearly half of the renew school teachers were new to their school in 2013. Roughly a quarter of

teachers were new in other district schools that year. This may help explain the initial decline in value-added in these schools. Moreover, we might have expected achievement to subsequently rise simply as new teachers gained experience.

Between-School Reallocation. In Table 4, we investigate the reallocation of students toward higher value-added schools. In Newark, there is a significant positive relationship between a school's value-added and its subsequent enrollment growth. The relationship between value-added and enrollment growth holds in both the tested grades (4-8) and younger grades (1-3). Looking over the entire 2011-2016 period, the estimates suggest that in Newark a school with 0.10 higher value-added in 2011 gained roughly 40 additional students.

Enrollment changes were much less related to differences in value-added in the *Abbott* districts and the rest of New Jersey. For these districts, there was no statistically significant relationship between enrollment growth and value-added differences in most years, and the magnitude of the coefficient in each year and over the whole period (2011 to 2016) was much smaller than in Newark.

Figure 5 plots enrollment growth in grades four to eight from 2011 to 2016 against school math value-added in 2011 for all Newark schools. There was substantial variation in value-added across schools in Newark, ranging from below -0.30 to above 0.40, and a clear upward slope indicating that higher value-added schools grew faster. The graph emphasizes that much of this relationship is due to the growth of charter schools and closing of less effective district schools. Also, we highlight schools that were oversubscribed in the first year of universal choice (2015). Since most of these schools had higher value-added, there would have likely been even more reallocation toward high value-added schools had it not been for these capacity constraints preventing even larger enrollment shifts.

Decomposition. Table 5 decomposes the change in value-added from the pre-reform baseline years (2010/2011) to 2016 into within- and between-school components separately for Newark and the other *Abbott* districts. Between the baseline period (2010/2011) and the most recent available year (2016), average value-added in Newark grew by 0.07 SD in English. In English, that is a sizeable gain, comparable to the impact of being assigned to an experienced versus novice teacher (Gordon, Kane and Staiger, 2006). However, 62 percent of the difference in English was due to the reallocation of students from lower to higher value-added schools.

In math, the average achievement gain was largely unchanged in Newark, at -0.036 SD. However, improvements due to the reallocation of students from low to high-value-added schools (0.043) essentially offset a within-school decline of a larger magnitude (-0.08). In other words, Newark's math value-added would have declined if not for the shift in enrollment toward higher achievement growth schools. The *Abbott* districts saw negligible gains from between school movements: 0.007 in math and 0.008 in English (both statistically significant but small).

To illustrate the importance of the reallocation of students in Newark, Figure 6 plots annual estimates of the cumulative change in value-added attributable to the within-school and between-school components. While improvements due to between-school reallocations have grown steadily (to about 0.043 SD per year in both subjects), there was a u-shaped pattern of within-school changes with substantial initial declines followed by a large increase in 2015. In math, this was followed by a large decline in 2016, but in ELA, within-school improvements continued into the most recent year.

Most of the between-school improvement in value-added was driven by enrollment shifts between existing schools. Our decomposition suggests that school closures were important part of the between-school improvement in Newark, simply because many closed schools had below-

average value-added prior to closure. Figure 7 ranks all Newark schools by their math value-added, and highlights the K-8 schools that closed between 2012 and 2014. Closed schools tended to rank lower in value-added, although some schools that were not closed ranked even lower. Figure 8 plots the difference in closure year school value-added between a closed school and the school the student moved to for all students in closed schools between 2012 and 2014. Students in closed schools with the lowest value-added moved to schools with substantially higher value-added, while those in closed schools with relatively high value-added saw little change in their school's value-added. Additionally, students who moved to charter schools saw consistently larger gains in value-added, regardless of their closed school's value-added.

Theoretically, the students from the closed schools would not necessarily have experienced the same higher achievement growth as their new classmates. We explore this question in Appendix A, and find evidence that closed school students do, on average, experience higher achievement growth after moving schools.

Conclusion

Although much has been written about the political controversies surrounding the Newark reforms, this is the first systematic accounting of their impact on student achievement growth. By 2016, average student growth in Newark improved significantly relative to the rest of the state in English and was not significantly changed in math. However, our findings also highlight the difficulty of turning around existing schools, as reflected in substantial declines in achievement growth within both charter and district schools in the early years of reform. Although there are examples of district turnaround yielding immediate academic improvements (such as New Orleans, Louisiana (Harris & Larsen, 2016) and Lawrence, Massachusetts

(Schueler, Goodman & Deming, 2016)), it is not uncommon for districts undergoing reform to see initial declines in student achievement growth (the Tennessee Achievement School District (Zimmer et al., 2015)). For Newark, the within-school rebound in achievement growth between 2014 and 2016 suggests that the tide may have turned (particularly in English).

As noted in Figure 2, Newark was not the only low-income district in New Jersey that witnessed an increase in achievement growth in 2015 and 2016. To the extent that the gains reflect policy shifts that Newark and other districts adopted, the improvement is encouraging. However, because the rise coincided with the introduction of a new assessment, there is some risk that the change is an artifact of measurement—a change in test scaling or testing procedures that disproportionately benefited students attending high poverty schools. We have tested a number of possibilities: the effect of accommodation policies on students with disabilities; floor or ceiling effects on the PARCC; a boost in scores for English language learners created by better text-to-speech options on the computer-based exams; changes in the treatment of incomplete items in the scoring of NJASK and PARCC. None of these hypotheses can account for the rise in achievement by many low-income districts in New Jersey. Nonetheless, even if a skeptic were to attribute the 0.07 gain in ELA achievement in the *Abbott* districts between 2014 and 2016 to some unspecified measurement artifact which also benefited Newark, the change in achievement growth in Newark was still twice as large as the other *Abbott* districts.

Still, the most distinctive element of the Newark reforms started in their first year (before the change in assessments) and continued through 2016: the steady shift in student enrollment toward higher value-added schools (both charter and district schools). That shift was helped along by the closure of some of the district's lowest value-added schools—which essentially forced parents to choose from among a set of schools with, on average, higher value-added

scores (there may be room for further improvement using this strategy given some of the lowest value-added schools remain open). There was also steady expansion of the charter sector over this time, with an emphasis on expanding enrollment at the high-value-added charters. As a result, even in English where there was sharp increase in achievement growth within existing schools, over sixty percent of the improvement was due to between-school reallocation of students. And, in math, the between school movement was necessary to convert a within-school loss in achievement growth to ensure that overall achievement growth remained on par with the district's relative performance in 2010/2011.

The experience in Newark has shown that re-allocation of market share can be an important contributor to productivity growth in K-12 education, as it has been in many other industries. Nevertheless, other cities considering Newark-style reforms should remember two facts that make Newark unusual: First, reallocation was not always "market-based," as school closures required some students to move involuntarily. Closing schools is politically difficult. Other districts, not under state control, may be less successful in making such hard choices. Second, Newark started the reform process with access to an unusually effective set of charter schools nearby, which is not present in most other cities. Therefore, it would be an oversimplification to suggest that most of Newark's progress was due solely to greater parental choice. In Newark, the positive effects of parental choice were enhanced by a series of difficult, but generally well-targeted school closures and ready access to an unusually effective charter sector.

References:

- Abdulkadiroglu, Atila, Joshua D. Angrist, Peter D. Hull and Parag A. Pathak. 2016. Charters without lotteries: Testing takeovers in New Orleans and Boston. *American Economic Review*, 106(7): 1878-1920.
- Abdulkadiroglu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane and Parag A. Pathak. 2011. Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics*, 126(2): 699-748.
- Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak and Christopher R. Walters. 2010. Inputs and impacts in charter schools: KIPP Lynn. *American Economic Review: Papers and Proceedings*, 100(2): 239-43.
- Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak and Christopher R. Walters. 2012. Who benefits from KIPP? *Journal of Policy Analysis and Management*, 31(4): 837-860.
- Baily, Martin N., Charles Hulten, and David Campbell. 1992. Productivity dynamics in manufacturing plants. *Brookings Papers on Economic Activity, Microeconomics*, 187–267.
- Baker, Bruce D. and Mark Weber. 2015. Research note: On average, are children in Newark doing better? *New Jersey Education Policy Forum*.
- Bartelsman, Eric, John Haltiwanger and Stefano Scarpetta. 2013. Cross-country differences in productivity: The role of allocation and selection. *American Economic Review*, 103(1): 305–334.
- Brummet, Quentin. 2014. The effect of school closings on student achievement. *Journal of Public Economics*, 119(C): 108-124.
- Calefati, Jessica. 2012. Principals selected for 8 struggling schools in Newark. *The Star Ledger*. Retrieved from http://www.nj.com/news/index.ssf/2012/05/principals_selected_for_8_stru.html
- Carlson, Deven and Stephane Lavertu. 2015. School closures and student achievement: An analysis of Ohio's urban district and charter schools. Thomas B. Fordham Institute.
- Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. Healthcare exceptionalism? Performance and allocation in the U.S. health care sector. *American Economic Review*, 106(8): 2110-2144.
- Chetty, Raj, John N. Friedman and Jonah E. Rockoff. 2014. Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9): 2593-2632.
- Chingos, Matthew M., Grover J. Whitehurst and Michael R. Gallaher. 2015. School districts and student achievement. *Education Finance and Policy*, 10(3): 378-398.

Center for Research on Education Outcomes. 2015. Urban charter school study report on 41 regions.
<https://urbancharters.stanford.edu/download/Urban%20Charter%20School%20Study%20Report%20on%2041%20Regions.pdf>

Conover, William J. 1999. *Practical nonparametric statistics (volume 3)*. New York: Wiley.

Dee, Thomas. 2012. School turnarounds: Evidence from the 2009 stimulus. National Bureau of Economic Research Working Paper No. 17990.

Engberg, John, Brian Gill, Gema Zamarro and Ron Zimmer. 2012. Closing schools in a shrinking district: Do student outcomes depend on which schools are closed? *Journal of Urban Economics*, 71(2): 189-203.

Foster, Lucia, John C. Haltiwanger and C.J. Krizan. 2001. Aggregate productivity growth: Lessons from microeconomic Evidence. In Charles R. Hulten, Edwin R. Dean and Michael J. Harper (Eds.), *New Developments in Productivity Analysis*, pp. 303–372. National Bureau of Economic Research.

Foster, Lucia, John Haltiwanger and Chad Syverson. 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1): 394–425.

Fulbeck, Eleanor, Martyna Citkowicz, Candace Hester, David Manzeske, Melissa Yisak and Ryan Eisner. 2016. *Newark Public Schools and Newark Teachers Union teacher contract evaluation: Year 1 report*. American Institutes for Research.

Gordon, Robert, Thomas J. Kane and Douglas O. Staiger. 2006. Identifying effective teachers based on performance on the job. Brookings Institution, Hamilton Project.

Gill, Brian, Ron Zimmer, Jolley Christman and Suzanne Blanc. 2007. State takeover, school restructuring, private management, and student achievement in Philadelphia. Santa Monica, CA: RAND Corporation and Research for Action.

Harris, Douglas N. and Matthew F. Larsen. 2016. The effects of the New Orleans post-Katrina school reforms on student academic outcomes. Tulane University, Education Research Alliance for New Orleans.

Heissel, Jennifer A. & Helen F. Ladd. 2016. School turnaround in North Carolina: A regression discontinuity analysis. *Economics of Education Review*.

Hess, Frederick M. 2007. Adequacy judgments and school reform. In Martin R. West & Paul P. Peterson (Eds.), *School Money Trials*, pp. 159-194. Washington, DC: Brookings Institution Press.

Kamenetz, Anya. (2013). Cami Anderson And Zuckerberg, the unlikely duo behind Newark schools' revitalization. *Fast Company*. Retrieved from <http://www.fastcompany.com/3007765/cami-anderson-and-zuckerberg-unlikely-duo-behind-newark-schools-revitalization>

Kane, Thomas J. and Douglas O. Staiger. 2008. Estimating teacher impacts on student achievement: An experimental evaluation. *National Bureau of Economic Research Working Paper No. 14607*.

Kane, Thomas J., Antoniya M. Owens, William H. Marinell, Daniel R.C. Thal and Douglas O. Staiger. 2016. *Teaching Higher: Educators' perspectives on common core implementation*. Center for Education Policy Research, Harvard University.

Kemple, James. 2015. High school closures in New York City: Impacts on students' academic outcomes, attendance, and mobility. The Research Alliance for New York City Schools.

Kim, Juli, Bryan C. Hassel, Elaine Hargrave, Lyria Boast, Christen Holly and Shonaka Ellison. 2015. Early Lessons from Newark's Experience with Charter Schools. Public Impact.

McGlone, Peggy. 2013. Newark school restructuring includes plans to put charters in district buildings. *NJ.com*. Retrieved from http://www.nj.com/education/2013/12/newark_school_plan_restructure.html

Meyer, Peter. 2013. Newark's superintendent rolls up her sleeves and gets to work. *Education Next*, 13(1): 2-5.

Mooney, John. 2013. Fine print: Newark charter school revocation. *NJ Spotlight*. Retrieved from <http://www.njspotlight.com/stories/13/04/15/fine-print-newark-charter-school-revocation/>

New Jersey Department of Education. (2016). *Historical Context: Overview of New Jersey's Statewide Testing Program*. Retrieved from <http://www.nj.gov/education/assessment/history.shtml>

Newark Public Schools (2013). Race to the Top District Application for Funding.

Newark Public Schools (2016). Newark Enrolls Frequently Asked Questions: School Matches. Retrieved from <http://newarkenrolls.org/frequently-asked-questions-school-matches/>

Papay, John and Molly Hannon. 2016. The Effect of School Turnaround Strategies in Massachusetts. Working paper, Brown University.

Russakoff, Dale. 2015. *The prize: Who's in charge of America's schools?* New York, NY: Houghton Mifflin Harcourt.

Schueler, Beth E., Joshua S. Goodman and David J. Deming. 2017. Can states take over and turn around school districts? Evidence from Lawrence, Massachusetts. *Educational Evaluation and Policy Analysis*, 39(2): 311-332.

Strunk, Katherine O., Julie A. Marsh, Ayesha K. Hashim, Susan Bush-Mecenas and Tracey Weinstein. 2016. The impact of turnaround reform on student outcomes: Evidence and insights from the Los Angeles Unified School District. *Education Finance and Policy*, 11(3): 251-282.

Ujifusa, Andrew. (2015). 'Opt-out' push gains traction amid common-core testing. *Education Week*, 34(29):16-17.

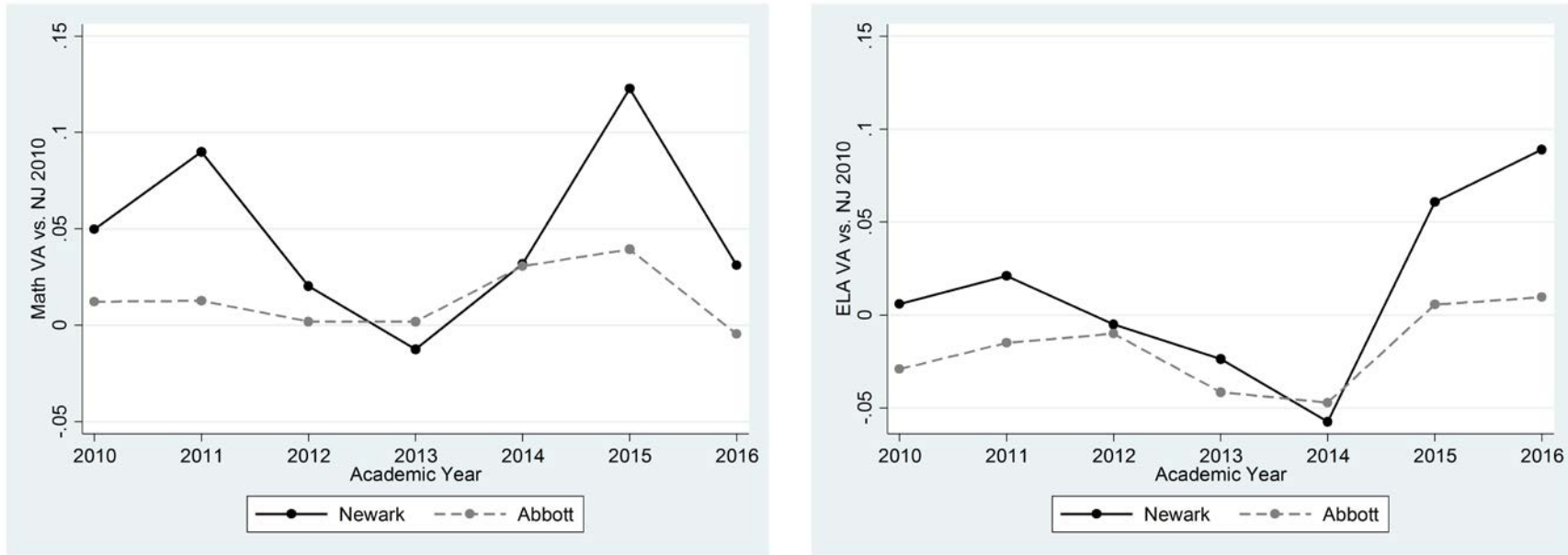
Young, Viki M., Daniel C. Humphrey, Haiwen Wang, Kristin R. Bosetti, Lauren Cassidy, Marjorie E. Wechsler, Elizabeth Rivera, Samantha Murray and Diane Whitmore Schanzenbach. 2009. Renaissance schools fund-supported schools: Early outcomes, challenges, and opportunities. SRI International, Menlo Park, CA.

Zimmer, Ron, Gary T. Henry and Adam Kho. (2017). The effects of school turnaround in Tennessee's achievement school district and innovation zones. *Educational Evaluation and Policy Analysis*.

Figure 1. Timeline of Recent Events and School Reforms in Newark

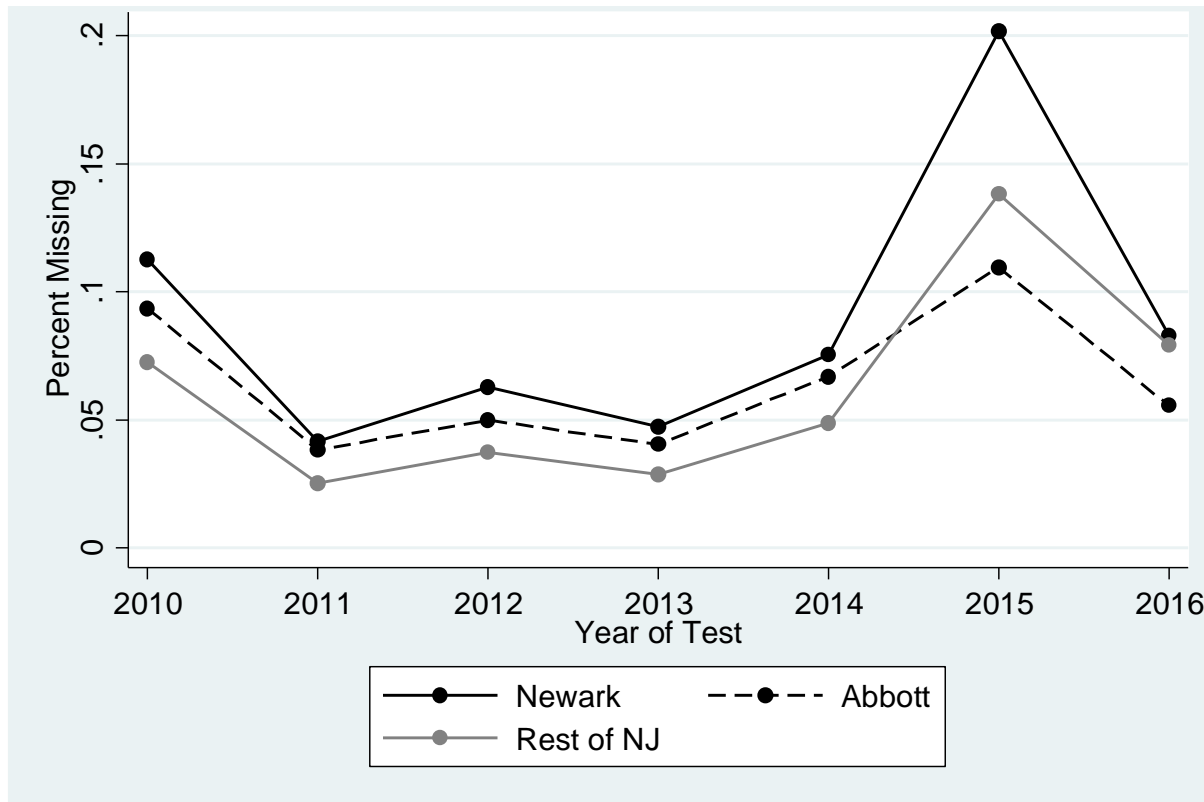
1985	First <i>Abbott v. Burke</i> State Supreme Court finance ruling leads to identification of 31 disadvantaged "Abbott districts"
1995	State takes control of Newark schools
Sept. 2010	\$100 million StartUp:Education donation announced on Oprah Winfrey Show
May 2011	Governor Chris Christie appoints Cami Anderson as Newark Superintendent
Spring 2012	Decision to consolidate twelve schools into eight "renew" schools and expand pre-K announced
2011-12	Nearly 25% of principals replaced leading up to the 2011-12 school year Central office reorganized New district-wide systems for registration, hiring, and data and accountability implemented Blended learning models and interim assessments with data use training piloted Extended learning time implemented in subset of schools
Spring 2012	Four K-8 schools closed at end of school year
Sept. 2012	Eight K-8 schools "renewed" (four absorb students from closed schools)
Nov. 2012	Agreement reached on new pay-for-performance teacher contract
Sept. 2013	A subset of high schools "renewed"
Spring 2013	Three K-8 schools closed at end of school year
Fall 2013	"One Newark" universal enrollment plan announced
Spring 2014	Families submit school choice preferences for 2014-15 placements Seven K-8 schools closed at end of school year
Sept. 2014	Eight additional K-8 schools "renewed"
Spring 2015	Families submit "Newark Enrolls" school choice preferences for 2015-16 placements
Spring 2015	First administration of Common Core-aligned PARCC exam in New Jersey
Summer 2015	Cami Anderson resigns and Chris Cerf assumes Superintendent role
Summer 2016	State announces plan to fully restore local control of Newark schools by 2017-18

Figure 2. Trends in Math and ELA Value-added in Newark and the Abbott Districts Relative to the Rest of New Jersey



Note: Value-added is calculated here by first estimating equation (1), taking the residuals from that regression and adding to them the estimated district-year fixed effects (i.e. A'_{ijt} in the text description). Then, in an equation similar to equation (2), we regress A'_{ijt} on the interactions between a Newark indicator and year dummies, interactions between an Abbott indicator and year dummies, and year fixed effects. We plot coefficients from this second regression here.

Figure 3. Percent of Students Missing Valid Math Test Scores Over Time



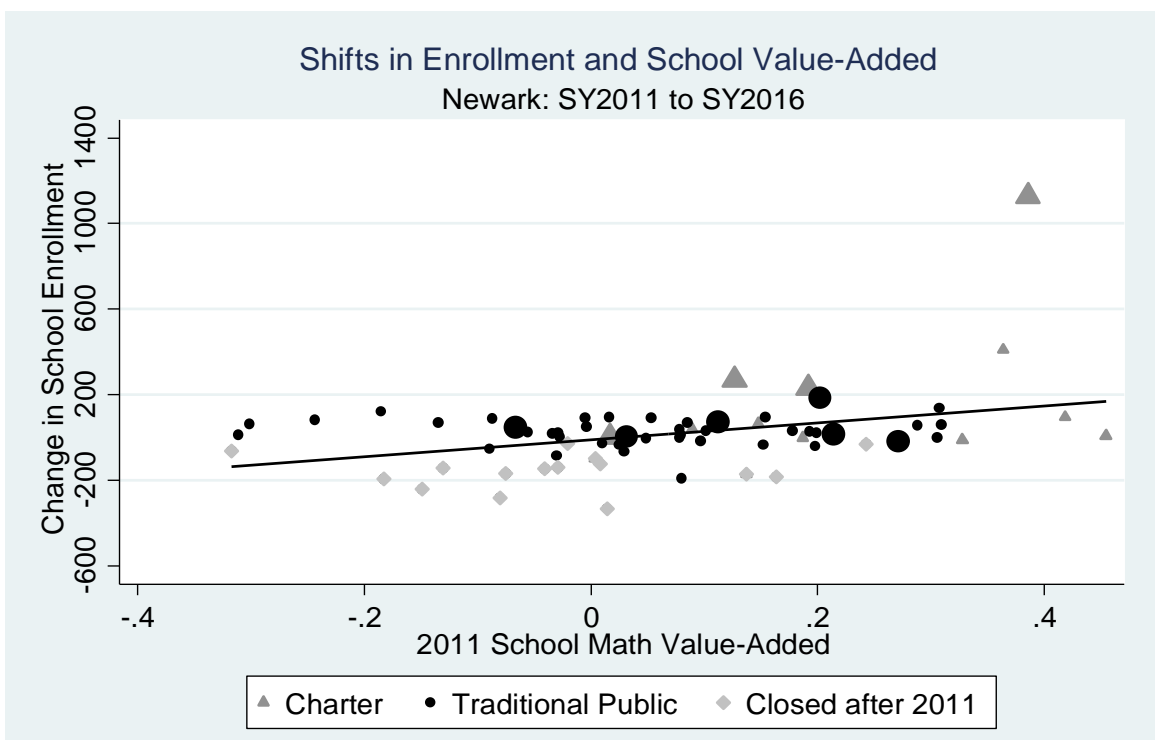
Note: This figure plots the percentage of students in grades 4-8 missing math test scores in the academic years 2009-10 through 2015-16 in Newark, Abbott districts, and the rest of the state. The plot for ELA was similar.

Figure 4. Change in Percent Missing Valid Test Scores and Change in Value-added by School Between 2014 and 2015



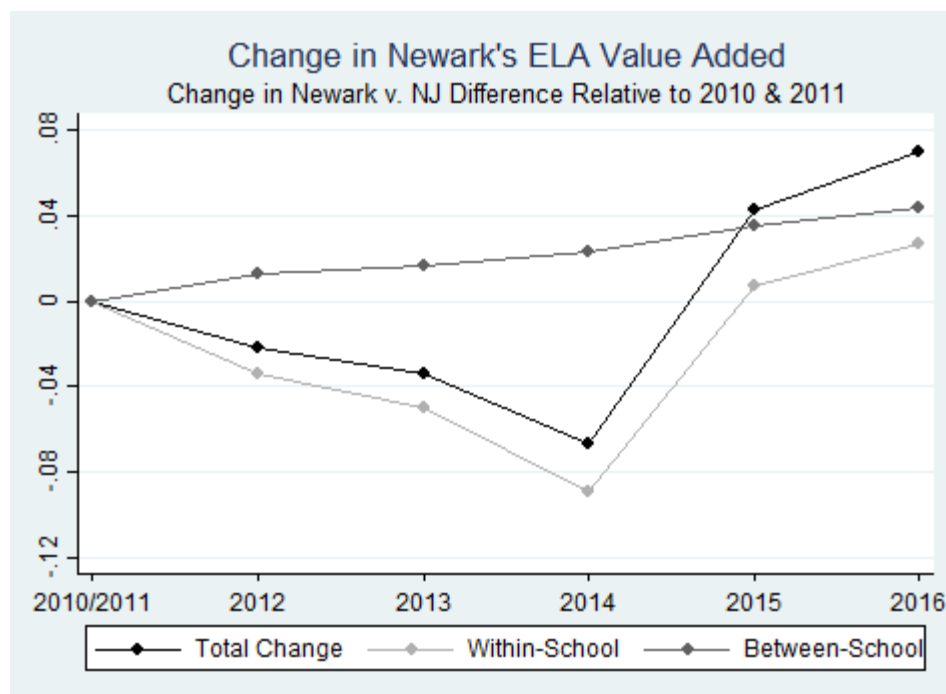
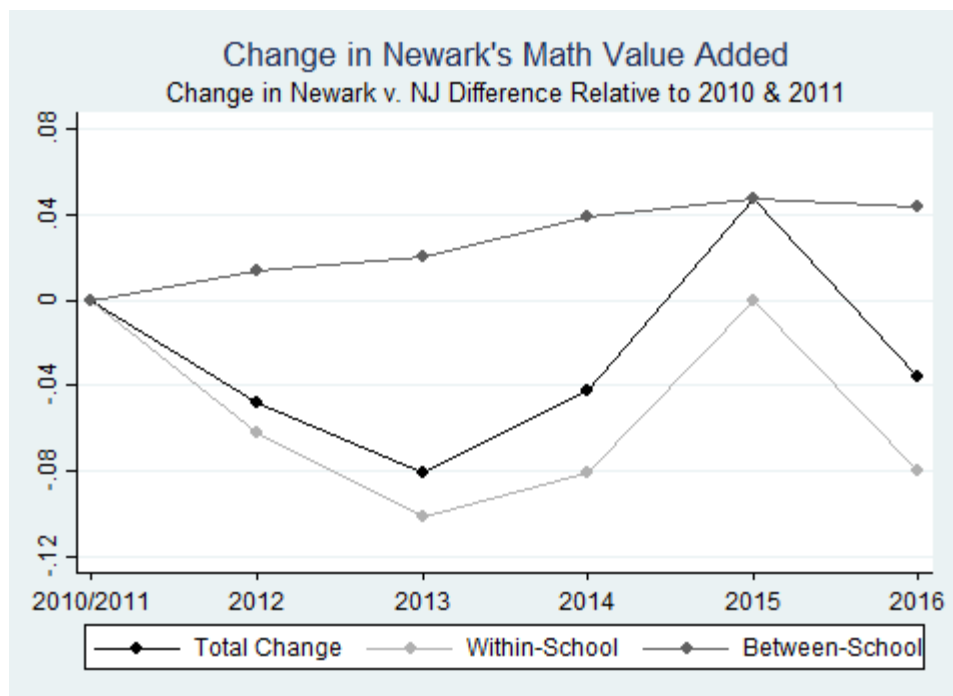
Note: Changes in value-added and the percentage of students missing test scores are calculated by subtracting the 2013-14 academic year from the 2014-15 academic year.

Figure 5. School Math 2011 Value-added and Enrollment Changes from 2011 to 2016 in Newark



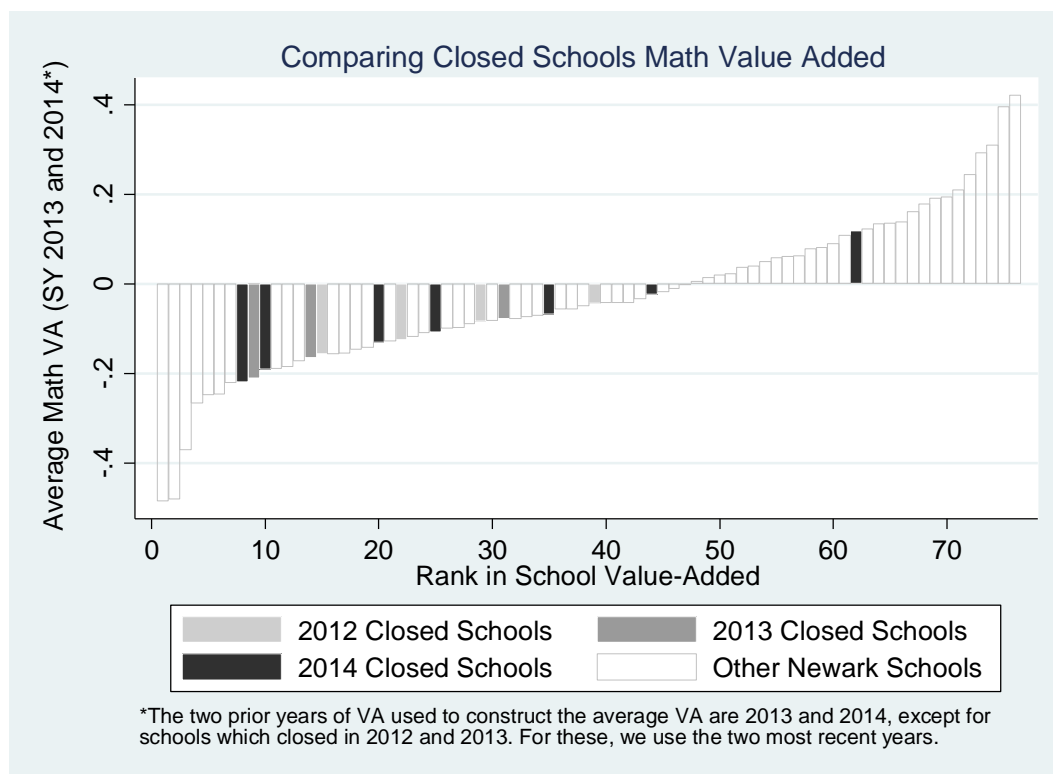
Note: Enrollment changes are calculated by subtracting 2010-11 enrollment from 2015-16 enrollment. Enrollment is based on tested grades (i.e. 4-8). Larger markers indicate schools that were oversubscribed in 2015, according to the 2015-16 Newark Enrolls Guidebook for Families.

Figure 6. Decomposing Newark's Change in Value-added Relative to New Jersey in 2010/2011



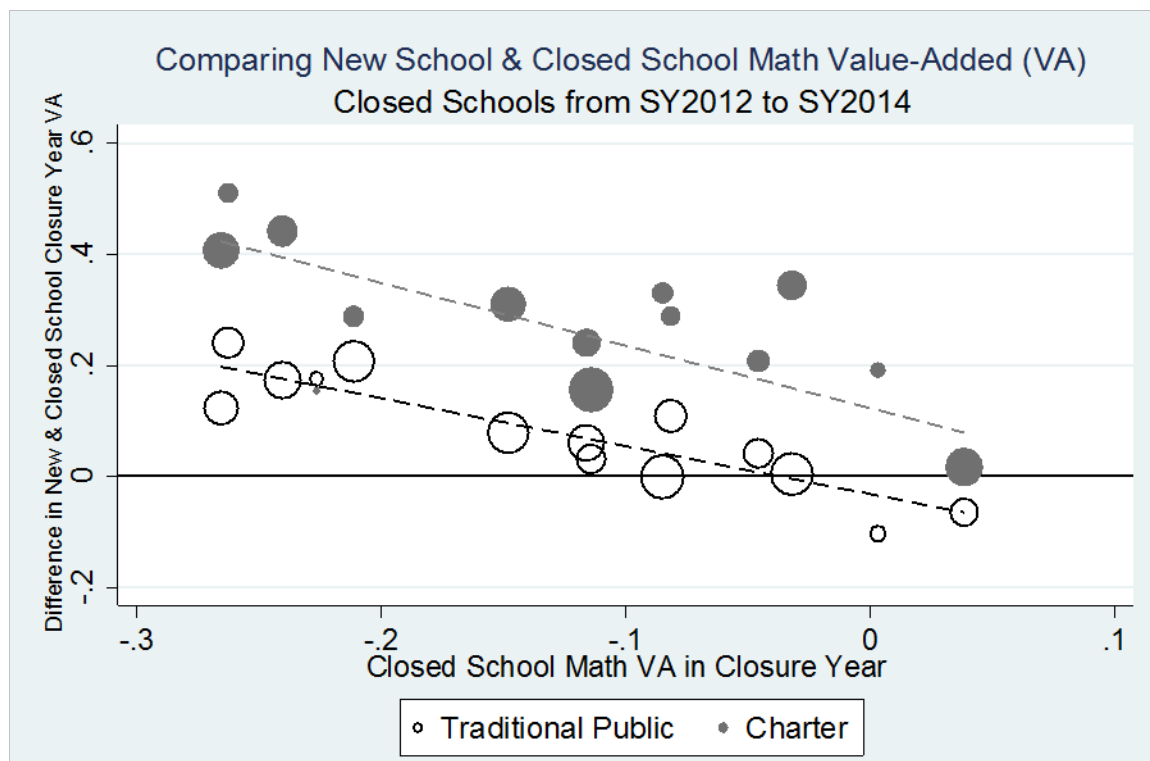
Note: This figure plots the decomposition of Newark's change in value-added, relative to New Jersey in the baseline period, into its within and between-school components using a version of Equation (4) where instead of estimating a "long difference" between 2009-10/2010-11 and 2015-16, we estimate the year-to-year change.

Figure 7. Math Value-added Rank of Closed Schools Relative to Other Newark Schools



Note: This figure presents a school's average math value-added for all schools in the 2012-13 and 2013-14 academic years in rank order. The exception (*) is schools that closed in 2011-12 and 2012-13 for which we take the average value-added from the last two years they were open.

Figure 8. Change in School Math Value-added for Students Leaving Closed Newark Schools



Note: For each student leaving a closed school, we calculate the difference in school-level math value-added between her closed school and new school using value-added calculated in the closure year for both schools (e.g. 2011-12 value-added for a school that closed at the end of spring 2012). We plot the average difference for students who moved to a district school (transparent) and those who moved to a charter school (dark gray) on the y-axis against the closed school's closure year math value-added. The bubbles are weighted by the number of students they represent. Dashed lines (in dark gray and black) are fitted regressions for students who move to a charter school and district school, respectively.

Table 1. Student sample characteristics, 2009-10 to 2015-16

	Newark District	Newark Charter	Other Abbotts	Rest of New Jersey
N of students	11,716	3,698	80,030	365,680
N of schools	52	15	328	1,421
Male	0.512	0.460	0.508	0.512
White	0.079	0.007	0.107	0.626
African American	0.493	0.797	0.325	0.103
Asian	0.009	0.009	0.034	0.112
Latino	0.417	0.164	0.498	0.134
American Indian	0.001	0.000	0.001	0.001
Multiracial/Other	0.001	0.023	0.034	0.024
FRPL eligible	0.883	0.829	0.798	0.235
LEP	0.053	0.004	0.072	0.012
Has disability	0.165	0.106	0.147	0.156
Prior math achievement	-0.668	-0.158	-0.550	0.167
Prior ELA achievement	-0.668	-0.158	-0.550	0.167

Note: The reported sample sizes are averaged across all the years for students with valid ELA scores. The numbers with valid math scores were similar. Charter networks count as a single school in these data.

Table 2. Trends in Value-Added in Newark Schools Relative to the Rest of New Jersey

VARIABLES	Math			ELA		
	All	District	Charter	All	District	Charter
Newark v. NJ Difference in 2010&2011	0.068*** (0.023)	0.028 (0.018)	0.319*** (0.033)	0.017 (0.018)	-0.014 (0.014)	0.215*** (0.035)
Change in Newark v. NJ Difference Relative to 2010&2011:						
2012	-0.048** (0.020)	-0.050** (0.021)	-0.136*** (0.046)	-0.021 (0.018)	-0.022 (0.019)	-0.090** (0.045)
2013	-0.081*** (0.022)	-0.092*** (0.024)	-0.171*** (0.038)	-0.034** (0.015)	-0.046*** (0.016)	-0.097*** (0.027)
2014	-0.042 (0.026)	-0.061*** (0.022)	-0.126** (0.063)	-0.066*** (0.018)	-0.082*** (0.017)	-0.133*** (0.035)
2015	0.047 (0.035)	0.035 (0.027)	-0.095 (0.084)	0.043* (0.024)	0.031* (0.018)	-0.066 (0.057)
2016	-0.036 (0.025)	-0.057** (0.024)	-0.172*** (0.040)	0.070*** (0.026)	0.058*** (0.021)	-0.047 (0.048)

Note: Robust standard errors in parentheses (clustered at the school level); *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. These estimates were generated by a two-step process in which we first estimated a value-added model controlling for student demographics (indicators for race/ethnicity, gender, special education status, limited English proficiency status, and free and reduced price lunch status), prior achievement (including cubic polynomials of math and ELA achievement scores interacted with grade), peer covariates (school-grade cohort means for all demographic and prior achievement characteristics), district-by-year fixed effects, grade-by-year fixed effects, and interactions between our grade-by-year fixed effects and all student and peer covariates. We then estimate a second equation in which the outcome is the residual plus district-by-year fixed effects estimated in the first equation, and the predictors are a main effect for Newark, Newark-by-year indicators, and year fixed effects. For columns 2-3 and 5-6, in the second equation we include a main effect for Newark district schools and Newark charter schools, Newark district school-by-year indicators, Newark charter school-by-year indicators, and year fixed effects.

Table 3. Within-School Changes in Value-Added in Newark (Including School Fixed Effects)

	Math				ELA			
	All	Charter	Renew	Other	All	Charter	Renew	Other
Change in Newark v. NJ Difference Relative to 2010&2011:								
2012	-0.064*** (0.019)	-0.132*** (0.047)	-0.016 (0.023)	-0.057** (0.024)	-0.034* (0.017)	-0.082* (0.047)	-0.034*** (0.011)	-0.024 (0.021)
2013	-0.107*** (0.021)	-0.152*** (0.032)	-0.171*** (0.037)	-0.090*** (0.027)	-0.052*** (0.014)	-0.072*** (0.024)	-0.132*** (0.031)	-0.039** (0.017)
2014	-0.079*** (0.022)	-0.132** (0.062)	-0.054 (0.054)	-0.076*** (0.024)	-0.091*** (0.015)	-0.126*** (0.034)	-0.108** (0.051)	-0.085*** (0.018)
2015	-0.002 (0.030)	-0.101 (0.082)	0.075*** (0.029)	0.015 (0.032)	0.007 (0.020)	-0.056 (0.055)	0.024 (0.034)	0.023 (0.020)
2016	-0.098*** (0.021)	-0.190*** (0.037)	-0.025 (0.033)	-0.081*** (0.027)	0.025 (0.019)	-0.049 (0.045)	0.035 (0.030)	0.049** (0.024)

Note: Robust standard errors in parentheses (clustered at the school level); *** p<0.01, ** p<0.05, * p<0.1. These estimates were generated by a two-step process in which we first estimated a value-added model controlling for student demographics (indicators for race/ethnicity, gender and free and reduced price lunch status), prior achievement (including cubic polynomials of math and ELA achievement scores interacted with grade), peer covariates, district-by-year fixed effects, grade-by-year fixed effects, and interactions between our grade-by-year fixed effects and student and peer covariates. We then estimate a second equation in which the outcome is the residual from the first equation plus district-by-year fixed effects and the predictors are a main effect for Newark, Newark-by-year indicators, year fixed effects, and school fixed effects. Columns 2-4 and 6-8 each respectively are estimated based on a single regression in which we split up the main effect of enrollment in Newark into three categories: Newark charter schools, renew schools, and other schools.

Table 4. Regressing Growth in School Enrollment on School Math Value-Added

	Grades 4-8			Grades 1-3		
	Newark	Abbott	Rest of NJ	Newark	Abbott	Rest of NJ
2011 to 2012	55.45 (34.46)	29.03** (13.80)	0.668 (27.51)	47.66 (29.85)	18.84 (15.77)	-0.107 (11.83)
2012 to 2013	130.0** (51.43)	32.52 (32.15)	11.86 (22.40)	124.7** (51.65)	23.33 (32.07)	11.40 (13.88)
2013 to 2014	123.7*** (46.09)	6.794 (21.74)	27.63 (26.74)	61.33 (55.49)	17.04 (27.94)	14.48 (14.42)
2014 to 2015	122.8* (68.15)	22.79 (53.24)	4.136 (11.10)	174.5 (115.5)	-22.66 (25.94)	-15.68* (9.362)
2015 to 2016	165.1** (81.72)	46.50** (21.80)	14.31 (10.66)	29.68 (57.22)	25.57** (11.19)	-6.906 (7.685)
2011 to 2016	394.6** (187.3)	93.25 (58.46)	-27.70 (36.42)	405.1* (236.4)	-11.22 (40.55)	3.112 (18.65)

Note: The reported estimates are coefficients from a school-level regression of change in enrollment on Empirical Bayes estimates of baseline value-added. Abbott and Rest of NJ regressions include district fixed effects. Standard errors in the Newark regressions are robust. Standard errors in the Abbott and Rest of NJ regressions are clustered at the district level. *p<.05, **p<.01, ***p<.001.

Table 5. Decomposing the Change in Value-Added from 2010/2011 to 2016

	Newark		Abbott	
	Math	ELA	Math	ELA
Total change in VA	-0.036 (.007)	0.070 (.007)	-0.014 (.003)	0.026 (.003)
Within School	-0.080 (.006)	0.026 (.006)	-0.021 (.003)	0.018 (.003)
Total Between School	0.043 (.003)	0.043 (.003)	0.007 (.001)	0.008 (.001)
<i>Between existing schools</i>	0.049 (.005)	0.038 (.005)	0.002 (.001)	-0.001 (.001)
<i>Cross</i>	-0.017 (.005)	-0.004 (.006)	0.001 (.001)	0.002 (.001)
<i>School entry</i>	-0.001 (.001)	0.002 (.001)	-0.001 (.000)	0.004 (.000)
<i>School closing</i>	0.013 (.002)	0.008 (.002)	0.005 (.001)	0.003 (.001)

Note: The change in the average value-added of the schools attended by Newark students is broken down into two components: improvements (or declines) in value-added within existing schools and changes due to the movement of students between schools with differing value-added. The change in value-added due to movement between schools is further broken down into four components: enrollment shifts between schools that existed throughout 2011 through 2016, the opening of new schools (school entry), the closing of schools and a final component due to a cross-product of changes in enrollment by changes in value-added. See text for further details.

Appendix A: School Closure Effects on Student Achievement

Students in Newark schools that closed subsequently attended schools with higher average value-added. In other words, students moved to schools where the *average* student was making faster progress. In this section, we explicitly test whether students from closed schools themselves subsequently outpaced similar students elsewhere in New Jersey.

If student learning was not disrupted when a student moved between schools, value-added is a valid estimate of a school's average treatment effect, and that effect applies to all subgroups of students, then we would expect the achievement of students from closed schools to rise by the difference between their new and old schools' value-added. However, those assumptions may not be true. For instance, moving schools as a result of a closure may generate short-term achievement disruptions. Second, schools absorbing closed students may not maintain the same productivity (e.g. they hire novice teachers to accommodate expansion, the influx of new closed school students disrupt other students' learning, etc.). Third, if treatment effects are heterogeneous, students from closed schools may not experience the same gains as the average student at their new school.

We estimate the effect of closure on achievement growth for students in three cohorts of schools that closed at the end of the 2011-12, 2012-13, and 2013-14 school years (referred to by their spring years from here on). The sample is limited to students who have a valid test score and lagged test score in the year of closure. Together, this sample includes 14 schools and 2,216 students in grades four to eight in the closure year (four schools in the 2012 cohort, three in 2013, and seven in 2014).

We use a difference-in-difference model to estimate the closure effect. The first difference is growth before versus after closure for students in closure cohorts. The second

difference is the growth of students in the rest of the state over this same period of time. We use the following model:

$$Y_{it} = \alpha + \sum_{k=-3}^3 \beta_{ik} s_{ik} + \delta_{gt} + \lambda_c + \varepsilon_{it}$$

Y_{it} are the same student-level value-added estimates calculated using the two-step procedure involving equations 1 and 2 as the main outcome of interest (See footnote 11). δ_{gt} are a set of grade-by-year fixed effects and λ_c are dummy variables for the three closed school cohorts (the excluded group is the rest of the state). s_{ik} is a set of dummy variables for whether a student is in a closed cohort and the student is k years from closure (e.g. -1 corresponds to one year before closure). β_{ik} are the coefficients of interest for each year relative to the closure year.

Appendix Table A1 reports that students in closed schools had lower average value-added than students across the rest of the state. In the main specification pooling cohorts in columns 1 and 5, students in closed schools had value-added between 0.10 and 0.15 SD below the state average in math and 0.10 to 0.17 SD below the state average in ELA, depending on the cohort.

Second, there appear to be parallel trends in value-added on average between the closed school students and the rest of the state in the pre-closure period. None of the pre-closure coefficients using the pooled models are statistically different than zero in either subject. However, there is evidence of non-parallel trends for the 2013 cohort in the pre-period. In Figure A1, this cohort of closures appears to have experienced a dip in value-added in the closure year, particularly in math. In contrast, the 2012 and 2014 cohorts had more stable pre-closure value-added trends on par with the state.

Post-closure, we find that students from closed schools had a statistically significant increase in value-added in the year following closure in math and ELA. Specifically, these students grew 0.14 SD in math and 0.11 SD in ELA more in the first years after closure than they did pre-closure, controlling for statewide changes in growth over this period. This large gain in value-added relative to the closure year is consistent through the post-closure period in math and ELA, suggesting that closed school students did reap the benefits of moving to higher value-added schools.

Figure A1 plots the Table A1 coefficients to show trends in value-added estimates for each year on average (top panels) and for each cohort (bottom panels).¹⁷ The 2013 and 2014 closure cohorts both experienced a jump in value-added one year after closure, increasing by 0.18 SD and 0.22 SD in math and 0.09 and 0.23 SD in ELA, respectively. In ELA, these gains were sustained in the post-closure period, while in math, they dropped somewhat becoming statistically insignificant but remaining positive by the last year estimated for each cohort. However, the 2012 cohort fared worse than the other two cohorts in the years following closure. Students' value-added was no better off within two years after closure. However, by three years after closure, the 2012 cohort had gained 0.15 SD in math relative to change in the rest of the state.

Overall, students in Newark's closed schools appeared to benefit academically from closure in both math and ELA. This growth was sustained for three years after closure on average, suggesting that schools receiving closed students were able to transfer their higher average value-added to new students. However, the effects vary across cohorts. While students in the 2013 and 2014 cohorts experienced immediate value-added gains as a result of closure,

¹⁷ Some cohorts have more post- or pre-closure years than others because our data are limited to the 2010 to 2016 window.

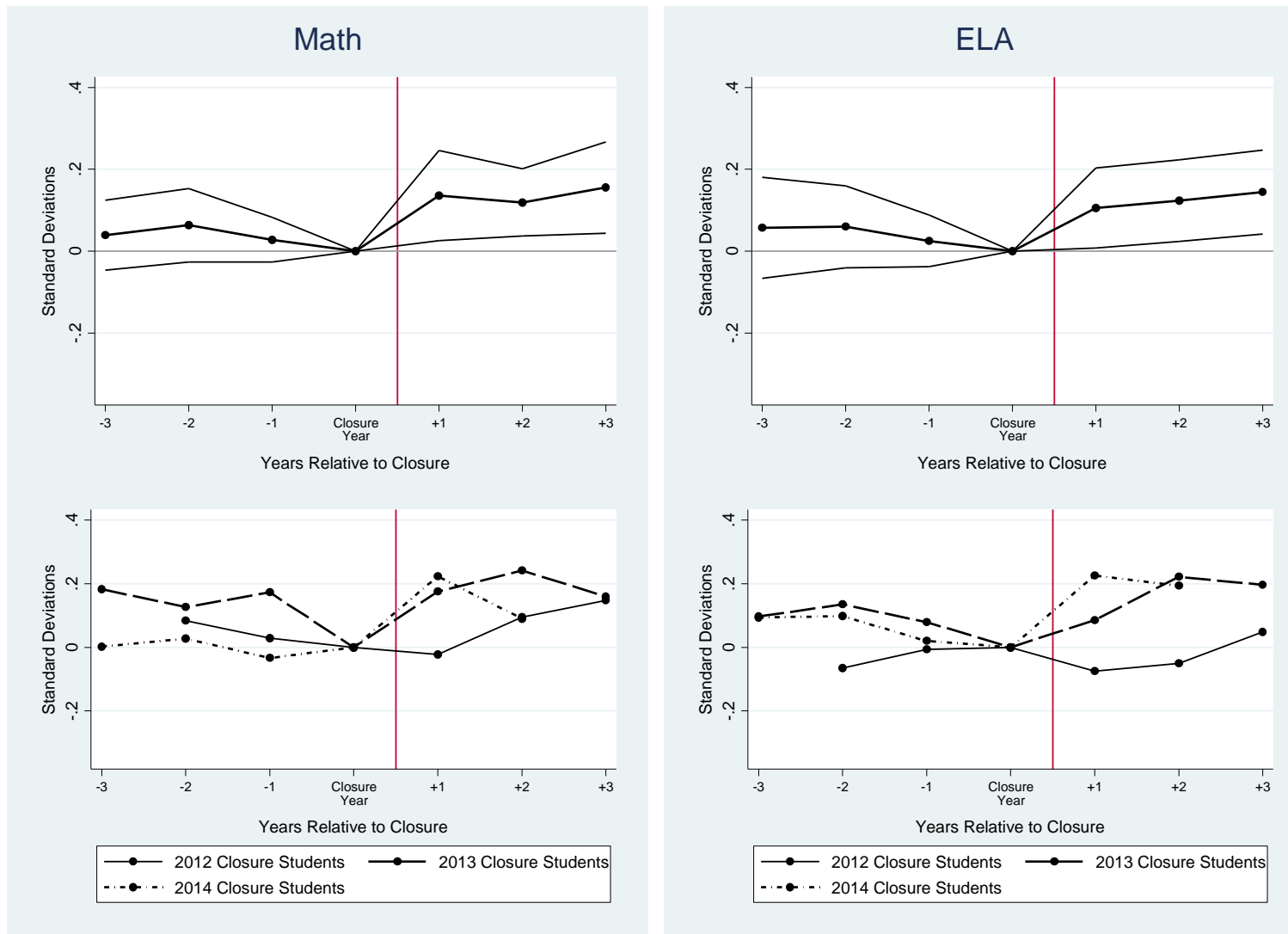
those in the 2012 cohort did not initially. Perhaps importantly, students in the 2012 cohort of closed schools were combined with students in existing schools targeted for renewal and undergoing extensive reforms which could explain the difference in the 2012 closure cohort's outcomes.

Table A1. Changes in Value-Added for Students in Closed Schools in Newark

Change in Closed Schools v. NJ (Relative to Year of Closure)	Math			ELA				
	Pooled	2012 Cohort	2013 Cohort	2014 Cohort	Pooled	2012 Cohort	2013 Cohort	2014 Cohort
3 Years Before Closure	0.039 (0.043)		0.182** (0.090)	0.002 (0.053)	0.057 (0.063)		0.097 (0.073)	0.094 (0.085)
2 Years Before Closure	0.063 (0.046)	0.083 (0.073)	0.126** (0.052)	0.027 (0.068)	0.060 (0.051)	-0.065 (0.096)	0.135*** (0.045)	0.098* (0.059)
1 Year Before Closure	0.028 (0.028)	0.029 (0.060)	0.172*** (0.047)	-0.034 (0.028)	0.025 (0.032)	-0.006 (0.068)	0.079* (0.042)	0.020 (0.041)
1 Year After Closure	0.135** (0.056)	-0.023 (0.074)	0.175*** (0.054)	0.223*** (0.077)	0.105** (0.050)	-0.075 (0.071)	0.085** (0.036)	0.226*** (0.069)
2 Years After Closure	0.119*** (0.042)	0.095 (0.058)	0.241*** (0.070)	0.090 (0.058)	0.124** (0.051)	-0.051 (0.069)	0.221*** (0.060)	0.193*** (0.070)
3 Years After Closure	0.155*** (0.057)	0.147** (0.063)	0.160 (0.118)		0.145*** (0.052)	0.049 (0.071)	0.197*** (0.063)	
2012 Cohort	-0.150*** (0.034)	-0.118*** (0.039)			-0.167*** (0.043)	-0.077 (0.059)		
2013 Cohort	-0.102*** (0.036)		-0.172*** (0.033)		-0.098*** (0.034)		-0.133*** (0.020)	
2014 Cohort	-0.128*** (0.033)			-0.120*** (0.041)	-0.113*** (0.041)			-0.151*** (0.051)
Grade-by-year FE	x	x	x	x	x	x	x	x
Observations	3238680	3233664	3233069	3235075	3230593	3225562	3224976	3226989

Note: Robust standard errors in parentheses clustered at the school level; *** p<0.01, ** p<0.05, *p<0.1. Value-added for the outcome variable is calculated using the same two-step approach based on equation (1) and (2) that was used in the decomposition (See footnote 11).

Appendix Figure A1: Estimated Effects of Closure on Student Value-Added



Note: The top panels plot the estimated effects of closure and 95% confidence intervals from the pooled sample results (columns 1 and 5 of Table A1). The bottom panels plot estimated effects for each closure cohort (columns 2-4 and 6-8 from Table A1).