

Working PAPER

BY HANLEY CHIANG, STEPHEN LIPSCOMB, AND BRIAN GILL

Is School Value Added Indicative of Principal Quality?

November 2012

Revised December 2014

Is School Value Added Indicative of Principal Quality?

Hanley Chiang
(corresponding author)
Mathematica Policy Research
955 Massachusetts Ave.
Suite 801
Cambridge, MA 02139
(617) 674-8374
hchiang@mathematica-mpr.com

Stephen Lipscomb
Mathematica Policy Research
955 Massachusetts Ave.
Suite 801
Cambridge, MA 02139
(617) 674-8371
slipscomb@mathematica-mpr.com

Brian Gill
Mathematica Policy Research
955 Massachusetts Ave.
Suite 801
Cambridge, MA 02139
(617) 301-8962
bgill@mathematica-mpr.com

December 2014

We are grateful to Duncan Chaplin and to conference participants at the 2012 annual meetings of the Association for Education Finance and Policy and the Association for Public Policy Analysis and Management for their helpful comments. This study extends analyses undertaken for the Team Pennsylvania Foundation and the Pennsylvania Department of Education and reported in Lipscomb, Chiang, and Gill (2012). The text reflects the views and analyses of the authors alone and does not necessarily reflect views of Mathematica Policy Research, the Pennsylvania Department of Education, or the Team Pennsylvania Foundation. All errors are the responsibility of the authors.

ABSTRACT

States across the country are developing systems for evaluating school principals on the basis of student achievement growth. A common approach is to hold principals accountable for the value added of their schools—that is, schools’ contributions to student achievement growth. In theory, school value added can reflect not only principals’ effectiveness, but also other school-specific influences on student achievement growth that are outside of principals’ control. In this paper, we isolate principals’ effects on student achievement growth and examine the extent to which school value added captures the effects that principals persistently demonstrate. Using longitudinal data on the math and reading outcomes of 4th through 8th grade students in Pennsylvania, our findings indicate that school value added provides very poor information for revealing principals’ persistent levels of effectiveness.

*This manuscript has been accepted for publication in *Education Finance and Policy*.*

1. INTRODUCTION

In recent years, policymakers have shown a keen interest in evaluating the effectiveness of school principals using performance measures based on student test scores. An increasingly common approach is to measure principal performance using measures of school “value added,” which capture schools’ contributions to student achievement growth. In this article, we empirically assess the extent to which these models provide information about principal effectiveness rather than about factors beyond principals’ control that shape school performance.

Interest in evaluating principals is rooted in the widely held notion that effective leadership is an important characteristic of successful schools. This notion has a long history in qualitative studies of effective schools (Purkey and Smith 1983). Some recent quantitative evidence indicates that school performance is higher when principals are more experienced (Clark, Martorell, and Rockoff 2009; Dhuey and Smith 2013),¹ have greater organizational management skills (Grissom and Loeb 2011), and demonstrate greater ability to recruit and retain high-quality teachers while removing low-quality teachers (Branch, Hanushek, and Rivkin 2012; Loeb, Kalogrides, and Bêteille 2012).

Although principal characteristics and practices might help to identify effective principals, policymakers are particularly interested in using student outcomes directly to measure principal quality for purposes of accountability and incentives. States across the country have begun mandating the incorporation of student achievement growth into principal evaluations. As a condition for receiving either federal Race to the Top funds or enhanced flexibility under the Elementary and Secondary Education Act, more than 40 states have agreed to hold principals

¹ Other research papers, such as Buck (2012) and Dhuey and Smith (forthcoming), find little relationship between principal experience and effectiveness at improving student achievement.

School Value Added and Principal Quality

accountable for student achievement growth. Nevertheless, they have faced challenges in developing outcomes-based measures of principal quality for use in performance evaluations.

In the public eye, principals are often viewed as effective when they lead schools that have high test scores. But test score levels generally are related more strongly to student and family characteristics than to a principal's performance. Value-added models (VAMs) have the potential to provide effectiveness information for principal evaluations because they seek to account for student, family, and neighborhood influences. VAMs have been examined extensively in the context of teacher and school effectiveness.² If a VAM fully accounts for influences beyond teachers' control, it can provide valid measures of a teacher's value added or "effectiveness"—that is, the teacher's contribution to student achievement growth.³ Similarly, if a VAM fully accounts for out-of-school influences on achievement, it can provide valid measures of a school's value added.⁴

Only recently have researchers applied the value-added methodology to principal effectiveness (Branch, Hanushek, and Rivkin 2012; Cannon, Figlio, and Sass 2012; Coelli and Green 2012; Dhuey and Smith 2013, forthcoming; Grissom, Kalogrides, and Loeb forthcoming; Lipscomb, Chiang, and Gill 2012). Estimating principal value added is not as straightforward as estimating teacher or school value added. The key analytical challenge of devising a principal

² For example, see Aaronson, Barrow, and Sander (2007); Chetty, Friedman, and Rockoff (2011); Deming (2014); Deutsch (2012); Glazerman et al. (2010); Goldhaber and Hansen (2010); Kane and Staiger (2008); Ladd and Walsh (2002); McCaffrey et al. (2009); Rivkin, Hanushek, and Kain (2005); and Rothstein (2010).

³ Throughout this article, we use the terms "value added" and "effectiveness" interchangeably, with both terms denoting contributions to student achievement growth.

⁴ An additional concern with the validity of school value-added estimates is that a school's estimated effectiveness in a given year may be influenced by its effectiveness in previous years, because schools, unlike teachers, serve many of the same students in consecutive years. Recent studies by Deutsch (2012) and Deming (2014), however, have validated commonly-used school VAMs by showing that they are unbiased predictors of experimentally identified school effectiveness estimates from school choice lotteries in Chicago and Charlotte-Mecklenburg.

School Value Added and Principal Quality

VAM is to disentangle principals' true contributions to student achievement growth from the influence of other school-level factors beyond principals' control. The authors of existing studies have addressed this challenge by using school leadership changes to identify a principal's effectiveness relative to other principals who have served at the same school. By controlling for school-level factors outside of principals' control that do not change over time, this approach can yield plausibly valid estimates of principal effectiveness. However, a major limitation of this approach is that it can only evaluate principals of schools that experience a leadership transition during the period in which performance is being evaluated. Therefore, this method cannot provide a useful basis for identifying principal quality in actual evaluation systems, which must include all principals.

In contrast, school value-added estimates can be generated for all principals, but they are imperfect indicators of principal quality. School value added combines principals' contributions to improved student outcomes with the contributions of other school staff and resources. If principals exert a strong influence on a school's effectiveness, then holding them accountable for the overall effectiveness of their school may be sensible from a management standpoint. Holding principals accountable for overall school effectiveness makes less sense if most of the variation in school effectiveness is due to factors outside of a principal's control.

Several states receiving Race to the Top funds and enhanced flexibility under the Elementary and Secondary Education Act have proposed to use measures of school-wide effectiveness as indicators of principal effectiveness. For example, Ohio, North Carolina, Pennsylvania, and Tennessee are proceeding to use school value-added estimates from the Education Value-Added Assessment System (EVAAS) in principal evaluations. School value-added scores have been or will be a determinant of performance bonuses for principals in several

School Value Added and Principal Quality

incentive programs, including the Teacher Advancement Program and the Teacher Incentive Fund.

In this article, we examine whether a school VAM is a valid tool for measuring principal effectiveness. We base our conclusions on the degree to which schools' effectiveness estimates are correlated with the effects on student achievement that their principals persistently demonstrate. The core of our analysis assesses the strength of the association between two types of value-added estimates: (1) school value-added estimates and (2) principal value-added estimates that isolate principals' effects from other persistent, unobserved differences across schools. To obtain these estimates, we use a longitudinal database of all Pennsylvania students in grades 4 through 8 from 2007–2008 to 2012–2013. After obtaining these value-added estimates, we employ a regression model to assess the extent to which differences in schools' value added predict differences in the value added of their principals.

Two key elements of our empirical strategy are designed to isolate the portion of a principal's effect that is consistent across time and student samples (and therefore presumably of greatest importance to evaluators). First, the school and principal value-added estimates come from different time periods. Our analysis therefore provides information about the degree to which school effectiveness estimates are indicators of persistent differences in principals' effectiveness. Second, we compare school and principal effectiveness estimates obtained from distinct groups of students. This avoids correlated sampling errors—situations in which there is a spurious correlation between school effectiveness and principal quality because, for instance, an unusually bright cohort of students is falsely inflating both a school's value-added estimate and the value-added estimate of its principal.

School Value Added and Principal Quality

Our findings pertain to a specific population of principals. To have value-added estimates, principals in the analysis must have been involved in leadership transitions. They either became the new principal of a school or were replaced by an incoming principal during the final four years of our analysis period—2009–2010 to 2012–2013. Our findings thus pertain to principals who either are in the first four years at their current positions or are being compared with other principals in the first four years at their positions when their effectiveness is measured. This does not severely limit the external validity of our findings; in 2012–2013, 55 percent of school principals in Pennsylvania were in their first four years at their current positions among those with students in the grades included in our principal VAMs.

Our findings indicate that school value added provides very poor information for evaluating principals' persistent levels of effectiveness. In both math and reading, there is no statistically significant relationship between school value added and principals' persistent levels of effectiveness. The magnitudes of the relationships are also small, implying that no more than 7 percent of any given difference in true school value added between two schools reflects persistent differences in the effectiveness of their current principals.

Our study contributes to the literature on performance-based indicators of leadership quality that are based on individual-level data by assessing the validity of measures that could be used in actual evaluations. In the most closely related study, Lipscomb, Chiang, and Gill (2012) found that principal and school effectiveness estimates are moderately correlated when those estimates pertain to the same school years and are obtained using the same student samples. We do not emphasize comparing school and principal effectiveness within the same time period and for the same students in this study because any observed relationship could reflect, in part, completely nonpersistent principal effects and correlated sampling errors. That is, we are not merely focused

School Value Added and Principal Quality

on whether school and principal value-added estimates come to the same (possibly erroneous) conclusions; instead, we analyze how closely school value added reflects principals' *true* and *persistent* levels of effectiveness.

The rest of the relevant research literature primarily has focused on describing the methodology for estimating principal effects and examining characteristics of the effectiveness distribution. Branch, Hanushek, and Rivkin (2012) estimate that math scores in Texas are 0.11 standard deviations higher in schools led by principals whose within-school VAM estimate is one standard deviation higher. Dhuey and Smith (forthcoming) measure the cumulative effects of middle-school principals in British Columbia, Canada over three grade levels, finding a within-school standard deviation of principal effectiveness that is 0.36 in math and 0.21 in reading. In a subsequent paper, Dhuey and Smith (2013) measure the effects of North Carolina principals, finding that principal-school match quality accounts for a significant portion of the variation in principal value added. Cannon, Figlio, and Sass (2012) also conclude that principal match quality matters, based on evidence from Florida that the persistence of a principal's value-added estimate declines when a principal changes schools. Coelli and Green (2012) explore models that allow principal effects to vary during the years of a principal's tenure at a school, finding that principals need several years as a school's leader to have their full effect on student outcomes.

Our findings do not contradict the evidence presented by other studies showing that principals differ meaningfully in their effectiveness. Instead, our paper shows that differences in principal effectiveness are not strongly predicted by measures of school value added. Moreover, as demonstrated by Grissom, Kalogrides, and Loeb (2012), school value-added measures predict non-test evaluation scores, such as observation-based and survey-based measures, more closely than principal value-added measures do. When combined with our findings, this result suggests

School Value Added and Principal Quality

that non-test evaluation measures may also be misattributing to principals the influence of school factors outside a principal's control.

2. EMPIRICAL METHODS

Conceptual Framework

The key objective of the empirical analysis is to determine the extent to which school value added predicts principals' effectiveness. Toward this objective, it is useful to consider a simple framework that decomposes school value added into various components. Let Sva_{spt} be the total true contribution (without estimation error) of school s to student learning—that is, the effectiveness or value added of school s —in grade g under the leadership of principal p in year t . The school's value added reflects the value added of its principal (Pva_{pgt}) in that grade and year as well as the influence of all other school-level factors (F_{sgt}) outside of the principal's control:

$$Sva_{spt} = Pva_{pgt} + F_{sgt} . \tag{1}$$

Measures of school effectiveness are informative of principal quality only insofar as they reflect Pva_{pgt} rather than F_{sgt} . Differences in F_{sgt} across schools could arise from several types of factors beyond principals' control. Some portion of the variation in teacher quality across schools could be outside of principals' discretion. For example, aspects of a school's physical location (such as accessibility or proximity to an education school) could make the school more or less attractive to good teachers. The composition of a school's teaching force might also reflect hiring decisions made by district offices. In addition, variation in F_{sgt} could stem from differences in financial resources, especially if schools from different districts are compared.

Of the variation in school effectiveness that is due to principals' effects, not all of the sources of this variation are of equal interest in evaluating principals. We assume evaluators

School Value Added and Principal Quality

want to gauge effects on achievement that principals' *persistently* demonstrate. In contrast, transitory impacts—that is, evidence of effectiveness that a principal exhibits in one year but not in the next year—are not relevant to predicting which individuals will be good school leaders in subsequent years.

To consider the variation in principals' effectiveness, we conceptualize total true principal effectiveness in a particular grade and year as the sum of four components:

$$Pva_{pgt} = \theta_p + \theta_{pg} + \theta_{pt} + \theta_{pgt}, \quad (2)$$

where θ_p is the component of a principal's effect that is common across years and grades; θ_{pg} is the component of a principal's effect that is specific to grade g but persistent across years; and θ_{pt} and θ_{pgt} represent the principal's nonpersistent impacts, respectively, in specific years and grade-year combinations. We define each of the four components of Pva_{pgt} to be independent of the other components. Moreover, we assume that the transitory components (θ_{pt} and θ_{pgt}) are independent of F_{sgt} and that $Cov(\theta_{pg}, \theta_{pg'}) = 0$ for $g \neq g'$, $Cov(\theta_{pt}, \theta_{pt'}) = 0$ for $t \neq t'$, and $Cov(\theta_{pgt}, \theta_{pg't'}) = 0$ for $g \neq g'$ or $t \neq t'$.

In our notation, we assume that evaluators are interested in measuring $[\theta_p + \theta_{pg}]$ for each principal. The first of these two components represents a persistent ability of principals to improve student outcomes across the grades at their schools. The second component recognizes that principals may be more effective in some grades than in others on a persistent basis. For example, a principal might be particularly knowledgeable about identifying appropriate curricula for the early elementary grades but not the upper elementary grades.

This framework yields the following key question to be addressed by the analysis: *To what extent do differences in school value added across schools predict differences in the persistent*

School Value Added and Principal Quality

effectiveness of their current principals? In other words, the objective is to determine the extent to which differences in Sva_{spgt} predicts differences in $[\theta_p + \theta_{pg}]$.

We focus the analysis on predicting $[\theta_p + \theta_{pg}]$ with a linear function of Sva_{spgt} . Consider a scenario in which both Sva_{spgt} and $[\theta_p + \theta_{pg}]$ could be observed perfectly without error. If we ran a linear regression of $[\theta_p + \theta_{pg}]$ on Sva_{spgt} , the coefficient on Sva_{spgt} , denoted by β , would measure the difference in the persistent component of principal value added that would be predicted from observing a one-unit difference in school value added. A coefficient of one would imply that school value added was an unbiased linear predictor of principals' effectiveness.

The size of β is determined by two key factors. These factors are highlighted by noting that

$$\beta = \frac{Cov(\theta_p + \theta_{pg}, Sva_{spgt})}{Var(Sva_{spgt})} = \frac{Cov(\theta_p + \theta_{pg}, \theta_p + \theta_{pg} + \theta_{pt} + \theta_{pgt} + F_{spgt})}{Var(Sva_{spgt})} = \frac{Var(\theta_p + \theta_{pg}) + Cov(\theta_p + \theta_{pg}, F_{spgt})}{Var(Sva_{spgt})} \quad (3)$$

First, β is larger when variation in principals' persistent levels of effectiveness ($Var(\theta_p + \theta_{pg})$) represents a greater fraction of total variation in school value added ($Var(Sva_{spgt})$). Conversely, variation in school value added due to non-principal factors or nonpersistent components of principal quality drives down β . Second, β is larger if more effective principals in a grade are assigned to schools with more positive additional influences on student learning in that grade ($Cov(\theta_p + \theta_{pg}, F_{spgt}) > 0$). Compensatory assignment—with more effective principals assigned to schools in which other factors depress learning growth—will tend to mask principals' impacts on their schools.

In practice, we cannot directly observe $[\theta_p + \theta_{pg}]$. At best, we can obtain estimates of principal value added based on a finite set of years and students. These estimates reflect not only

School Value Added and Principal Quality

$[\theta_p + \theta_{pg}]$ but also transient components of principal effectiveness and random sampling error stemming from any idiosyncratic characteristics of the particular students used in estimation.

School value added also must be estimated on a particular sample of years and students. Therefore, school value-added estimates also reflect transient components of principal effectiveness and sampling error, among other factors. If the years or students used to estimate school value added overlap with those used to estimate principal value added, then the two sets of estimates will be correlated, in part, due to reflecting the same values of θ_{pt} or θ_{pgt} or the same idiosyncratic student characteristics. In this case, the coefficient from a linear regression of principal value-added estimates on school value-added estimates would be biased upward, measuring more than just the relationship between school value added and $[\theta_p + \theta_{pg}]$.

Our empirical strategy avoids these biases by estimating principal and school value added on *independent* samples—separate sets of years and students. As a result, the school value-added estimates can be associated with the principal value-added estimates only as a result of being associated with $[\theta_p + \theta_{pg}]$. In the remainder of this section, we discuss the models used for estimating principal and school value added and our approach to assessing the relationship between these estimates.

Estimating Principal Value Added

To identify principals' effects on student achievement, we exploit leadership transitions within schools—that is, instances in which one principal replaces another at a school—and assess the within-school changes in student outcomes induced by these leadership transitions. Variants of this strategy have been used in all prior research on principal value-added estimators.

School Value Added and Principal Quality

To implement this approach, we estimate the following principal VAM for the outcome test score, y_{ispgt} , of student i in grade g within school s led by principal p in year t :

$$y_{ispgt} = \mathbf{X}_{ispgt} \boldsymbol{\delta} + Pva_{pg\bullet} + \alpha_{sg\bullet} + \varepsilon_{ispgt} \quad (4)$$

where \mathbf{X}_{ispgt} is a vector of covariates (without an intercept), $Pva_{pg\bullet}$ is a grade-specific principal fixed effect, $\alpha_{sg\bullet}$ is a grade-specific school fixed effect, and ε_{ispgt} is a student-level error term.

The principal and school fixed effects are coefficients on vectors of principal and school dummy variables, respectively, and represent average effects across all years used in the analysis.⁵ As we discuss later, for the purposes of our empirical strategy it is advantageous to estimate Equation (4) separately by grade.

The outcomes of interest are math and reading scores on state assessments, which we standardize to have mean zero and standard deviation one within the full statewide population of test takers in each grade and year. The student-level covariates include prior-year test scores in both math and reading, gender, race/ethnicity, free meals and reduced-price meals participation, English language learner (ELL) status, special education status, and whether the student's outcome and baseline tests had modifications. We also control for school-by-year-by-grade averages of various student-level characteristics (gender, race/ethnicity, subsidized lunch status, ELL status, and special education status) and year dummies.

⁵ Because students might be observed in more than one school in a given year due to mobility, the “dummy” variables are not purely dichotomous. For each student-by-year observation, we allow the value of a school “dummy” variable to vary continuously from 0 to 1 to represent the implied fraction of the school year in which the student was attending the given school in the given year. For each student-by-year observation, the value of a principal dummy variable is identical to the value of the dummy variable for the school that the principal leads in that year; therefore, principal dummy variables can also vary continuously from 0 to 1. We do not have exact measures of the fraction of the year in which a student attends a school; we impute this fraction with the reciprocal of the number of schools in which the student is observed in a given year.

School Value Added and Principal Quality

To estimate Equation (4), we use data on outcome scores from the school years 2008–2009 to 2012–2013 for only those students attending a school that ever experienced a leadership transition during that period. Principals in charge of fewer than 20 student-years during the analysis period are excluded from the sample.

Each principal's estimated fixed effect, $P\hat{v}a_{pg}$, serves as the principal's value-added estimate. Differences in value added across principals are measured in terms of standard deviations of student-level scores. Because the model conditions on school fixed effects, only changes over time in the identities of principals leading a school are used to identify the principal fixed effects. This purges the estimated principal effects of all unobserved, school-specific influences on student achievement that remain invariant over the period of analysis, such as persistent between-school differences in resources or neighborhood quality.

Despite improving the causal validity of the estimated principal effects, there are several costs to including school fixed effects in the principal VAM. First, the VAM generates effectiveness estimates only for principals who have led schools with a leadership transition during the analysis period. Second, for those principals with estimates, the VAM allows each principal to be compared only to a limited set of other principals. Naturally, the most direct comparisons are between principals who have served at the same school. If, after controlling for all covariates, student outcomes at a given school are better under a successor than under a predecessor, then the successor will have a more positive value-added estimate than the predecessor.

Beyond these direct comparisons, there is a somewhat broader—but still limited—group to which each principal can be compared. Comparisons can be made among principals who have served in the same connected network of schools, where a network is a set of schools in which

School Value Added and Principal Quality

every school has had at least one principal transfer to at least one other school in the network during the analysis period. We use the phrase “network” to refer, interchangeably, to either the set of schools connected by these transfers or the set of principals who ever served in any of those schools. Figure 1 provides a hypothetical example of how comparisons can be made within a network of two schools (schools A and B) and three principals (principals 1, 2 and 3). Based on this scenario, the VAM makes a comparison within school B to estimate that principal 3 is more effective than principal 1 by 0.2 units, and makes a comparison within school A to estimate that principal 1 is more effective than principal 2 by 0.2 units. By the assumption of transitivity, principal 3 is deemed to be more effective than principal 2 by 0.4 units. In contrast, without the inclusion of school fixed effects, principal 2 would be regarded as more effective than principal 3 due to the fact that student outcomes are generally better at school A than at school B.

As noted by Dhuey and Smith (forthcoming), the raw estimates of Equation (4) force one principal per network to have a value-added estimate of zero. We follow previous studies in re-centering value-added estimates by network, so that the resulting estimate for every principal is expressed relative to the average principal in the same network. We recalculate standard errors of the value-added estimates accordingly.

As we discuss in Section 3, the vast majority of principals with VAM estimates from Equation (4) belong to networks that are no larger (and usually smaller) than the scenario shown in Figure 1. Therefore, the principal VAM generally cannot be applied to a real-world evaluation system that seeks to assess a principal’s effectiveness relative to a much broader comparison group. Nevertheless, Equation (4) represents the best available method for obtaining principal effectiveness estimates purged of time-invariant, school-specific influences beyond the principals’ control, and we can use these estimates to assess the usefulness of more widely

School Value Added and Principal Quality

applicable performance measures—measures of school effectiveness—as tools for evaluating principals. We turn next to our strategy for estimating school effectiveness.

Estimating School Value Added

We estimate the effectiveness of each principal’s school in the 2007–2008 school year—a year prior to the period on which the principals’ own effectiveness measures are based. Using student outcome data from that year, we estimate a school VAM of the following form, separately by grade:

$$y_{ispg,2007-08} = \mathbf{X}_{ispg,2007-08} \boldsymbol{\delta} + Sva_{spg,2007-08} + \varepsilon_{ispg,2007-08} \quad (5)$$

where $Sva_{spg,2007-08}$ is the grade-specific fixed effect for school s , capturing the value added of that school in 2007–2008. All other variables of Equation (5) are defined as before. We control for the same covariates as in the principal VAM except for the year dummies and school-by-year-by-grade averages of student characteristics, given that Equation (5) is estimated on one year of outcome data. The schools in the estimation sample for Equation (5) consist only of schools that have at least 20 students in the given grade in 2007–2008 and are led by principals who have principal value-added estimates from Equation (4). For direct comparability with the principal value-added estimates, we group school value-added estimates from 2007–2008 into networks according to the networks to which their *principals* belong in the 2008–2009 to 2012–2013 period (the years of the principal VAM). We then re-center the school value-added estimates by network.

Estimating the Association Between School and Principal Value Added

In the final stage of the analysis, we examine the extent to which principals' value added in the 2008–2009 through 2012–2013 school years can be predicted by the value added of the schools that they led in an earlier year, 2007–2008. Our basic approach is to estimate a regression at the principal-by-grade level, with grade-specific principal value-added estimates regressed on school value-added estimates from the same grade.

Our empirical strategy removes factors that could otherwise lead to spurious associations between the principal and school value-added estimates. As stated earlier, the advantage of estimating principal and school value added from different sets of years is that completely transitory elements of principal effectiveness (θ_{pt} and θ_{pgt}), which are likely not of interest to evaluators, cannot contribute to a correlation between the two types of value-added estimates. Another advantage is that the same *students* do not contribute to both the school and principal value-added estimates being paired. This prevents correlation in the sampling error of the two estimates.

To do so, we pair grade-specific school value-added estimates from 2007–2008 with same-grade principal value-added estimates from 2008–2009 through 2012–2013. This approach yields close-to-independent samples, with the only overlap consisting of a small number of students who repeat the same grade across years. Formally, our dataset consists of the pairs $\{S\hat{v}a_{spg,2007-08}, P\hat{v}a_{p,g,\bullet}\}$, which we refer to as principal-grade observations, for each of the grades g from 4 to 8.

To estimate the association between school and principal value added, we pool together all principal-grade observations across all grade levels and estimate the following specification:

$$P\hat{v}a_{p,g,\bullet} = \gamma(S\hat{v}a_{spg,2007-08}) + \mathbf{D}_{p,g} \boldsymbol{\mu} + \omega_{p,g} \quad (6)$$

School Value Added and Principal Quality

where $\mathbf{D}_{p,g}$ is a vector of network dummies. Inclusion of network dummies appropriately forces comparisons to be made only among principals in the same network. To improve the efficiency of the estimates of Equation (6), we weight each principal-grade observation by the inverse of the squared standard error of the principal value-added estimate.

Of central interest in the study is whether the coefficient estimator from Equation (6), $\hat{\gamma}$, provides a consistent estimator of the true relationship (β) between school value added and the persistent component of principal effectiveness. To examine this, it is instructive to express each set of value-added estimates in terms of various components. Let T denote the set of five school years (2008–2009 through 2012–2013) on which the principal VAM is estimated. Using Equation (2) and the fact that $P\hat{v}a_{p,g,\bullet}$ measures a principal's effectiveness with some sampling error, we have

$$P\hat{v}a_{p,g,\bullet} = \theta_p + \theta_{p,g} + \frac{1}{5} \sum_{t \in T} \theta_{pt} + \frac{1}{5} \sum_{t \in T} \theta_{p,g,t} + u_{p,g}, \quad (7)$$

where $u_{p,g}$ represents sampling error and all other terms are defined as in Equation (2). A similar decomposition based on Equations (1) and (2) yields

$$S\hat{v}a_{spg,2007-08} = \theta_p + \theta_{pg} + \theta_{p,2007-08} + \theta_{pg,2007-08} + F_{sg,2007-08} + v_{sg,2007-08} \quad (8)$$

where $v_{sg,2007-08}$ represents sampling error and all other terms are defined as in Equations (1) and (2).

Using the facts that θ_{pt} and θ_{pgt} are uncorrelated across years and the sampling errors from (7) and (8) are uncorrelated with each other, $\hat{\gamma}$ converges in probability to the following:

$$\hat{\gamma} \rightarrow \frac{\text{Cov}(P\hat{v}a_{p,g,\bullet}, S\hat{v}a_{spg,2007-08})}{\text{Var}(S\hat{v}a_{spg,2007-08})} = \frac{\text{Var}(\theta_p + \theta_{pg}) + \text{Cov}(\theta_p + \theta_{pg}, F_{sg,2007-08})}{\text{Var}(Sva_{spg,2007-08}) + \text{Var}(v_{sg,2007-08})}$$

School Value Added and Principal Quality

$$= \beta \times \frac{\text{Var}(Sva_{spg,2007-08})}{\text{Var}(Sva_{spg,2007-08}) + \text{Var}(v_{sg,2007-08})}. \quad (9)$$

Thus, $\hat{\gamma}$ converges to the true desired relationship, β , multiplied by a factor less than one, with that factor representing the reliability of the school value-added estimates. This result illustrates the well-known fact that measurement error in the independent variable—which, in this case, stems from the use of a finite student sample to estimate school value added—leads to attenuation bias in its coefficient estimate.

As done in Jacob and Lefgren (2005, 2008), we correct the attenuation bias by converting $S\hat{v}a_{spg,2007-08}$ to an empirical Bayes (EB) estimate. The EB adjustment “shrinks” or pulls the ordinary least squares (OLS) estimate toward the average estimate in a defined group of principals by a factor equal to the unreliability of the OLS estimate. Substituting the EB estimate for the OLS estimate of school value added on the right-hand side of Equation (6) exactly offsets the attenuation bias, enabling $\hat{\gamma}$ to be a consistent estimator of β .

To construct the EB estimates of school value added, we first take the within-network sample variance of the OLS estimates and subtract the average squared standard error of those estimates.⁶ This difference, denoted by V_{true}^{school} , measures the variance of *true* school value added—that is, the variance that would occur if school value added were observed without error. The EB estimate of the value added of school s in grade g is then equal to the corresponding OLS estimate multiplied by $(V_{true}^{school}) / (V_{true}^{school} + SE_{sg}^2)$, where SE is the standard error of the OLS estimate. This approach implicitly shrinks the OLS estimates toward zero, given that the OLS

⁶ In these calculations, each principal-grade observation continues to be weighted by the inverse of the squared standard error of the principal value-added estimate, for consistency with how Equation (6) is estimated.

School Value Added and Principal Quality

estimates of school value added average to zero within each network. We then substitute the EB estimates of school value added in place of the OLS estimates in Equation (6).

In short, our empirical strategy produces a consistent estimator of the extent to which school value added can predict persistent elements of a principal's effectiveness. The coefficient estimate from Equation (6) can be used to answer the following question: *For any given difference in true school value added between two schools, what proportion of that difference, on average, reflects persistent differences in the effectiveness of their current principals?*

As an alternative way to gauge how strongly school value added signals principals' effectiveness, we also transform the regression coefficient into an R-squared value, again adjusted for estimation error. The R-squared answers the following question: *Even without estimation error, what proportion of the variation in principals' subsequent effectiveness can be predicted from information revealed by school value added?* To obtain the R-squared value, we square the regression coefficient and multiply it by $(V_{true}^{school} / V_{true}^{principal})$, where V_{true}^{school} is defined above and $V_{true}^{principal}$ is the variance of true principal value added in the absence of sampling error, which we estimate using the same methods used to estimate V_{true}^{school} .

When interpreting the estimates of Equation (6), it is important to consider the types of principals to whom these results can be generalized. In order to have principal value-added estimates, all principals in the analysis must be involved in a leadership transition between some pair of successive years during the analysis period of the principal VAM. This means that in at least one of the final four years of that period—2009–2010 through 2012–2013—the principals in the analysis either started leading a school they had not led before or were replaced by an incoming principal. Therefore, the results of this paper indicate how well school value added predicts principals' effectiveness when principal value added is measured in the principals' first

School Value Added and Principal Quality

four years at their positions or is measured relative to others who are in their first four years at their positions. As indicated above, 55 percent of Pennsylvania principals leading schools serving grades 4 through 8 in 2012–2013 began their current position no more than four years earlier. Thus, our findings are relevant to a majority of the state’s principal workforce.

3. DATA AND DESCRIPTIVE STATISTICS

Data Sources

Our data come from longitudinally-linked student-level files obtained from the Pennsylvania Department of Education on all public school students in the state. The first set of files contains student achievement scores from the Pennsylvania System of School Assessment (PSSAs) in math and reading for grades 3 to 8 from 2006–2007 to 2012–2013. The PSSAs are standardized tests that Pennsylvania uses for compliance with federal school accountability policies. Nearly 97 percent of all Pennsylvania students in these grades have a PSSA scaled score from 2012–2013.⁷ We link these scores to a second set of files containing administrative records from the Pennsylvania Information Management System (PIMS) on students in grades 4 to 8 from 2007–2008 to 2012–2013. PIMS data include information on students’ gender, race/ethnicity, free and reduced-price meal participation, ELL status, and special education category. PIMS also allows us to link students to the schools they attended during the year and the principal that led each respective school. Altogether, we have current and prior achievement scores and other

⁷ The PSSAs were available in both regular and modified versions until 2012-2013 when the modified version was discontinued. Two percent of all students with scores took the modified test when it was offered, which was intended for some special education students. We include these scores in the VAMs with an indicator for taking the modified version. Of students without any PSSA score, nearly half took the Pennsylvania Alternate System of Assessment (PASA). The PASA is intended for students with severe cognitive impairments.

School Value Added and Principal Quality

information on students from six school years. We use these data to study school effectiveness in the first year and principal effectiveness across the five subsequent years.

Principal Transitions

As indicated by Table 1, 41 percent of Pennsylvania's schools serving 4th through 8th graders experienced at least one leadership change between 2008–2009 and 2012–2013.⁸ The annual rate of school leadership changes during these years ranged between 10 and 12 percent. A total of 3,269 principals served at 2,532 total schools during these five years; of these principals, 1,929 served at one of the 1,034 schools where a leadership transition occurred.⁹ We estimate value-added scores for these 1,929 principals using the estimation strategy described in Section 2, which yields effectiveness data for 59 percent of all principals in our sample.

Relative comparisons of principal effectiveness are made within small connected networks of schools and principals. In Table 2, we report the number of networks and their sizes. We focus on grade-specific networks in the table because our principal effectiveness models are estimated separately by grade. Of 2,155 connected networks representing 5,238 principal-grade observations, 1,629 networks (76 percent) include just one school, which means that the school

⁸ The percentage of Pennsylvania's schools serving 4th through 8th graders that experienced at least one leadership transition increases to 48 percent if we include data from 2007–2008, which we use for calculating school effectiveness estimates.

⁹ Some schools appear in the data to be led either by co-principals or jointly by a single principal. We restricted our sample to only include principals, schools, and years in which a single principal led a single school. This reduces the number of principals involved in leadership changes that we report in any pair of adjacent years in Table 1. For instance, the principal VAM would record one total transition between 2008–2009 and 2012–2013 for a school led by principal A in 2008–2009, jointly by principals A and B in 2009–2010, and then by principal B alone starting in 2010–2011. Transitions would not be recorded between adjacent years because 2009–2010 would be excluded for that school.

School Value Added and Principal Quality

experienced a leadership transition but none of the principals involved were observed at any other school. Only 57 networks include four or more schools.

The high frequency of single-school networks in the data means that most principals are being compared to their predecessor or successor only. Although such a narrow set of comparisons would not suffice for a real evaluation system, it still permits a clear test of the validity of school value added. A *necessary* (although not sufficient) condition for school value added to be a valid measure of principal effectiveness is that it should identify which principals are better than others in a subsequent time period, even if each principal is just being compared to one other principal. Studies to validate teacher value-added measures (Kane et al. 2013; Chetty, Friedman and Rockoff 2014) have taken a similar approach, examining whether teacher value-added estimates can predict future performance differences across teachers that are rigorously estimated (through experimental or quasi-experimental methods) but confined to narrow comparisons within the same grade and school.

We can include only a subset of the 5,238 principal-grade observations on 1,929 principals in the final analysis due to several necessary sample restrictions. First, we require both that a principal's own effectiveness estimate must be based on assessment data from at least 20 students and that another principal in the same network must have an effectiveness estimate based on at least 20 students taking the PSSA. We exclude estimates for principals not meeting these thresholds because those estimates are likely to be heavily influenced by the performance of a few students. As indicated by Table 3, imposing this restriction reduces our sample to 5,059 principal-grade observations on 1,881 principals. Second, we can only use effectiveness data on principals who led a school with students in the same grade in 2007–2008, reducing our sample further to 2,001 principal-grade observations on 802 principals. Third, each principal network in

School Value Added and Principal Quality

the final analysis sample must include at least two principals after the sample is restricted to principals who had led schools in 2007–2008. This restriction removes from the sample many principals who were originally in two-principal networks because the other principal was removed through either of the first two restrictions. This restriction ensures that all principals in the final analysis sample have at least one other principal to serve as a basis for comparing effectiveness estimates. Following these reductions, our final analysis sample includes 673 principal-grade observations on 291 principals.

Characteristics of Principals and Students

We provide means and standard deviations on several principal and student characteristics in Tables 4 and 5, respectively. Table 4 shows that the professional and demographic background characteristics of principals in the final sample were generally similar to those of the statewide population of principals. The few differences observed were small; for instance, principals in the final sample were slightly less likely than all principals in the state to be white (80 versus 87 percent) or to have at least a master’s degree (78 versus 85 percent). Principals in the final sample had an average of two more years of total experience in K–12 education, which is not surprising given that they needed to have led a school in both 2007–2008 and at least one subsequent year.

The student-level characteristics in Table 5 include current- and prior-year PSSA scores, gender, race/ethnicity, free and reduced-price meal participation, ELL status, and special education status. These variables are used in the principal and school VAMs. In the first column of data, we show statistics for Pennsylvania students in grades 4 to 8 from 2007–2008 to 2012–2013. Average test scores in z -score units are not exactly zero because we only include students

School Value Added and Principal Quality

with complete data on all of the characteristics. In the next two columns, we show statistics on the same variables for students attending schools led, respectively, by principals with effectiveness estimates and by principals in the final analysis sample. The last column shows descriptive statistics for students in grades 4 to 8 in 2007–2008 who contribute to school VAM estimates used in the final analysis sample. Overall, the data indicate that school leadership changes are more frequent in schools with lower-performing students. In addition, the samples for the final analysis have larger proportions of students eligible for free meals and black students than are typical across Pennsylvania, suggesting that our results are based on students who are more socioeconomically disadvantaged.

4. RESULTS

Main Findings

Table 6 presents findings from the estimation of Equation (6), assessing the extent to which school value added predicts subsequent principal value added. We show separate results for math and reading outcomes. The coefficient estimates—0.07 for math and -0.03 for reading—are statistically insignificant.¹⁰ The point estimates indicate that no more than 7 percent of any given difference in value added between two schools reflects persistent differences in the effectiveness of their current principals. As an alternative interpretation of this relationship, the implied R-squared value for the math estimate indicates that information about principals' persistent levels of effectiveness revealed by school value added explains only 1 percent of the variation in principals' subsequent effectiveness. In both subjects, school effectiveness does not appear to

¹⁰ We also estimated models where school effectiveness in reading (math) was used to predict principal effectiveness in math (reading). The relationships were statistically insignificant in both subjects.

School Value Added and Principal Quality

provide evaluators with useful information for predicting the subsequent performance of principals.

Our central finding that school value added is a poor predictor of principal value added would be meaningless unless our sample actually had considerable variation in principal value added. In fact, principals in the sample vary substantially in their value added, even within the same network. We find that the within-network standard deviation of principal effects—after removing variation attributable to random sampling error—is 0.14 student-level standard deviations in math and 0.11 student-level standard deviations in reading. To put this variation in context, the standard deviation of principal effects in our sample is at least 80 percent of the size of the standard deviation of teacher effects based on Hanushek and Rivkin's (2010) synthesis. Nevertheless, very little of this variation in principal effectiveness can be predicted by school value added.

We next conduct several additional analyses to assess the robustness of the main findings, including exploring four alternative VAM specifications, different samples of principal-grade observations, several subgroups of principals, and measures of principal value added that account explicitly for the possibility that principals' effects are only gradually manifested after starting to lead a school.

Alternative VAM Specifications

There is currently debate among researchers about how to measure school and principal effects using a VAM, so it is important to explore the sensitivity of our main findings to alternative approaches. For example, Ehlert, Koedel, Parsons, and Podgursky (2013) compared school effects obtained under different VAM specifications. While the correlations across

School Value Added and Principal Quality

specifications typically were at or above 0.90, specifications that did not aggressively control for student background characteristics were more likely to assign higher effectiveness scores to schools serving larger shares of students with more-advantaged characteristics. The authors recommend a two-step estimation strategy where school effects (and by extension principal effects) are completely orthogonal to background characteristics. They argue that school fixed effects otherwise are potentially biased because the coefficients on the student background variables are attenuated due to relatively little within-school variation in the covariates.

We replicate our main analyses using the Ehlert et al. (2013) approach along with three additional conceptualizations¹¹: (1) using student growth percentiles as the outcome variable in lieu of including covariates in the VAMs, (2) estimating principal value added without conditioning on students' prior-year scores, and (3) estimating principal value added omitting a principal's first year leading a school. The student growth percentiles model (see Betebenner 2009) is a well-known approach that is used in Colorado and in other districts and states.¹² The specification that excludes students' prior-year scores avoids any potential concern that those prior-year scores are endogenous to the current principal's effectiveness if the principal was also leading the same school in the prior year. Identification of principal effects thus comes from within-school changes in test score *levels* associated with leadership transitions. Finally, the specification that excludes the first year of a principal's tenure as a school's leader from the

¹¹ To apply the Ehlert et al. (2013) approach, the first step for both the school and principal VAMs is to regress student outcome scores on all covariates and to obtain the residuals. The second step is to regress the residuals on school dummy variables (in the case of the school VAMs) or both school and principal dummy variables (in the case of the principal VAMs).

¹² We carry out a nonparametric version of this model by first calculating the percentile of each student's outcome score among all students in the same grade and year who earned the same scale score on the previous year's assessment in the same subject. We then regress these percentiles on school dummy variables (in the case of school VAMs) or both school and principal dummy variables (in the case of principal VAMs).

School Value Added and Principal Quality

principal VAM purges any immediate decline in either student achievement or the school environment that may result from principal transitions. To the extent that the transition year is marked by some degree of upheaval, excluding this year in the principal VAM could raise the correlation between value added estimates of principals and the schools they formerly led.

In Table 7, we report the findings using alternative specifications for math and reading. Across all of these specifications, school value added is never a statistically significant predictor of subsequent principal value added. In all but one estimate, the magnitudes of the relationships between school and principal value added are small, implying that no more than 12 percent of any given difference in true school value added serves as a signal of principals' persistent levels of effectiveness.¹³

Using Reverse Predictions

The next analysis uses an alternative sample of principal-grade observations that is constructed to capture a similar relationship between school value added and principals' persistent levels of effectiveness as in Table 6. The alternative sample comes from "reverse predictions" that use future school effectiveness estimates to predict principal effectiveness in previous years. That is, for each of the grades 4 to 8, we pair a school's effect in that grade from the 2012–2013 year with the principal's effect between 2007–2008 and 2011–2012 based on the same grade. Reverse predictions can be used in our analytic framework because measuring the relationship between school value added and principals' persistent levels of effectiveness should

¹³ The only exception is the analysis that excludes the first year after leadership transitions from the principal VAM. The math coefficient (0.51) is large but also noisy.

School Value Added and Principal Quality

not depend on the chronological order of when each estimate is obtained, provided they are still based on different time periods and student samples.

We show findings based on the reverse predictions in Table 8. The estimated relationships for both subjects continue to be small and statistically insignificant. Because this robustness analysis is designed to apply the same empirical method to a different set of observations, the findings—along with those from using alternative VAM specifications—reinforce our conclusion that school value added provides poor information about principals' persistent levels of effectiveness.

Principal Subgroups

In Table 9, we report findings from estimating Equation (6) separately for elementary school principals, middle school principals, and principals outside of Philadelphia and Pittsburgh. We exclude Philadelphia and Pittsburgh because officials in those districts presumably have the greatest opportunity to make compensatory assignments of more effective principals to less effective schools by virtue of managing the most schools in Pennsylvania. Since principal transitions brought about by compensatory assignments reduce the observed correlation between school effectiveness and principal quality, excluding these districts may increase the magnitude of the estimates. In all three analyses, however, we cannot reject the null hypothesis that school effectiveness estimates provide no indication of principals' persistent levels of effectiveness.

Accounting for the Gradual Manifestation of Principals' Effects

We next consider the possibility that the full effects of principals are not manifested immediately after assuming a school's leadership due to the lingering effects of the predecessor

School Value Added and Principal Quality

principal. For instance, although staffing decisions could be a key channel through which principals affect student outcomes (Branch et al. 2012), it may take time for a newly arriving principal to replace teachers who were hired by the previous principal. Likewise, a newly arriving principal may only gradually be able to change attitudes, expectations, and other aspects of a school's culture that were fostered by the predecessor.

School value added ought to be judged by how well it predicts principals' full effects, because those effects are the most complete representation of a principal's capabilities as a leader. However, it is possible that the principal value added estimates in our analysis may not yet encapsulate a principal's full effect. Our analysis is based on comparisons between departing principals and successors who started their positions within the last four years, so there may not have been sufficient time for successors to fully undo the influences of the predecessors. The concern, then, is that the variance in the principal value added estimates may be artificially compressed, leading to a downward biased coefficient on school value added when predicting principal value added.

Our solution is to inflate the estimated within-school differences in principal value added to the extent that those differences are likely to be artificially compressed. Adapting the framework used by Coelli and Green (2012), we defined the *leadership effect* at school s in year t , L_{st} , to be the combined effect of decisions and actions made by the current and prior principals on a school's effectiveness in the current year. The leadership effect is a weighted average of the

School Value Added and Principal Quality

previous year's leadership effect and the full underlying effect of the current principal (θ_p), with weight ρ on the previous year:¹⁴

$$L_{st} = \rho L_{s(t-1)} + (1 - \rho)\theta_p. \quad (10)$$

Our existing estimate for the within-school difference in value added between two principals is actually the difference in leadership effects between the period under a successor and the period under a predecessor. Consider a school in which we observe exactly one predecessor ($p=a$) and one successor ($p=b$) during the five school years ($t=1, \dots, 5$) of the principal VAM, with the successor starting in year t^* . Normalizing $L_{s1} = \theta_a$ and conducting repeated forward substitution of Equation (10), we can see that for every year $t \geq t^*$ the leadership effect is $L_{st} = \rho^{t-t^*+1}\theta_a + (1 - \rho^{t-t^*+1})\theta_b$. Therefore, the average difference in leadership effects between the successor's period ($t=t^*, \dots, 5$) and the predecessor's period ($t=1, \dots, t^*-1$), which is the quantity captured in our existing principal value added estimates, is

$$\frac{1}{5-t^*+1} \sum_{t=t^*}^5 L_{st} - \frac{1}{t^*-1} \sum_{t=1}^{t^*-1} L_{st} = \left(\frac{1}{5-t^*+1} \sum_{t=t^*}^5 (1 - \rho^{t-t^*+1}) \right) (\theta_b - \theta_a). \quad (11)$$

If $\rho > 0$ —that is, if the predecessor's influence lingers into the successor's tenure—then our estimated difference in the principals' value added is less than the full underlying difference in their effectiveness, $\theta_b - \theta_a$. However, Equation (11) makes clear how to recover the full underlying difference in effectiveness: multiply the estimated difference by the inflation factor

$$\left((1 / (5 - t^* + 1)) \sum_{t=t^*}^5 (1 - \rho^{t-t^*+1}) \right)^{-1}.$$

¹⁴ For ease of exposition, in this framework we abstract away from idiosyncratic components of principal quality, such as completely transitory or grade-specific components. Introducing those components would not yield extra insights from this framework.

School Value Added and Principal Quality

Our data do not afford a sufficiently long time series to estimate ρ with precision. For this sensitivity analysis, we borrow the estimate of ρ from Coelli and Green (2012), who find $\hat{\rho} \approx 0.7$ for achievement outcomes. We then inflate the within-school differences in the principal value added estimates by the inflation factor described above.¹⁵

The results, shown in Table 10, yield similar conclusions as our main findings.¹⁶ School value added has no significant relationship with the effectiveness that principals demonstrate in future years. Given the negative point estimates, we cannot have any confidence in the claim that a school whose value added exceeds another has the more effective principal.

Using Value-Added Estimates Based on Multiple Grades and Subjects

Thus far, our analyses have assessed the extent to which grade-specific school value added can predict the principal's subsequent value added in the same grade. As described earlier, requiring the school and principal value-added estimates that are paired with each other to be based on the same, single grade (from different years) effectively ensures that the same students do not contribute to both estimates and, hence, avoids spurious correlations between the estimates. However, in practice, the school value-added estimates that state and district officials see are likely to be those that reflect a school's effect averaged across all tested grades that it

¹⁵ We do not apply the inflation factor to the school value-added estimates. School value-added estimates do not, by definition, adjust for principals' length of service since those estimates are focused on the effectiveness of a school's entire staff. Moreover, given that school value added is the independent variable for predicting principal value added, inflating the variance in that independent variable would actually *decrease* the coefficient on that variable—the opposite of our intention.

¹⁶ For the analyses in Table 10, we narrow the sample to principals who served in a school with only one other principal (predecessor or successor) during the period of the principal VAM, and we narrow the definition of a network to be a single school only. This guarantees that every principal's value-added estimate is based solely on a comparison with one other principal who served at the same school. Due to this limitation in the sample and network type, we regard this approach as a sensitivity analysis only.

School Value Added and Principal Quality

serves. Therefore, as a more direct evaluation of school value-added estimates that are likely to be used in practice, we assess how well a school's average effect across all tested grades in 2007–2008 predicts the future value added—also averaged across all tested grades—of the school's principal.¹⁷ Despite having the advantage of using value-added estimates that are more aligned with practice, this analysis—unlike the main analysis—is vulnerable to bias from correlated sampling errors. For instance, an unusually bright group of fourth graders that inflates a school's value-added estimate in 2007–2008 could, upon reaching fifth grade the next year, inflate the same principal's subsequent value-added estimate, inducing a spurious correlation between the two estimates.

The findings, shown in Table 11, are nevertheless consistent with our main findings. The first two columns of results, which focus on math and reading separately, fail to reject the hypothesis that a school's value added is uninformative of the same principal's subsequent value added. The analysis in the final column uses composite value added estimates that are averaged across the two subjects, given that such composites are also used in practice.¹⁸ Again, school value added has no statistically significant relationship with a principal's subsequent value added. Across all of the analyses in Table 11, no more than 14 percent of any given difference in

¹⁷ We take the weighted average of a school or principal's unshrunk grade-specific value-added estimates, with weights equal to the effective number of students contributing to each value-added estimate. As in all of our analyses, the school value-added estimates are then shrunk with the usual empirical Bayes adjustment.

¹⁸ For each school and principal, we first take the simple average of the unshrunk math and reading value-added estimates. We calculate the sampling variance of the unshrunk composite estimate as one-fourth of the sum of three terms: the sampling variance of the math value-added estimate, the sampling variance of the reading value-added estimate, and twice the covariance of the math and reading value-added estimates (calculated as the covariance of reading and math test-score residuals from the school value-added model, divided by the effective number of students contributing to the school or principal's value-added estimate). Finally, the school value-added estimates are shrunk with the usual empirical Bayes adjustment.

School Value Added and Principal Quality

value added between two schools represents persistent differences in the effectiveness of the schools' current principals.

Concurrent Relationships between School and Principal Value-Added Estimates

In our final analysis, we show how a more naïve approach to evaluating the validity of school value added would have come to a qualitatively different conclusion than our main analysis. For this naïve approach, we estimate the association between school and principal value-added estimates that come from the same years and student samples. The results, shown in Table 12, indicate that 12 to 14 percent of the variance in principal value added can be explained by concurrent estimates of school value added (with implied correlation coefficients of 0.35 to 0.37), compared with R-squared values of no more than 1 percent under our main approach. In short, school and principal value-added estimates based on the same data have a moderate degree of consistency with each other.

Why does the naïve approach yield far larger relationships between school and principal value added than our main approach does? Two factors likely explain the difference. First, the sampling errors of the two sets of estimates are likely to be correlated in the naïve approach. Because the two estimates are based on the same students, whenever a school's value-added estimate is erroneously high or low due to fluctuations in the types of students enrolled, the principal's value-added estimate will also likely be erroneously high or low. In our data, error variance constitutes about 5 to 10 percent of the total variance in principal value-added estimates. If most of this error variance is also shared by the school value-added estimates, then this would account for much of the higher R-squared in the naïve approach. Second, under the naïve approach, the school and principal value-added estimates that are paired with each other

School Value Added and Principal Quality

reflect the same transitory principal effects (θ_{pt}) and (θ_{pgt}). Thus, the association between the two sets of estimates is inflated by principal effects that are not of key importance to evaluators.

5. CONCLUSION

A common method of using student achievement data to evaluate principals is to estimate school VAMs, which capture the contributions of entire schools to student achievement growth. However, school value-added estimates reflect a mix of school-level factors that might or might not be under principals' control. The usefulness of school VAMs as a tool for principal evaluations depends critically on whether differences in effectiveness across schools primarily reflect differences in principal quality or variation in factors over which principals have little discretion.

In this paper, we have examined the relationship between school effectiveness and the persistent level of effectiveness demonstrated by the school's principal. Using student-level longitudinal data from Pennsylvania, we measure school effectiveness and principal quality on the basis of school VAMs and principal VAMs, respectively. We identify principal quality from within-school changes in student achievement associated with leadership transitions, thereby purging the principal effectiveness estimates of all school-specific influences that are invariant over time. Moreover, we measure school effectiveness based on distinct years and students from those used to measure principal effectiveness. Therefore, any observed association between the school and principal value-added estimates must reflect effects on student achievement that principals persistently demonstrate.

We conclude that there is little evidence to support the contention that school effectiveness is a useful tool for assessing principals' persistent levels of effectiveness. Across multiple ways of specifying school and principal value-added models, school value added is never a statistically

School Value Added and Principal Quality

significant predictor of principal value added. Moreover, the magnitudes of the estimated relationships are small. In our main estimates, no more than 7 percent of any given difference in value added between two schools reflects persistent differences in the effectiveness of their current principals.

There are two potential reasons for the weak association between school effectiveness and principals' persistent levels of effectiveness. One possibility is that more-effective principals are assigned to schools with other factors that depress student learning growth, thereby masking the principals' contributions. We cannot confirm or rule out the presence of such compensatory assignments—despite our rough, initial attempt by excluding Pennsylvania's two largest districts in Table 9—given that school effectiveness measures cannot disentangle principals' contributions from unobserved, school-specific factors. Future research could probe more deeply into the ways in which principals are assigned to schools. Nevertheless, the presence or absence of compensatory assignments does not change our key conclusion that school value added likely provides a poor signal of principal quality under prevailing practices for how principals are assigned to schools. A second possibility, which we believe is more likely, is that school value added primarily reflects a combination of influences on student learning outside of the principals' control as well as effects that principals do not consistently demonstrate from one year to the next.

REFERENCES

- Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics* 25(1): 95–135.
- Branch, Gregory, Eric Hanushek, and Steven Rivkin. 2012. Estimating the effect of leaders on public sector productivity: The case of school principals. NBER Working Paper No. 17803.
- Betebenner, Damian. 2009. Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice* 28(4): 42-51.
- Buck, S. 2012. Estimating the effects of principal quality and experience. Unpublished paper, University of Arkansas.
- Cannon, Sarah, David Figlio, and Tim Sass. 2012. Principal quality and the persistence of school policies. Unpublished paper, Northwestern University.
- 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.
- Clark, Damon, Paco Martorell, and Jonah Rockoff. 2009. School principals and school performance. CALDER Working Paper No. 38, Urban Institute.
- Coelli, Michael, and David A. Green. 2012. Leadership effects: School principals and student outcomes. *Economics of Education Review* 31(1): 92–109.
- Deming, David J. 2014. Using school choice lotteries to test measures of school effectiveness. NBER Working Paper No. 19803.
- Deutsch, Jonah. 2012. Using school lotteries to evaluate the value-added model. Unpublished paper, Mathematica Policy Research.

School Value Added and Principal Quality

Dhuey, Elizabeth, and Justin Smith. forthcoming. How important are school principals in the production of student achievement? *Canadian Journal of Economics*.

Dhuey, Elizabeth, and Justin Smith. 2013. How school principals influence student learning. Unpublished paper, University of Toronto.

Ehlert, Mark, Cory Koedel, Eric Parsons, and Michael J. Podgursky. 2013. The sensitivity of value-added estimates to specification adjustments: Evidence from school- and teacher-level models in Missouri. *Statistics and Public Policy*, 1(1): 19-27.

Glazerman, Steven, Susanna Loeb, Dan Goldhaber, Douglas Staiger, Stephen Raudenbush, and Grover Whitehurst. 2010. *Evaluating teachers: The important role of value-added*. Washington, DC: Brown Center on Education Policy at Brookings.

Goldhaber, Dan, and Michael Hansen. 2010. Using performance on the job to inform teacher tenure decisions. *American Economic Review: Papers and Proceedings* 100(2): 250–255.

Grissom, Jason, Demetra Kalogrides, and Susanna Loeb. Forthcoming. Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*.

Grissom, Jason, and Susanna Loeb. 2011. Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal* 48(5): 1091–1123.

Hanushek, Eric, and Steven Rivkin. 2010. Generalizations about using value-added measures of teacher quality. *American Economic Review: AEA Papers and Proceedings* 100: 267–271.

Jacob, Brian, and Lars Lefgren. 2005. Principals as agents: Subjective performance measurement in education. NBER Working Paper No. 11463.

Jacob, Brian, and Lars Lefgren. 2008. Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics* 26(1): 101–136.

School Value Added and Principal Quality

Kane, Thomas J., Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger. 2013. *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Seattle: Bill & Melinda Gates Foundation.

Kane, Thomas J., and Douglas Staiger. 2008. Estimating teacher impacts on student achievement: An experimental evaluation. NBER Working Paper No. 14607.

Ladd, Helen F., and Randall P. Walsh. 2002. Implementing value-added measures of school effectiveness: Getting the incentives right. *Economics of Education Review* 21(1): 1–17.

Lipscomb, Stephen, Hanley Chiang, and Brian Gill. 2012. *Value-added estimates for phase 1 of the Pennsylvania teacher and principal evaluation pilot: Final report*. Cambridge, MA: Mathematica Policy Research.

Loeb, Susanna, Demetra Kalogrides, and Tara Bételle. 2012. Effective schools: Teacher hiring, assignment, development, and retention. *Education Finance and Policy* 7(3): 269–304.

McCaffrey, Daniel F., Tim R. Sass, J. R. Lockwood, and Kata Mihaly. 2009. The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4(4): 572–606.

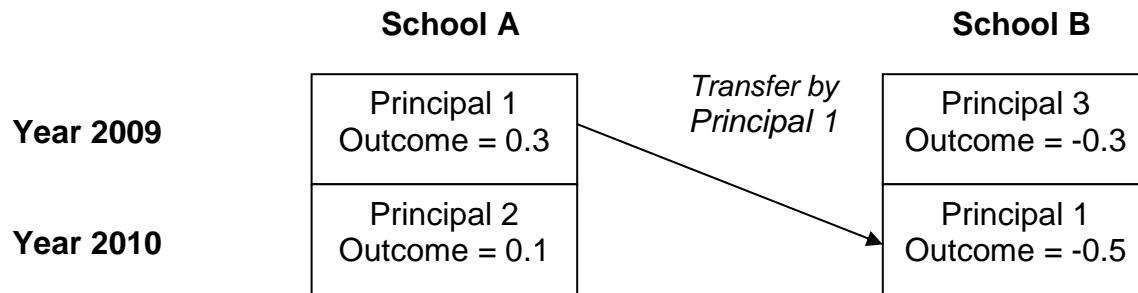
Purkey, Stewart C., and Marshall S. Smith. 1983. Effective schools: A review. *The Elementary School Journal* 83(4): 426–452.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. Teachers, schools, and academic achievement. *Econometrica*, 73 (2): 417–458.

Rothstein, Jesse. 2010. Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics* 125(1): 175–214.

School Value Added and Principal Quality

Figure 1. Hypothetical Connected Network



Note: Each cell of this figure represents a combination of year and school. Within each cell, the figure shows the identity of the principal in charge and the mean student outcome (in z-score units) after adjusting for all covariates (except the school and principal dummies) in the principal VAM.

School Value Added and Principal Quality

Table 1. Principal Transitions Across Schools Serving Students in Grades 4 to 8, 2007–2008 to 2012–2013

	Principals			Schools		
	Total	Involved in a Leadership Change (number)	Involved in a Leadership Change (percentage)	Total	Change in Leadership Occurred (number)	Change in Leadership Occurred (percentage)
Pairs of Adjacent Years						
2007–2008 and 2008–2009	2,426	433	18	2,274	233	10
2008–2009 and 2009–2010	2,401	448	19	2,253	245	11
2009–2010 and 2010–2011	2,419	496	21	2,241	275	12
2010–2011 and 2011–2012	2,407	417	17	2,282	233	10
2011–2012 and 2012–2013	2,405	458	19	2,265	254	11
Period for Estimating Principal Effectiveness, 2008–2009 to 2012–2013						
Period for Estimating Principal Effectiveness, 2008–2009 to 2012–2013	3,269	1,929	59	2,532	1,034	41
All Available Years, 2007–2008 to 2012–2013						
All Available Years, 2007–2008 to 2012–2013	3,565	2,378	67	2,626	1,492	48

School Value Added and Principal Quality

Table 2. Distribution of Network Sizes

Size Category: Number of Schools in the Grade-Specific Network	Number of Grade-Specific Networks with the Specified Size	Total Number of Principal-Grade Observations in Networks with the Specified Size	Average Number of Principals per Network
1	1,629	3,428	2.1
2	388	1,115	2.9
3	81	347	4.3
4	45	260	5.8
5	11	83	7.5
8	1	5	5
Any	2,155	5,238	2.4

Table 3. Sample Restrictions and Resulting Sample Sizes for the Analysis

Group	Number of Principal-Grade Observations	Number of Distinct Principals
1. Involved in a leadership change at some time from 2008–2009 to 2012–2013	5,238	1,929
2. Responsible for at least 20 students and in a network with another principal responsible for at least 20 students (and in Group 1)	5,059	1,881
3. Led a school that spans the same grade in 2007–2008 (and in Group 2)	2,001	802
4. At least one other principal in the same network meets the sample criteria (and in Group 3): final analysis sample	673	291
Distribution of Principal-Grade Observations by Network Size		
In networks with 2 principals in final sample	458	
In networks with 3 principals in final sample	156	
In networks with 4 principals in final sample	44	
In networks with 5 principals in final sample	15	

Table 4. Summary Statistics for Principal Characteristics

	Principals who Led a School Containing Any of the Grades from 4 to 8 in Any Year from 2008–2009 to 2012–2013		
	All Principals in PA	Principals Involved in Transitions	Principals in Final Analysis Sample
Master’s degree (proportion)	0.742 (0.438)	0.737 (0.440)	0.675 (0.469)
Doctorate degree (proportion)	0.104 (0.305)	0.092 (0.289)	0.107 (0.310)
Total years of experience in K–12 education (average)	20.1 (9.8)	18.9 (10.3)	22.3 (10.1)
White (proportion)	0.869 (0.337)	0.845 (0.362)	0.802 (0.399)
Female (proportion)	0.489 (0.500)	0.507 (0.500)	0.552 (0.498)
Number of principals	3,269	1,929	291

Note: Standard deviations are listed in parentheses below the mean for each variable in each sample.

Table 5. Summary Statistics for Student-Level Variables

	All Students in PA with Complete Data, 2007–2008 to 2012–2013	Students Contributing to Principal VAM Estimates, 2008–2009 to 2012–2013		Students Contributing to School VAM Estimates in Final Analysis Sample, 2007–2008
		Attending Schools Led by Any Principals with VAM Estimates	Attending Schools Led by Principals in Final Analysis Sample	
Current-Year PSSA Score (z-score units)				
Math	0.046 (0.979)	-0.028 (0.989)	-0.029 (1.008)	-0.064 (1.005)
Reading	0.042 (0.984)	-0.026 (1.004)	-0.020 (1.027)	-0.064 (1.021)
Prior-Year PSSA Score (z-score units)				
Math	0.042 (0.978)	-0.012 (0.993)	-0.014 (1.005)	-0.064 (1.014)
Reading	0.036 (0.984)	-0.025 (1.002)	-0.016 (1.022)	-0.061 (1.025)
Student Background (proportions)				
Female	0.490 (0.500)	0.490 (0.500)	0.490 (0.500)	0.491 (0.500)
Black	0.139 (0.346)	0.175 (0.380)	0.196 (0.397)	0.231 (0.421)
Hispanic	0.076 (0.265)	0.088 (0.283)	0.088 (0.284)	0.099 (0.298)
Eligible for free meals	0.322 (0.467)	0.374 (0.484)	0.368 (0.482)	0.361 (0.480)
Eligible for reduced-price meals	0.063 (0.242)	0.059 (0.235)	0.049 (0.216)	0.053 (0.223)
English language learner	0.020 (0.138)	0.024 (0.153)	0.027 (0.162)	0.028 (0.164)
Special education	0.157 (0.364)	0.161 (0.368)	0.162 (0.369)	0.163 (0.369)
Number of student-year observations	2,573,228	748,936	111,121	55,072

Note: Standard deviations are listed in parentheses below the mean for each variable in each sample. The summary statistics in the table include students in grades 4 to 8.

PSSA = Pennsylvania System of School Assessment; VAM = value-added model.

Table 6. Extent to Which School Value Added Predicts Principal Value Added in Subsequent Years: Main Findings

Independent Variables (in student z-score units)	Dependent Variable: Principal Value Added, 2008–2009 to 2012–2013 (in student z-score units)	
	Math (1)	Reading (2)
School Value Added, 2007–2008	0.07 (0.06)	-0.03 (0.07)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.01	0.00
Adjusted Standard Deviations Within Networks (expressed in student z-score units):		
Principal value added	0.14	0.11
School value added	0.16	0.12
Number of Principal-Grade Observations	673	673
Number of Distinct Principals	291	291

Note: Standard errors clustered by principal are listed in parentheses below each coefficient estimate. No estimates are statistically significant at the 5 percent level. All models control for network-by-grade fixed effects. An adjusted standard deviation of value added within networks is the square root of the difference between the total within-network variance of the value-added estimates and the average squared standard error of the value-added estimates.

School Value Added and Principal Quality

Table 7. Extent to Which School Value Added Predicts Principal Value Added in Subsequent Years: Alternative VAM Specifications

Independent Variables	Dependent Variable: Principal Value Added, 2008–2009 to 2012–2013	
	Math	Reading
Two-Step Value-Added Model for Principals and Schools		
	Student z-score units	Student z-score units
School Value Added, 2007–2008 (student z-score units)	0.06 (0.06)	-0.07 (0.09)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.00	0.01
Student Growth Percentiles Model for Principals and Schools		
	Student percentile points	Student percentile points
School Value Added, 2007–2008 (student percentile points)	0.06 (0.07)	-0.08 (0.08)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.00	0.01
No Prior-Year Scores in the Principal Value-Added Model		
	Student z-score units	Student z-score units
School Value Added, 2007–2008 (student z-score units)	0.122 (0.065)	0.12 (0.07)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.02	0.02
Drop First Year After Transition in the Principal Value-Added Model		
	Student z-score units	Student z-score units
School Value Added, 2007–2008 (student z-score units)	0.51 (0.27)	-0.04 (0.11)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.04	0.00

Note: Standard errors clustered by principal are listed in parentheses below each coefficient estimate. No estimates are statistically significant at the 5 percent level. All models control for network-by-grade fixed effects. All models except the final model are based on a sample with 673 principal-grade observations and 291 distinct principals; the final model is based on a sample with 589 principal-grade observations and 256 distinct principals.

School Value Added and Principal Quality

Table 8. Extent to Which School Value Added Predicts Principal Value Added in Prior Years

Independent Variables (in student z-score units)	Dependent Variable: Principal Value Added, 2007–2008 to 2011–2012 (in student z-score units)	
	Math	Reading
School Value Added, 2012–2013	0.07 (0.08)	0.14 (0.09)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.00	0.02
Number of Principal-Grade Observations	649	649
Number of Distinct Principals	288	288

Note: Standard errors clustered by principal are listed in parentheses below each coefficient estimate. No estimates are statistically significant at the 5 percent level. All models control for network-by-grade fixed effects.

School Value Added and Principal Quality

Table 9. Extent to Which School Value Added Predicts Principal Value Added in Subsequent Years: Principal Subgroups

Independent Variables (in student z-score units)	Dependent Variable: Principal Value Added, 2008–2009 to 2012–2013 (in student z-score units)	
	Math	Reading
Elementary-School Principals	Grades 4–5	Grades 4–5
School Value Added, 2007–2008	0.04 (0.05)	-0.07 (0.09)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.00	0.01
Number of Principal-Grade Observations	423	423
Number of Distinct Principals	241	241
Middle-School Principals	Grades 6–8	Grades 6–8
School Value Added, 2007–2008	0.13 (0.16)	0.05 (0.09)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.02	0.00
Number of Principal-Grade Observations	250	250
Number of Distinct Principals	150	150
Excluding Philadelphia and Pittsburgh	Grades 4–8	Grades 4–8
School Value Added, 2007–2008	0.05 (0.07)	0.00 (0.08)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.00	0.00
Number of Principal-Grade Observations	468	468
Number of Distinct Principals	223	223

Note: Standard errors clustered by principal are listed in parentheses below each coefficient estimate. No estimates are statistically significant at the 5 percent level. All models control for network-by-grade fixed effects.

School Value Added and Principal Quality

Table 10. Extent to Which School Value Added Predicts Principal Value Added Measures that Account for the Gradual Manifestation of Principals' Effects

Independent Variables (in student z-score units)	Dependent Variable: Principal Value Added, 2008–2009 to 2012–2013 (in student z-score units)	
	Math	Reading
School Value Added, 2007–2008	-0.10 (0.16)	-0.35 (0.22)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.00	0.03
Number of Principal-Grade Observations	306	306
Number of Distinct Principals	130	130

Note: In this analysis, principal value-added estimates are adjusted to correct for the lingering influence of a school's previous principal during the tenure of a subsequent principal. Standard errors clustered by principal are listed in parentheses below each coefficient estimate. No estimates are statistically significant at the 5 percent level. All models control for network-by-grade fixed effects.

School Value Added and Principal Quality

Table 11. Extent to Which School Value Added Predicts Principal Value Added in Subsequent Years, Using Value-Added Estimates Averaged Across Grades and Subjects

Independent Variables (in student z-score units)	Dependent Variable: Principal Value Added, 2008–2009 to 2012–2013 (in student z-score units)		
	Math	Reading	Both Subjects Combined
School Value Added, 2007–2008	0.12 (0.07)	0.10 (0.08)	0.14 (0.09)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.01	0.01	0.01
Number of Principals	291	291	291

Note: Robust standard errors are in parentheses. No estimates are statistically significant at the 5 percent level.

School Value Added and Principal Quality

Table 12. Extent to Which School Value Added Predicts Principal Value Added in the Same Years: Naïve Approach

Independent Variables (in student z-score units)	Dependent Variable: Principal Value Added, 2008–2009 to 2012–2013 (in student z-score units)	
	Math (1)	Reading (2)
School Value Added, 2008–2009 to 2012–2013	0.24* (0.01)	0.27* (0.01)
Fraction of the Variance in True Principal Value Added that Can Be Predicted by True School Value Added	0.12	0.14
Number of Principal-Grade Observations	5,059	5,059
Number of Distinct Principals	1,881	1,881

Note: Standard errors clustered by principal are listed in parentheses below each coefficient estimate.

* Statistically significant at the 5 percent level.

About the Series

Policymakers require timely, accurate, evidence-based research as soon as it's available. Further, statistical agencies need information about statistical techniques and survey practices that yield valid and reliable data. To meet these needs, Mathematica's working paper series offers policymakers and researchers access to our most current work.

For more information, contact Hanley Chiang, senior researcher, at hchiang@mathematica-mpr.com, Stephen Lipscomb, senior researcher, at slipscomb@mathematica-mpr.com, or Brian Gill, senior fellow, at bgill@mathematica-mpr.com.

www.mathematica-mpr.com

**Improving public well-being by conducting high-quality,
objective research and surveys**

PRINCETON, NJ - ANN ARBOR, MI - CAMBRIDGE, MA - CHICAGO, IL - OAKLAND, CA - WASHINGTON, DC

MATHEMATICA
Policy Research

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.