Educational Researcher
http://er.aera.net

Who Would Stay, Who Would Be Dismissed? An Empirical Consideration of Value-Added Teacher Retention Policies

Marcus A. Winters and Joshua M. Cowen
EDUCATIONAL RESEARCHER 2013 42: 330 originally published online 9 July 2013
DOI: 10.3102/0013189X13496145

Who Would Stay, Who Would Be Dismissed? An Empirical Consideration of Value-Added Teacher Retention Policies

Marcus A. Winters[1] and Joshua M. Cowen[2]

Several states have recently adopted or are pursuing policies that deny or revoke tenure from teachers who receive poor evaluation ratings over time based in part on quantitative measures of performance. Using data from the state of Florida, we estimate such value-added measures to consider the future effectiveness and number of teachers who would have been dismissed under different versions of these policies. Students assigned to teachers who would have been dismissed according to a value-added policy made considerably smaller academic improvements than did students assigned to teachers who would have avoided dismissal. Critically, however, we show that specific policy design determines the extent of the potential for value-added to improve the overall quality of the teaching workforce.

Keywords: accountability; economics of education; educational policy; educational reform; regression analyses

Introduction

Several states have recently adopted plans to deny tenure to, or revoke it from, teachers who receive poor evaluation ratings over time. Although the details differ, these plans generally allow or require administrators to remove teachers based in part upon a measure of their influence on student learning. Encouraging states to adopt policies intended to remove ineffective teachers was a major piece of the federal Race to the Top grant competition.

Supporters of such policies commonly believe that low-performing teachers are protected from termination by overly restrictive due process requirements. New dismissal policies may represent fundamental changes for the teaching profession, which has for several decades offered substantial job security. Such policies are designed to improve teacher quality and, ultimately, educational outcomes for students.

There are few who argue publicly that ineffective teachers should remain in the classroom. There is considerable disagreement, however, about how administrators should identify poorly performing teachers for dismissal. Strong supporters of teacher rights have also been concerned that teachers with unsatisfactory performance be given a chance to improve over time. What makes the latest teacher dismissal plans especially controversial is that they assess teachers in part based on quantitative measures of their performance, primarily through so-called value-added models of teacher effectiveness (VAMs). Although there are meaningful differences in the statistical strategies employed between them, value-added approaches generally attempt to isolate a common teacher-specific component to student test score outcomes. To put the point differently, value-added models generally predict individual student achievement based on a set of observable characteristics, and then assign any differences between actual and predicted test scores to the student's teacher in a given year.

Important issues continue to be raised over analysts' ability to estimate precise and unbiased measures of teacher value-added (see, e.g., Rothstein 2009, 2010, and McCaffrey, Sass, Lockwood, & Mihaly, 2009). There is particular concern that the estimates are so noisy—prone, in other words, to significant measurement error—that teacher dismissal policies based on value-added would arbitrarily remove many average or even effective teachers. Nonetheless, prior estimates showing that value-added measures contain information that can predict a teacher's future classroom performance suggest that the procedure might be used to improve upon the current system's ability to identify ineffective teachers for reward or dismissal (e.g., Goldhaber & Hansen, 2010; Jacob, 2011). More broadly, several studies have used value-added estimates to link teacher effectiveness to hiring and retention patterns (e.g., Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008; Boyd, Lankford, Loeb, Ronfeldt, & Wyckoff,

[1]University of Colorado Colorado Springs
[2]University of Kentucky, Lexington, KY

Educational Researcher, Vol. 42 No. 6, pp. 330–337
DOI: 10.3102/0013189X13496145

Table 1
Policies Linking Teacher Evaluation Results to Dismissal

Policy	General Dismissal Policy: Dismissal Directly Tied to Ineffective Teaching	General Dismissal Policy: Student Achievement Primary Determinant	Dismissal Criteria Include: Consecutive Years of Lowest Performance Rankings (# Years)	Dismissal Criteria Include: Nonconsecutive Lowest Rankings or Multiple Next-Lowest Rankings	Dismissal Criteria Include: Average Low Rankings or Other Multiyear Measure	Dismissal Criteria Include: Multiple Criteria	Dismissal Criteria Include: Unspecified
States	CO, CT, DE, DC, FL, HI, IL, IN, LA, MI, NJ, NV, NY, OH, OK, OR, PA, RI, TN	CO, DE, DC, FL, HI, LA, MI, NV, OH, OK, PA, RI, TN	DE, DC, FL, IN, NV, NJ, NY, OK, RI, TN (2); MI (3)	CO, DE, FL, HI, IL, IN, LA, OH	DE, IL, IN, NJ, OK, PA	DE, FL, HI, NJ, OK, TN	CT, OR
Number of states	19	13	11	8	6	6	2

Source: National Council on Teacher Quality (2011, 2012); Education Commission of the States (2013); New Jersey TEACHNJ Act of 2012.

2011; Cowen & Winters, 2013; Feng & Sass, 2011; Goldhaber, Gross, & Player, 2011; Goldhaber & Hansen, 2010; Hanushek, Kain, & Rivkin, 2004; Jacob, 2011; Krieg, 2006; West & Chingos, 2009). Despite the imprecision of the estimates, several simulations in the literature predict that improvements in overall quality may result from replacing ineffective teachers as measured by value-added with new applicants (Goldhaber & Hansen, 2010; Rothstein, 2012; Staiger & Rockoff, 2010; Winters & Cowen, 2013).

There remains, however, much to learn about the potential effects of such policies from a policymaking perspective. How many teachers would be removed by policies based on value-added? How would the teachers removed by such policies have performed relative to their peers in later years? What are the implications for policy design or the performance standard for the quality and number of teachers in danger of being dismissed? Although value-added evaluations represent but one area of debate over teacher quality and the teaching profession, addressing these questions is of considerable importance at this time when several states and local school districts have already adopted or are aggressively pursuing value-added to inform teacher retention decisions.

In this article, we consider these issues concretely using a rich administrative dataset from Florida. Looking back in time, we identify teachers who would have been removed in previous years under different politically feasible versions of value-added-based dismissal policies. We then evaluate the later effectiveness of teachers who would have been removed if particular dismissal policies had been in place. Our findings indicate that students assigned to teachers who 1 or 2 years earlier would have been dismissed according to a value-added-based policy made considerably smaller academic improvements than did students assigned to teachers who would have avoided dismissal under such a system.

Critically, however, we show that policy design determines the potential for value-added to improve the teacher quality distribution by removing teachers who will tend to be low-performing in the future. As Table 1 illustrates, nearly all states that have implemented these reforms focus on removing teachers that have received low effectiveness ratings in 2 consecutive years. We consider the quality of teachers removed under such a system and compare it to a policy design that is instead based on the teacher's average performance over a 2-year period. Although each of the policy types considered utilize 2 years of performance data, when they utilize the same percentile cutoff, their impact on the teacher quality distribution varies dramatically.

To have a meaningful impact on teacher quality throughout a school system, policies that remove teachers for below-standard performance in consecutive years—the modal style adopted thus far—must set a much higher performance standard than a policy that removes teachers whose performance over a 2-year period is below a given percentile. Consecutive-year policies that set relatively low percentile cutoffs for satisfactory performance will tend to remove very few teachers because even very bad teachers might score above the threshold in 1 of 2 years due to random fluctuation in the estimates of their effectiveness. Importantly, we demonstrate that when the two policy-types are designed to dismiss a similar number of teachers, the quality of teacher dismissed is very similar under each design. For such policies to remove similar numbers of teachers, however, the criteria for performance must be markedly different.

Our results make clear that value-added measures contain information that can help to identify teachers who will prove to be ineffective in later school years. In this, the article provides cautious optimism that a system in which teachers are removed in part based on these teacher evaluations may yield improvements in teacher quality over time. But our evidence also indicates that no system of evaluation will eliminate flaws from the measurement of teacher ability. The quality and number of teachers dismissed under value-added policies depends heavily on policy design. We assign no normative preference for a particular percentile cutoff or policy design in this article. Rather, we intend the results to inform policymakers and administrators about the implications of policy design when considering the design and implementation of such policies.

Data Sources

To begin our study, we constructed a set of data on districts, schools, teachers, and students over 5 years in the state of Florida. These data, obtained from Florida's K20 Data Warehouse, contain the test scores for each student who was administered the state's mandated reading test[1] from 2004–05 through 2008–09. The data include a unique student identifier used to track student performance over time as well as a unique identifier for the teacher that we use to measure teacher effects in the model.

We restrict our analysis to include only fourth and fifth grade teachers. It is easier to match a student to an individual teacher in these early grade levels. In addition, because testing in Florida begins in the third grade, we must further restrict our analysis to only fourth and fifth grade students in order to control for the student's prior test score.

Because even in elementary grades students are often matched to multiple teachers, we develop a protocol for identifying a single observation of a student matched to their teacher most responsible for their math or reading achievement in a given year. First, we only include teachers listed as the head of a self-contained classroom. If a student is still observed to be attached to multiple teachers, we then assign them to particular course numbers. Students are first matched to the teacher in the course listed as "fourth grade" or "fifth grade," and about 85% of students are matched to this teacher. Remaining students are matched to courses specific to elementary math or reading, depending on the analysis. The progression assigns the student (in order) to the teacher listed as language arts elementary, reading elementary, and finally language arts K–5. About 96% of students in our dataset are matched to a teacher according to these progressions, and remaining students are excluded from the analyses.

The analysis includes 15,152 fourth and fifth grade teachers observed in Florida public schools in 2006. The final estimation sample includes 227,014 fourth and fifth grade students enrolled in Florida public schools in 2009.

Methodological Approach

The first step of our procedure is to estimate a simple value-added measure of teacher effects on student reading test scores for each year. We utilize the student-level dataset to estimate a series of regressions taking the form:

$$y_{ijt} = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 X_{ijt} + \alpha_3 Z_{ijt} + \lambda_j + \varepsilon_{ijt}, \qquad (1)$$

where y_{ijt} is the test score of student i assigned to teacher j during school year t; X is vector of observed characteristics about the student, including the student's grade level; Z is a vector of demographic characteristics for students in the classroom that includes percentage male, percentage by race/ethnicity, percent of students with an individualized education program (IEP), percent with limited proficiency in English, and average incoming reading or math score; λ is a teacher fixed-effect; ε is stochastic term. We estimate Equation (1) separately for each grade level and year under consideration using observations from all students and their teachers for which data are available. We capture from Equation (1) the estimated teacher fixed effect. We then adjust the resulting fixed effect estimate according to the empirical Bayes estimator (e.g., Jacob & Lefgren, 2008). This adjusted fixed effect is the value-added measure of a teacher's effectiveness in a given year.

We collapse the resulting dataset to the teacher-year level. This allows us to identify those teachers who had estimated value-added scores below a given level in a particular year and thus would be dismissed under various forms of the policy. We next use the teacher's value-added score from each year to determine which teachers would have been removed under different versions of a dismissal policy. However, there are several different ways to design such a policy. In particular, school administrators choose the standard above which a teacher's value-added score must reach in order to avoid an "Unsatisfactory" or "Ineffective" rating (different states have different names for these lowest ranks) and also in what ways such a rating leads to dismissal—perhaps in negotiation with representatives from the teachers' union. In order to consider the influence of such decisions on program effects, we study two potential policy design types operating under three potential percentile cutoffs.

The percentile cutoff is based on the teacher's yearly effectiveness as measured by value-added relative to that of the distribution of teachers throughout the state. We consider policies that set the percentile cutoff, S, such that a teacher's value-added score (λ) is above the 5th, 10th, or 25th percentile of all teachers that school year in order to receive a satisfactory rating.

The first policy design that we consider identifies a teacher for dismissal if she receives unsatisfactory ratings in consecutive years. This policy design is the general application of those adopted in most states per Table 1. Here we define:

$$Dismiss^{consecutive} = \begin{Bmatrix} 1 & if\ \lambda_{jt} \le S_t \ \&\ \lambda_{jt-1} \le S_{t-1} \\ 0 & Otherwise \end{Bmatrix}. \qquad (2a)$$

We also evaluate the performance of an alternative policy that removes teachers whose average value-added score over the previous 2 years is below a given percentile for all teachers in the state during those 2 years. Only a few states have adopted this design as part of their dismissal plans (Table 1), but such a policy and those using similar multiyear measures remain the most visible alternative framework to consecutive-based policies currently in practice. In this alternative, we define:

$$Dismiss^{avg} = \begin{Bmatrix} 1 & if\ \dfrac{\lambda_{jt} + \lambda_{jt-1}}{2} \le S_t \\ 0 & Otherwise \end{Bmatrix}. \qquad (2b)$$

After calculating both 2a and 2b, we then merge the respective *Dismiss* variable back to the student-level dataset in order to estimate the effect of a student being assigned to a teacher who would have been dismissed by any particular policy under consideration. We estimate a regression that uses the indicator for whether a teacher would have been previously dismissed under a given policy and standard to predict later student achievement. Formally, we estimate models taking the form:

Table 2
Regression Results—Effect of Assignment to Would-Have-Been Dismissed Teacher

	5th Percentile	10th Percentile	25th Percentile
		Consecutive Years	
One year later	-0.188	-0.148	-0.105
	[0.041]	[0.025]	[0.010]
Two years later	-0.096	-0.099	-0.093
	[0.043]	[0.022]	[0.010]
		Two Year Average	
One year later	-0.109	-0.088	-0.060
	[0.014]	[0.009]	[0.006]
Two years later	-0.078	-0.066	-0.044
	[0.014]	[0.010]	[0.006]

Note. Each cell represents coefficient estimates from a different regression (12 regressions total). Columns indicate the percentile above which a teacher's value-added score must reach in a given year to avoid an Unsatisfactory rating. Consecutive Years refers to a policy that dismisses a teacher who receives two Unsatisfactory ratings in consecutive years; Two Year Average refers to a policy that dismisses a teacher whose value-added score over the previous 2-year period is below a given percentile for all teachers. Dependent variable in all models is the student's reading test score in 2008–09, expressed in standard deviation units. All models control for student grade level, previous year's reading score, gender, race/ethnicity, individualized education program (IEP) status, English language learner status, and an indicator for whether the teacher possesses a master's degree. Coefficient of interest is an indicator for whether the teacher would have been dismissed according to a particular policy design at the end of the 2006–07 (Two Years Later) or 2007–08 (One Year Later) school year.

$$y_{ijt} = \beta_0 + \beta_1 y_{ijt-1} + \beta_2 X_{ijt} + \beta_2 Z_{ijt} + \beta_4 Dismiss_{ijt-k} + \mu_{ijt}, \quad (3)$$

where $Dismiss_{ijt-k}$ is an indicator that equals 1 if the teacher would have been removed from the classroom *k* years earlier according to a given policy; μ is a stochastic term clustered by school; and β_0 through β_4 are parameters to be estimated.

Results

Table 2 reports the results from estimating Equation (3) for each of the policies under consideration. Each cell represents coefficient estimates resulting from a separate regression. The results labeled "One Year Later" report coefficient estimates for regressions evaluating the relationship between a student being assigned to a teacher in 2008–09 who would have been removed under the policy at the end of the 2007–08 school year; and the second set of results evaluate the relationship between being assigned to a teacher in 2008–09 who would have been removed according to the policy at the end of the 2006–07 school year.

The results show a statistically significant and substantial negative relationship between assignment to a teacher who would have been previously dismissed according to a given policy and student achievement. For instance, assignment in 2009 to a teacher who would have been removed at the end of 2008 according to a policy that dismisses a teacher with a value-added score at or below the fifth percentile for consecutive years is related to an average 0.188 standard deviation decrease in achievement relative to assignment to a teacher who would not have been dismissed according to such a policy. As the standard that a teacher must reach in order to earn a satisfactory rating increases, the average performance decline associated with assignment to a teacher who would be dismissed declines. That result is expected, since a policy with a higher standard will tend to remove teachers with higher average quality.

In each case the coefficient estimate for being assigned to a teacher who would have been removed due to consecutive below-standard performance is substantially more negative than the respective coefficient for being assigned to a teacher who would have been removed under a policy based on 2-year average performance. Thus, the results demonstrate that when considered under the same percentile cutoff, a policy based on consecutive below-standard performance removes teachers who on average prove to be less effective than a policy based on below-standard performance over a multiyear period.

Figures 1 and 2 illustrate the distribution of value-added scores of teachers in 2009 who would or would not have been dismissed had a policy been in place 1 or 2 years earlier. For illustration, we show the effect of policies of each design that set the percentile cutoff at the 10th percentile of performance in terms of value-added to earn a satisfactory rating. The figures clearly show that teachers who would have been removed from each version of the policy perform worse, on average, 2 years later than the total teacher effectiveness distribution. However, the overlap between each plot also makes clear that each version of the policy would remove some teachers whose later performance would prove to be above average. As the regression results indicate, the distribution of later effectiveness for teachers who would have been removed under a policy based on consecutive below-standard performance appears to be well below that of the distribution of teachers who would have been removed under a policy based on 2-year average performance if each policy sets the same percentile cutoff.

Quality and Number of Teachers Dismissed Under Different Policies

That a policy based on consecutive years of below-standard performance tends to remove worse teachers on average than a policy based on average performance when each sets the same percentile cutoff is an important finding. However, the effectiveness of a dismissal policy on student achievement depends both on the quality of teacher and the number of teachers who are removed from the classroom. A policy that removes only the single worst teacher in a state, for instance, would have a large effect for the few students who would have been assigned to him, but an imperceptible effect on teacher quality throughout the school system.

Table 3 maps the assignment of unsatisfactory performance ratings and dismissal of teachers over a 3-year period for each policy under consideration. The mapping begins with 15,152

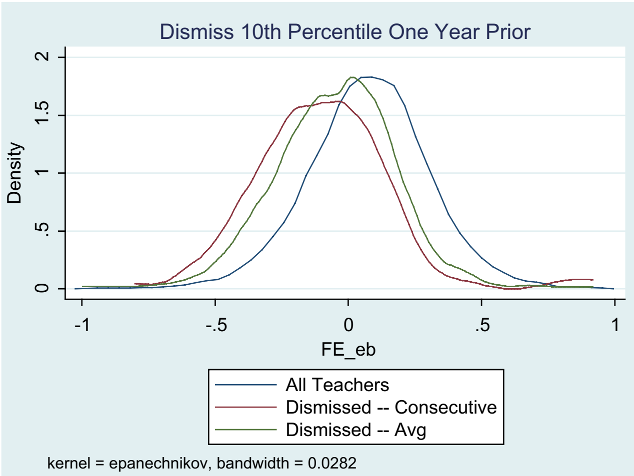

FIGURE 1. *Distribution of 2008–09 value-added scores for teachers dismissed under various policies at end of 2007–08*
For illustration purposes, the figures do not include the very few observations of teacher scores above 1 or below -1. All Teachers: Observations = 13,994, Mean = 0.062; Dismissed—consecutive: Observations = 88, Mean = -0.100; Dismissed—Avg: Observations = 593, Mean = -0.047.

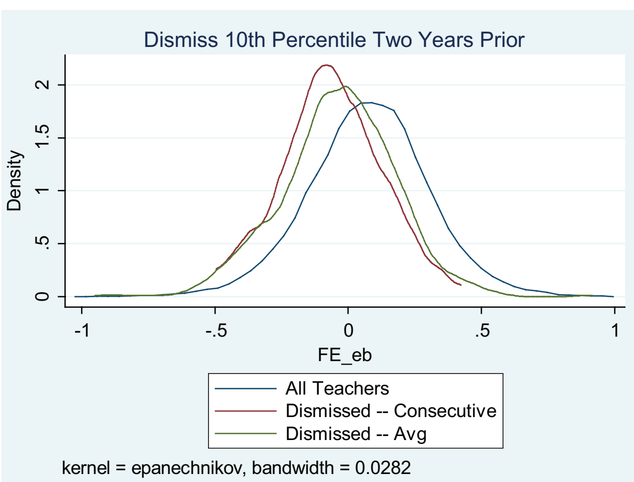

FIGURE 2. *Distribution of 2008–09 value-added scores for teachers dismissed under various policies at end of 2006–07*
For illustration purposes, the figures do not include the very few observations of teacher scores above 1 or below -1. All Teachers: Observations = 13,994, Mean = 0.062; Dismissed—Consecutive: Observations = 92, Mean = -0.086; Dismissed—Avg: Observations = 545, Mean = -0.048.

Table 3
Mapping Teacher Ratings and Dismissals

	Below Year 2006	Dismissed 2007	Dismissed and Observed 2008
		Consecutive Years	
5th percentile			
Teachers	754	50	31
Students			555
10th Percentile			
Teachers	1,509	155	88
Students			1,595
25th Percentile			
Teachers	3,774	844	523
Students			9,403
		Two Year Average	
5th percentile			
Teachers		493	292
Students			5,131
10th percentile			
Teachers		987	593
Students			10,557
25th percentile			
Teachers		2,468	1,639
Students			27,867
Teachers observed 2006	15,152		

Note. Table follows the hypothetical performance ratings for teachers based on different policy types and standards. Columns indicate the percentile above which a teacher's value-added score must reach in a given year to avoid an Unsatisfactory rating. Consecutive Years refers to a policy that dismisses a teacher who receives two Unsatisfactory ratings in consecutive years; Two Year Average refers to a policy that dismisses a teacher whose value-added score over the previous 2-year period is below a given percentile for all teachers.

fourth and fifth grade teachers observed in the dataset. We first consider the number of teachers removed under a policy based on consecutive below-standard performance. By definition, the percentage of teachers who receive a below-satisfactory rating according to such a policy about matches the percentile of the percentile cutoff needed to receive a satisfactory rating. However, a minority of teachers who receive an unsatisfactory rating under a given policy in one year will receive a below-satisfactory rating the next year and thus be dismissed. In addition, the third column of the table reports that only a fraction of teachers who would have been dismissed under such a policy were actually observed in Florida classrooms 2 years later. For instance, according to the table, only 50 of the 754 (6.6%) teachers with value-added scores below the 5th percentile also scored below that threshold in the following year and would have been identified for dismissal, and only about 62% of that number were actually observed in classrooms in 2008 absent the policy. Thus, had a policy based on a teacher scoring below the 5th percentile in consecutive years been in effect, only 31 of the original 15,152 fourth and fifth grade teachers throughout the state of Florida in 2006 would have been removed by such a policy—less than one teacher per two school districts in the state.

The second set of results maps the number of teachers who would have been dismissed under a policy based on 2-year average performance. As expected, a far larger number of teachers would have been removed under this policy than under a policy based on consecutive below-standard performance that set the

Table 4
Average Teacher VAM if Dismissed Teachers Replace With Average VAM Teachers

	Current	5th	10th	25th
Consecutive	0.037	0.040	0.043	0.057
Average		0.057	0.068	0.087
			Difference From Current	
Consecutive		0.003	0.006	0.020
Average		0.020	0.031	0.050

Note. The top panel of the table reports the average value-added score for teachers in 2009 under the current system and under policies that dismiss teachers under a particular policy design and percentile cutoff if all dismissed teachers were replaced by teachers with average value-added scores. The bottom set of results simply report the difference between average value-added score under the policy under consideration and under the current system. VAM = value-added models of teacher effectiveness.

Table 5
Characteristics of Policies Holding Fixed Percentage of Teachers to be Dismissed

Target Percentage of Teachers to Dismiss	5 Percent		10 Percent		25 Percent	
	Consecutive	Average	Consecutive	Average	Consecutive	Average
Percentile cutoff	17	5	27	10	47	25
Teacher dismissed	421	493	969	987	2,482	2,468
Teachers observed 1 year later	256	292	610	593	1671	1635
Average VAM 1 year later	-0.076	-0.050	-0.049	-0.047	-0.010	-0.018
Number dismissed under either policy	285		698		1,931	

Note. VAM = value-added models of teacher effectiveness.

same percentile cutoff. In fact, more than nine times the number of teachers would have been dismissed under an average-performance based policy with a standard set at the 5th percentile than would be removed under a policy with the same standard based on consecutive below-standard performance in consecutive years.

Table 4 puts the overall potential impact of the policies into context, taking into account both the quality and number of teachers removed, by calculating the difference in 2008 mean value-added score for the entire school system if we assume that dismissed teachers are replaced by teachers with average value-added scores. The table demonstrates that for any given percentile threshold the policy based on average effectiveness over 2 years would be expected to have a substantially larger impact on average teacher quality. This is because although teachers removed under the consecutive-year policy are of lower quality on average than those removed under the average-performance policy, too few teachers are removed under the consecutive-year policy to make much of a change in the overall teacher quality distribution.

We caution that the results reported in Table 4 are meant to illustrate the different potential effects of the policy types relative to one another, but they should not be interpreted as the anticipated effect of removing teachers based on any system considered on average teacher effectiveness. Simulating the overall effects of such policies on the teacher quality distribution requires taking several factors into account. Several recent articles have simulated the potential effects under such policies under a variety of assumptions (Goldhaber & Hansen, 2010; Rothstein, 2012; Staiger & Rockoff, 2010; Winters & Cowen, 2013).

Equalizing the Number of Dismissals Across Policies

Thus far, we have considered the relative impact of value-added-based dismissal policies with fixed percentile cutoffs but different definitions for an ineffective teacher—either a teacher who has not met the fixed percentile cutoff for consecutive years or a teacher whose average score over 2 years is below the fixed percentile cutoff. We have modeled the policy comparison in this way in order to mimic the way that policymakers appear to be considering the policy design issue.

The design of our policy comparison could create the impression that the difference in policy effects derives from differences in the precision of the estimates calculated using consecutive or average performance. That is not the case. In order to consider the relative precision of the estimates under the different policy designs, we must structure a comparison whereby equal numbers of teachers are dismissed. We can do so by identifying the percentile cutoff for the consecutive-year policy that would need to be set in order to dismiss the same number of teachers removed under a respective average-year policy.

Table 5 compares the required standards and the quality of teachers dismissed under each policy type with the goal of removing a given percentage of teachers. The table shows that a consecutive-year policy would need to considerably increase the percentile cutoff relative to an average policy in order to remove the same number of teachers. For instance, removing the bottom 10% of teachers, over a 2-year period would require setting the standard for satisfactory performance under the consecutive-year policy to about the 27th percentile compared to the 10th percentile for the average year policy. However, the table also demonstrates that the average quality of teacher removed under each system is similar when it is set to remove the same number of teachers.

Interestingly, the table shows that although the two types of policies remove similar quality teachers when they are designed to remove similar percentages of the workforce, many of the particular teachers who would be dismissed under one policy avoid dismissal under the other. From a policymaker's perspective, this is unimportant as long as the later effectiveness of dismissed teachers is expected to be similar. But from an individual teacher's perspective the difference can be considerable. Thus, each policy type can be structured to have similar effects. However, the necessary percentile cutoff to have a meaningful effect under each policy differs substantially. This is an important point for policymakers considering such policies to take to heart.

Discussion

Consistent with prior research, the results of our analysis in Florida suggest that even simple value-added assessments of a teacher's performance in one year contain information that can be used to predict student learning in that teacher's classroom in future years. In particular, we demonstrate that student performance was considerably lower in classrooms of teachers who would have been removed from the classroom had various versions of dismissal policies based on value-added been in operation.

However, our results further demonstrate that the number and quality of teacher removed under such dismissal plans depends heavily upon policy design. We have considered policies that dismiss teachers based on consecutively poor ratings as well as those based on average ratings over time. Although a small number of states (five, as of this writing) have incorporated combinations of the latter policies, the vast majority focus only on consecutive scores at the lowest performance level. This suggests that most states either consider this the only viable approach from a political or economic perspective or the only approach necessary from a policy perspective to achieve their aims. We have noted that given a state's particular dismissal goals, either of the consecutive or average-based approaches could be structured to have similar impacts on the size of its teaching workforce employed at a certain point in time. Critically, however, reconciling the two approaches in terms of the number of teachers dismissed comes at a cost of substantially changing the criteria for being ranked "ineffective" relative to the typical teacher in the workforce. Consecutive-based policies could remove a similar number of teachers as an average policy, but doing so requires setting the percentile cutoff substantially higher than is necessary under the average policy.

Most states that have adopted teacher dismissal policies in recent years have designed the system to remove teachers based on consecutive poor evaluation ratings based on value-added. Our results suggest that such policies will lead to relatively few dismissals of teachers who will later prove to be average or good relative to their peers. However, our results also suggest that unless the percentile cutoff is set quite high policymakers should limit their expectations for the effectiveness of such a policy on overall student achievement because it will tend to remove few teachers and many ineffective teachers will remain unidentified.

The descriptive empirical analysis provided in this article has some important limitations. First, by evaluating student achievement in prior years in the absence of a dismissal policy linked to teacher performance, we are not able to consider any effects that such a policy might have on changes in teacher or school over time. In particular, it is certainly possible that under such a policy teachers who receive a poor rating in one year would respond to the possibility of termination in either productive or unproductive ways. Second, a full consideration of the likely effects of such a policy would need to account for the sorts of teachers who would replace those dismissed by the system. We leave such an analysis for theoretical research, and in fact several recent articles have simulated such effects (Goldhaber & Hansen, 2010; Rothstein, 2012; Staiger & Rockoff, 2010; Winters & Cowen, 2013).

Finally, it is important to keep in mind the fact that our analysis has considered the teachers who would be removed under certain policies if teacher evaluations were based entirely on value-added assessment of their performance. However, all recent reforms along this line use value-added as only one part of a larger evaluation system. Furthermore, our analysis is only relevant to teachers whose value-added can be calculated, which excludes a large number of teachers who teach in untested grades or subjects. Thus, we argue that our analysis puts value-added to a particularly hard test in that we do not allow for qualitative, classroom-based evaluations that might identify cases where a teacher's value-added score does not correctly identify—or is unavailable as a measure of—his or her performance. The impact of the policies we consider here would be all-the-more understated relative to a distribution that includes such teachers.

Despite these limitations, our results provide cautious optimism that a system in which teachers are removed in part based on value-added teacher evaluations may yield improvements in teacher quality over time. But our evidence also indicates that no system of evaluation will eliminate flaws from the process by which administrators measure teacher ability. This is a point perhaps overlooked by the strongest supporters of value-added retention policies, but also overemphasized by the strongest critics.

In our data, the seemingly simple decision to employ consecutive years of outcomes rather than a multiyear average would have identified some of the same teachers for dismissal, but the two approaches would have ended the careers of other, different sets of teachers as well. This divergence stresses the importance of a transparent—and well-justified—set of criteria for implementing such a system of teacher evaluation. No system of evaluation in other fields—whether for doctors, airline pilots, law associates, or career civil servants—eliminates such subjectivity. Our results stress that in education, value-added approaches to teacher retention can provide some empirical evidence of teacher quality—perhaps more evidence than is currently available.

Interpreting this evidence, and acting on it, will remain the result of human decisions.

NOTE

[1]We focus on the reading test in our analysis. However, results are similar when the math test is utilized.

REFERENCES

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). Who leaves? Teacher attrition and student achievement. *NBER Working Paper no. 14022*.

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2011). The role of teacher quality in retention and hiring: Using applications to transfer to uncover preferences of teachers and schools. *Journal of Policy Analysis and Management*, *30*(1), 88–110.

Cowen, J. M., & Winters, M. A. (2013). Do charters retain teachers differently than traditional public schools? Evidence from elementary schools in Florida. *Education Finance and Policy*, *8*(1), 14–42.

Education Commission of the States. (2013). *State Policy Database*. Retrieved January 22, 2013, from www.ecs.org

Feng, L., & Sass, T. (2011). *Teacher quality and teacher mobility* (National CALDER Working Paper 57). Washington, D.C.: Urban Institute.

Goldhaber, D., Gross, B., & Player, D. (2011). Teacher career paths, teacher quality and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management*, *30*(1), 57–87.

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

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, *39*(2), 326–354.

Jacob, B. A. (2011). Do principals fire the worst teachers? *Educational Evaluation and Policy Analysis*, *33*(4), 403–434.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, *2*, 101–136.

Krieg, J. M. (2006). Teacher quality and attrition. *Economics of Education Review*, *25*(1), 13–27.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, *4*(4), 572–606.

National Council on Teacher Quality. (2011). *State of the states 2011: Trends and early lessons on teacher evaluation and effectiveness policies*. Retrieved January 22, 2013, from http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf

National Council on Teacher Quality. (2012). *State of the states 2012: Teacher effectiveness policies*. Retrieved January 22, 2013, from

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, *4*(4), 537–571.

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

Rothstein, J. (2012). *Teacher quality policy when supply matters* (Working Paper). Retrieved from http://gsppi.berkeley.edu/faculty/jrothstein/rothstein_teacherquality.pdf

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *The Journal of Economic Perspectives*, *24*(3), 97–117.

West, M. R., & Chingos, M. M. (2009). Teacher effectiveness, mobility and attrition in Florida. In M. G. Springer (Ed.), *Performance incentives: Their growing impact on American K-12 education*. Washington, DC: Brookings.

Winters, M. A., & Cowen, J. M. (2013). Would a value-added system of retention improve the distribution of teacher quality? A simulation of alternative policies. *Journal of Policy Analysis and Management*, *32*, 634–654.

AUTHORS

MARCUS A. WINTERS, PhD, is an assistant professor at the University of Colorado–Colorado Springs at 1420 Austin Bluffs Pkwy, Colorado Springs CO 80918; *mwinters@uccs.edu*. His research focuses on educational policy including teacher quality, school accountability, and school choice.

JOSHUA M. COWEN, PhD, is an assistant professor at the Martin School of Public Policy and Administration at the University of Kentucky; 433 Patterson Office Tower Lexington, KY 40506-0027; *joshuacowen@uky.edu*. His research focuses on areas of educational policy, including teacher quality, student mobility, school choice, and school accountability. He holds a doctorate from the University of Wisconsin–Madison.

Manuscript received January 10, 2013
Revisions received March 23, 2013, and May 29, 2013
Accepted June 6, 2013