

Will the Tide Lift All Boats? Examining the Equity Effects of Performance Funding Policies in U.S. Higher Education

Nathan Favero
Department of Public Administration & Policy
American University
favero@american.edu

Amanda Rutherford
School of Public and Environmental Affairs
Indiana University
aruther@indiana.edu

Abstract

This study considers whether performance funding policies systematically tend to harm some types of institutions of higher education while helping others. Building on theories of deck stacking and institutional stratification, a formal theoretical model of the effects of performance funding policies on individual institutions is developed and discussed. We find two types of likely policy effects—one which serves to improve overall institutional performance and another which exacerbates unevenness among institutions in terms of quality. We then conduct an initial empirical test of our theory, analyzing a cross-sectional time-series dataset of colleges and universities in the U.S. Our findings are somewhat mixed. The adoption of performance funding policies appears to have the ability to boost overall average levels of degree production in some instances. However, performance funding 2.0 policies are also associated with larger variance in degree production rates. We find some evidence that 2.0 policies also have heterogeneous effects on graduation and retention rates, whereby the benefits of these policies disproportionately accrue to institutions already positioned to perform well.

Key Words: performance funding, equity, performance, formal model

Will the Tide Lift All Boats? Examining the Equity Effects of Performance Funding Policies in U.S. Higher Education

As with many government institutions, accountability for efficient and effective performance is a central concern for public colleges and universities. Performance-funding policies for public institutions of postsecondary education have been discussed, if not implemented, in virtually every U.S. state. Such policies focus heavily on providing funding as determined by institutional performance metrics rather than the traditional input-based approaches such as total enrollments, credit hours, or similar measures.

The attraction towards using performance-based funding models is not unique to higher education but spans multiple policy areas. This approach stems from that notion that rather than dictating what specific structures or processes should be utilized to transform inputs to outputs, governments should measure performance outputs and hold individuals and their agencies accountable for what they produce. In theory, these policies allow for more freedom in processes which should lessen bureaucratic rules and red tape. Less burden in processes is expected to allow individual employees and institutions to have the flexibility to figure out what practices they think will work best in their particular setting (it is not always clear that procedural burdens are actually lessened; see Behn 2002). If what they try works, they will be rewarded; if not, they will have to try something else. Rewarding good performance also counters one of the most prevalent perceptions of government employees in the U.S.—the lazy bureaucrat. Those who are not intrinsically motivated to provide good service are expected to have a stronger incentive to do so as soon as there are rewards for good performance. In the context of higher education, it is often hoped (and expected) that performance funding will induce college and university employees to work harder and find better ways to serve students. Of course, the perception of

public agents as lazy and unmotivated is likely not to hold in many cases despite common rhetoric.

In spite of the popularity of performance funding policies among policymakers and the intuitive appeal of such programs, it is not clear that such policies are improving the performance of institutions. The first state performance funding policies in higher education were adopted decades ago, and existing studies of these policies have found little evidence to suggest that they are effective at improving performance, on average, within institutions (e.g., Hillman, Tandberg, and Gross 2014; Rutherford and Rabovsky 2014; Lahr et al. 2014; Hillman, Hicklin-Fryar, and Crespin-Trujillo 2017). However, what is often less clear in these studies is whether such policies produce unintended consequences in which some groups experience positive gains while others do not. Recent research has considered this in the context of within-institution changes for particular student groups. For example, Umbricht et al. (2015) find that institutions in Indiana became more selective and also enrolled fewer minority students. Kelchen and Stedrak (2016) determined that Pell Grant revenues per full-time equivalent student decreased under performance funding for four-year institutions, signaling that institutions may enroll fewer low-income students (see also Kelchen 2018 and Gándara and Rutherford 2017).

While these discussions have attracted growing attention to equity across student groups within institutions covered by performance funding, consequences related to equity *across* institutions has been virtually ignored in empirical research. We hypothesize it is likely that performance policies can reward those institutions already performing up to par while removing needed resources from low performing institutions. In this study, we consider the effects of performance funding policies not just on the overall performance of higher education institutions in a state but also on the differences among institutions within a state. Specifically, we wish to

investigate whether performance funding policies might have heterogeneous effects, positively affecting some institutions and negatively affecting others. Since performance funding policies fundamentally alter the way in which resources are allocated among colleges and universities, we suspect that institutions that receive relatively fewer resources under this distribution may suffer.

We consider this possibility in the following sections by reviewing existing theoretical perspectives that might elucidate policy making in the realm of higher education, developing a formal model and conducting Monte Carlo simulations, and analyzing an empirical test using a panel of U.S. higher education data. The Monte Carlo simulations demonstrate the possibility that average performance increases but also the possibility that the standard deviation of performance increases, which suggests that the net effect on low performing institutions depends on which force is stronger. Our empirical findings are somewhat mixed, but evidence in line with our expectations for heterogeneous effects on institutions is stronger for performance 2.0 policies—those that determine part of the base funding of an institution—as compared to performance 1.0 policies—those that generally provide a financial bonus to institutions. More specifically, we find that performance 2.0 policies widen the variance observed in the standard deviation of degree production but actually lower the standard deviation for graduation rates. Further, marginal effect plots by selectivity show little difference among institutions for degree production but more variation for graduation and retention rates in a direction that would benefit those institutions already performing well while potentially hurting less selective colleges and universities. Historically Black Colleges and Universities (HBCUs) are also more likely to see decreasing graduation rates compared to non-HBCUs in our data under 2.0 policies. These findings and additional implications are considered after presenting the data and analyses below.

The Development of Performance Funding Policies in Postsecondary Education

Performance-based funding models now occupy center stage in discussions among practitioners and scholars regarding the mechanisms by which to hold public institutions of higher education accountable. States have taken a number of approaches in the development and refinement of such policies—hence the high level of variance seen among policies—and research continues to evolve in an effort to determine the effects of these policies on various indicators of student access and performance.

Proponents of performance-based funding generally argue that institutions must move beyond the traditional focus on enrolling students (and receiving state appropriations according to such enrollments). Groups like the Lumina Foundation and Complete College America stress that institutions must be held accountable for the success of students post-enrollment, especially given that state allocations of funding are intended to use taxpayer dollars efficiently to produce expected outcomes (Complete College America 2016, Lumina Foundation 2016). In other words, because dropouts are perceived to cost taxpayers money and signal waste and inefficiency in the public bureaucracy that is higher education, states should allocate such dollars based on outputs and performance metrics. By doing so, institutions will be incentivized to focus on student outputs and will therefore develop strategies by which to improve institutional quality and efficiency (Alexander 2000). This argument resonates well with multiple stakeholder groups and suggests that policymakers are taking action. In 2015, the National Conference of State Legislatures reported that thirty-two states had some type of funding formula or policy and that an additional five states were in the process of transitioning to some type of policy (National Conference of State Legislatures 2015). That seventy-five percent of states had active performance policies in 2015 and discussions of these policies has only expanded indicates

widespread support among some groups for challenging the incentive structures at work in public postsecondary institutions.

Opponents of performance funding argue that these policies are likely to be ineffective for a number of reasons. First, many question whether institutions will be provided the leverage needed to react to these policies in developing promising practices by which to improve performance across the entire college or university (Rutherford and Rabovsky 2014). Second, much debate has occurred in trying to achieve consensus on how performance should be defined and operationalized as well as whether such indicators provide objective assessments of performance (Tam 2001, Moynihan 2008). Third, as policies are revised or removed and reenacted, it is possible that such instability will limit opportunity for significant long-term change (Dougherty, Natow, and Vega 2012; Dougherty and Natow 2014; Bell, Hicklin-Fryar, and Hillman 2018). This idea is particularly important given the slow-moving nature of large bureaucracies and the stickiness of performance data (Hicklin and Meier 2008, Rutherford and Meier 2015). For many institutions, the largest predictor of the graduation rate in year t is the graduation rate in year $t-1$; the same can be said for many other performance measures. Fourth, whether the share of budget allocations constitutes a meaningful amount that generates reaction and behavioral change can be consequential. For instance, should performance funding only be used for a small share of appropriations, then institutions may not invest the resources to react to such policies (Rutherford and Rabovsky 2014).

Scholarly research, using multiple data samples as well as various methodological approaches, on the effect of performance funding policies have largely supported dismal conclusions about whether performance funding is effective. Such policies are not significantly related to degrees awarded or graduation rates (e.g., Dougherty and Reddy 2013; Hillman,

Tandberg, and Gross 2014) and appear to have only limited linkages to the granting of certificates (Hillman, Tandberg, and Hicklin-Fryar 2015; Hillman, Hicklin-Fryar, and Crespin-Trujillo 2018), which are generally worth less than other types of degrees (Dadgar & Trimble, 2015). Perhaps partially in response to the first wave of these studies, advocates of performance funding have pushed states to raise the stakes of external accountability. This has taken form in the increasing shares of institutional funding determined by performance criteria. For example, states like Ohio and Tennessee now determine nearly 100 percent of allocations through performance funding models (Dougherty et al. 2014). Additionally, states have refined models to include a number of potential performance indicators. While most originally focused on degree production or graduation and retention rates (Dougherty and Reddy 2011), some states now include premiums for the enrollment and completion of in-state students, low socioeconomic status students, or students of color (Gándara and Rutherford 2017, Kelchen 2018). These additional complexities create interesting questions when assessing effectiveness or unintended consequences either for certain groups of students or certain types of institutions.

Potential Disparate Effects of Performance Funding

While scholars continue to empirically test the effectiveness of performance funding policies on student outcomes, important angles of these policies that have the potential to produce perverse incentives and promote institutional gaming are largely understudied. Though performance-based accountability mechanisms intend provide incentives for colleges and universities to increase performance, institutions may find shortcuts allowing them to produce better numbers on salient metrics without actually improving the quality of their educational services. Focusing on the case of Indiana, Umbricht, Fernandez, and Ortagus (2015) investigate perhaps the most commonly proposed way in which institutions might game performance

policies—admissions processes—and find that institutions do appear to become more selective following the implementation of a performance funding policy (see also the work by Dougherty et al. 2016 on colleges and universities in Indiana, Ohio, and Tennessee). If institutions need to increase retention, graduation, and degree production but are cognizant that much of student success depends on student- and family-level factors, they may simply select those students who are most likely to succeed on their own. This practice (known as creaming) may have consequences for students from less affluent or otherwise disadvantaged backgrounds, threatening goals of accessibility in public institutions. A recent study by Kelchen and Stedrak (2016) supports similar issues with selectivity; while the authors do not look at admissions directly, they do find that colleges subject to performance funding receive less Pell Grant revenue, suggesting some trade-offs in the composition of student groups recruited by the institutions (in other words, these colleges may recruit higher shares of high-income students and focus less on low-income students).

Even though these studies have helped to increase our understanding of the gaming that can occur in the implementation of performance-based funding, there is still very little research that considers equity *across* institutions rather than *within* institutions. Trade-offs considered under both lenses merit scholarly attention, but, to date, research has focused on the latter (though see early work by Fryar 2011). Only in the last few years have scholars considered disparate effects across institutions. For example, Li, Gándara, and Assalone (2018) find that MSIs in Texas and Washington receive the same or less per-student state funding following performance funding that their non-MSI counterparts (see also similar findings in Hillman and Corral 2017), and Hagood (2017) shows that high-resource institutions are more likely to benefit

from performance funding that lower resource institutions using data on public four-year institutions from all U.S. states.

It is also possible that the null affects found for student outcomes are somewhat a function of a zeroing out effect that occurs in analyses when some institutions gain and others lose. We argue here that some institutions—generally those with higher performance and more resources—are likely to experience gains while other institutions already struggling with poor performance suffer in states with performance funding models. This is likely to be the case even where funding models have become more complex to include additional categories of performance criteria (Hagood 2017); while additional options provide multiple avenues to meet performance standards, it is unclear whether the avenues receive equal weight or adequately account for contextual differences across institutions. Further, there may be additional opportunities to participate in gaming on one or more of the performance categories.

One example of the potentially disparate effects of accountability policies occurs in K-12 education (e.g., Abernathy 2007). Diamond and Spillane (2004), for example, illustrate that schools respond to accountability policies such that higher performing schools are more likely to respond by bolstering performance for all students as the policy intends while poor performing schools focus more narrowly on components of compliance. The authors argue that disparate effects are explained by the fact that a large portion of student educational outcomes are influenced by factors outside of the school such that institutional variance is a mirror of larger issues of disadvantage and segregation in society. Darling-Hammond (1994, 15) summarizes the issue of disparate impacts across institutions in writing:

Applying sanctions to schools with lower test score performance penalizes already disadvantaged students twice over: having given them inadequate schools to begin with,

society will now punish them again for failing to perform as well as other students attending schools with greater resources and more capable teachers. This kind of reward system confuses the quality of education offered by schools with the needs of the students they enroll; it works against equity and integration... by discouraging good schools from opening their doors to educationally needy students.

In other words, low performing schools are punished without an attempt to correct the resource disparity that already exists and that may be one of the causes of initial poor performance.

Deck Stacking and Institutional Stratification

In considering the argument of whether some institutions gain and others lose in performance-based funding systems, at least two theoretical foundations can help to lay out expectations and hypotheses. First, in considering the need for elected policymakers to retain some control over bureaucratic agencies, McCubbins, Noll, and Weingast (1987) argue that stacking the deck, or the use of processes to steer agency decisions and policy actions in a certain direction, favor some stakeholders over others. Of course, some political science scholars have claimed that early discussions of deck stacking were based on positivism in which problems are modeled away rather than empirically tested (Spence 1997, Balla 1998). Yet, even if policymakers are not intentionally attempting to stack the deck, it is likely that their experiences, values, and biases shape the way in which they form policies, and this may lead to the favoring of some groups over others (see, for example, the discussion of upper echelons theory by Hambrick and Mason 1984). It is also likely that policymakers will focus on constructing policies in a way that fits the needs of the organizations that are most salient to either actors in the policymaking process or the general public. By tailoring policies to fit these organizations, performance and accountability criteria may be less well suited to less prominent organizations.

This could certainly be the case in higher education, where four-year public institutions, particularly flagship or land grant universities, are often the most salient to a state's population and policymakers; those holding office are often graduates of these state institutions (Smallwood and Richards 2011). Thus, when crafting performance-based funding mechanisms, what performance information is collected and how it is accounted for in appropriations may be designed with these high capacity, well-resourced institutions in mind. This means that these institutions may be most likely to benefit from responding to policies and further increasing resources. On the other hand, other colleges in the state, including but not limited to regional schools and minority-serving institutions, may be less well situated to meet performance criteria as they have been interpreted. Instead, these institutions could struggle to meet performance standards, which will result in resources becoming even more constrained when appropriations are redistributed away from these colleges (Hicklin 2011). If the deck is already stacked against low performing institutions, it may not be reasonable to expect performance to improve given fewer resources.

A second theory that can help to explain why performance funding policies would be expected to lead to disparities across colleges and universities is institutional stratification. This is perhaps best outlined by Roscigno (2000) in explaining the linkages between family and school disadvantages and the argument that where one resides has a large effect on the services one uses. Institutions serving low-income students, for example, may face different performance challenges than those institutions that attract high-income students, who are also more likely to have resources to help with retention and graduation aside from any efforts made by the institution. Indeed, students of the highest socioeconomic levels may be most mobile, and they are also likely to self-select into institutions already viewed as high performers (Carnevale and

Rose 2003, Paulsen and St. John 2002). Those public institutions struggling to meet institutional or state-defined performance criteria have more limited resources and a more limited pool of students, particularly when located in education deserts. This again may stack the deck against these institutions and lead to an environment in which performance funding only makes resources more scarce and performance more difficult to achieve. Essentially, group disadvantages that exist outside of these organizational boundaries are likely to be duplicated, and perhaps aggregated, when institutions serve different types of students (Posselt et al. 2012, Cantwell and Taylor 2013).

This context also emphasizes that institutions of postsecondary institutions are indeed slow-moving bureaucracies (Hicklin and Meier 2008, Rutherford and Meier 2015). And while institutions situated in areas of educational abundance or those institutions with a high level of resources may have smaller room for improvement (going from 89 percent retention to 90 percent retention is arguably more difficulty than moving from 50 percent retention to 51 percent retention), they are better equipped to maintain and strengthen performance given the rewards generated by existing performance funding models.

Developing a Formal Model to Predict the Effects of Performance Funding on Institutions

If state performance funding policies alter the allocation of state resources among institutions, then such redistribution is likely have consequences for institutional performance. We argue that those institutions that have a starting point of high performance (as measured for state funding formulas) will experience a different trajectory than those institutions that start from poor performance such that the latter group will be harmed. We explore whether this claim has any validity in two stages—first through a formal theoretical model and then through empirical analysis on public four-year institutions of higher education between 1993 and 2013.

Model Assumptions

In order to formally model the effects of performance funding policies on institutions' performance, we must first make some assumptions about the determinants of organizational performance. To construct our model, we make five key assumptions:

1. Overall performance can be meaningfully represented with a single dimension. For the sake of simplicity, we consider performance along a single dimension, although in reality the performance of public organizations is almost certainly a multidimensional concept. We would argue that as long as one is not concerned with modeling how particular variables (e.g., institutional behaviors) affect various performance dimensions differently, the assumption of a single dimension is not particularly limiting. After all, multiple dimensions of performance can be combined into a single performance measure by, for example, averaging several measures of performance that each capture distinct dimensions (given the measures exist along comparable scales). To the extent that real-world performance funding policies do not accurately measure and reward all aspects of performance, the effectiveness of policies will likely be hindered. Because our formal model does not consider the possibility that performance dimensions are inaccurately accounted for, we are essentially modeling a best-case scenario for performance-based funding policies with regards to the complications inherent in measuring multidimensional performance.
2. Financial resources positively affect performance. We assume that when a state provides a greater amount of financial resources to an institution, the institution's performance increases. For example, Deming and Walter (2017) find that spending has large effects on both enrollment and degree completion in public institutions;

importantly this study identifies that changes in state funding have important consequences for core academic spending, which includes tutoring, advising, and mentoring (see also Jackson et al. 2016 and Lafortune et al. 2018). Others, such as Long (2016), note that schools with higher outcomes such as retention receive more state funding, meaning that funding often goes disproportionately to those institutions where students are already prepared to succeed. While it is unclear in this work whether higher outcomes drive larger appropriations or larger appropriations yield higher outcomes, states do tend to favor flagships and four-year institutions which can exacerbate inequities (Long 2016). Resources in our model are best conceptualized as spending per student. Because we realize the resource-performance link may not always hold in practice, we also try relaxing our resources assumption and simulate results for a version of our model in which resources have no effect on performance. Furthermore, we acknowledge that resources are not the only determinant of institutional performance; in the online appendix, we present the results of alternative simulations that explicitly account for other time-invariant institutional factors (e.g., historical prestige) that influence performance levels.

3. The behavior of institutional actors (administrators, faculty, and staff) affects performance. One of the key parameters we will vary in the model represents the extent to which institutional actors react positively to the incentives created by a performance funding policy. In other words, we will consider how the effects of performance funding policies differ depending on the degree to which we assume that institutional actors adopt behaviors that improve performance when a performance funding policy is in place. This assumption does not imply that actors will (or will

not) respond positively to performance funding policies. Instead, we will model a range of responses to performance funding policies, including responses which have no effect on performance and responses which have a strong, positive effect on performance.

4. Performance funding policies do not increase total state support for higher education but allocate a greater share of state resources towards those institutions that have performed well. We wish to compare the effects of (1) a (performance funding) policy that allocates X state dollars among various institutions with the institutional amounts based partially on institutional performance in the prior year versus (2) a policy that allocates X state dollars among all institutions equally without regard to institutional performance. In other words, we assume the state spends the same number of dollars per student on higher education in the state regardless of whether or not a performance funding policy has been adopted. We believe this is the ideal tradeoff to consider since it speaks directly to the question of whether performance-based funding policies allow states to stretch their dollars further when it comes to funding higher education institutions. As a practical matter, if all institutions exceed performance expectations, they are likely to all experience increases in resources in the short-term under real-world performance funding policies. However, we do not expect that legislatures' appetite for funding higher education in the long-term will be affected by the performance of higher education institutions. Here, we simply assume that total state support for higher education is constant (regardless of performance levels) but that the amount allocated to each institution varies depending on

performance. We also note that this assumption indicates that state resources are tied to absolute performance levels rather than year-to-year changes in performance.

5. Performance is measured perfectly. We assume that performance can be measured perfectly by the state, so adjustments to resource allocation can be based on actual performance. As such, it is impossible for institutions to engage in “gaming” behaviors that improve their performance metrics without actually improving the outcomes that are desired. The possibility of gaming behaviors certainly occurs in organizations and merits scholarly study, but in this study we wish to focus on other factors affecting the consequences of performance funding policies in our model.

When gaming behaviors do occur, we would expect the outcomes to be less favorable than those indicated by our model.

Formally, we assume that performance is determined by the following equation:

$$p_{i,t} = \beta_0 + \beta_1 r_{i,t} + \beta_2 b_{i,t} + x_{i,t}, \quad (1)$$

where $p_{i,t}$ is the performance of institution i at time t , $r_{i,t}$ is the level of institutional resources provided by the state, $b_{i,t}$ represents the level of positive behavior by institutional actors, $x_{i,t}$ is a term representing the effect of all other factors affecting the organization’s performance, and β_0 , β_1 , and β_2 are population parameters.

We assume that resources are allocated according the following formula:

$$r_{i,t} = \gamma + \alpha(p_{i,t-1} - \bar{p}_{t-1}) \quad (2)$$

where \bar{p}_t is the average performance among all institutions at time t , α is a parameter describing the strength of the performance funding policy (in other words, dosage, or the extent to which the resource allocation depends on past relative performance), and γ is a population parameter

indicating the average level of funding among all institutions.¹ When there is no performance funding policy, $\alpha = 0$. When $\alpha > 0$, the level of financial support an institution receives is positively related to its relative performance in the prior year ($p_{i,t-1} - \bar{p}_{t-1}$). By *relative* performance, we mean an institution's performance relative to the average level of performance among all institutions in a state in a given year (positive values indicate better-than-average performance while negative values indicate worse-than-average performance).² Varying allocations based on relative performance holds the total amount of state financial support constant (consistent with our fourth key assumption as explained above).

The following equation represents the effect of performance funding policies on the behavior of institutional actors:

$$b_{i,t} = \delta_{i,t} + \varepsilon\alpha \quad (3)$$

where ε is a population parameter representing the responsiveness of institutional actors to the incentives created by the performance funding policy (α) and $\delta_{i,t}$ indicates institutional behavior in the absence of performance funding. When $\varepsilon = 0$, institutional actors do not respond to performance funding policies in any way that affects performance. When $\varepsilon > 0$, institutional actors respond positively to the existence of performance funding policies such that performance improves.

Substituting equations (2) and (3) into equation (1) and rearranging some of the terms produces:

¹ Equation (2) implies that all institutions receive an equal amount of resources (funding per student) in the absence of a performance funding policy. Our model can still accommodate variation in funding levels (due to reasons other than performance funding) if we consider such variation to be a part of the $x_{i,t}$ term in equation (1) rather than a part of the resources term ($r_{i,t}$).

² Some smaller states have very few institutions of higher education, which may lessen concerns that performance funding policies will exacerbate inequality among institutions. Instead, the main sources of inequality will probably be the K-12 education system or uneven distribution of resources within a university or college.

$$p_{i,t} = \beta_1 \alpha (p_{i,t-1} - \bar{p}_{t-1}) + \beta_2 \varepsilon \alpha + x_{i,t} + \beta_0 + \beta_1 \gamma + \beta_2 \delta_{i,t}. \quad (4)$$

If we define $\theta = \beta_1 \alpha$, $p^*_{i,t-1} = p_{i,t-1} - \bar{p}_{t-1}$, $\tau = \beta_2 \varepsilon \alpha$, and $u_{i,t} = x_{i,t} + \beta_0 + \beta_1 \gamma + \beta_2 \delta_{i,t}$, we are left with:

$$p_{i,t} = \theta p^*_{i,t-1} + \tau + u_{i,t}. \quad (5)$$

θ represents the extent to which last year's performance affects this year's performance by altering the share of state resources received this year. τ indicates whether (and how much) institutions positively alter their behavior in response to the incentives produced by a given funding policy.³ In other words, higher values of τ indicate that university administrators are more positively responding to the performance funding scheme, perhaps pursuing new initiatives or programs that improve student outcomes because of a desire to acquire greater resources by scoring well on performance metrics. When there is no performance-based funding policy, $\theta = \tau = 0$. $u_{i,t}$ represents the effect of all factors unrelated to the funding policy on performance.

Monte Carlo Simulations

We conducted Monte Carlo computer simulations to determine how various values of the key parameters in equation (5) affect the distribution of performance outcomes for institutions. We assume there are 50 institutions of higher education in the state. Most U.S. states have far fewer than 50 four-year colleges or universities. We simulate a large state because it is easier to

³ Because equation (3) assumes that the size of the effect of a performance funding policy on institutional behavior depends (linearly) on the strength of the performance-based component of the allocation formula (or the degree to which funding levels vary based on past performance, as expressed in α), increasing the intensity of the performance funding policy (α) will increase both θ and τ . We do not emphasize this aspect of the model, however, and could have easily obtained equation (5) under the assumption that τ is fixed for all performance funding policies regardless of the magnitude of α ; under such an alternative derivation, τ would be interpreted as a fixed response to the presence of any performance policy. Under the formulation presented in the manuscript, τ represents the response to a one-unit increase in the intensity of a performance funding policy multiplied by the actual intensity of the performance funding policy that is in effect.

describe the tails of a distribution with a larger sample, but we have also obtained similar results for simulations that uses only 11 institutions (see online appendix). For our initial set of simulations, we assume that each value of $u_{i,t}$ is randomly and independently drawn from a normal distribution with a variance of one. We report statistics describing the distribution of institutional performance within the state in a policy's tenth year ($t = 10$) for various values of θ and τ .⁴ Reported values of statistics are median values obtained from 500 simulations.

Each cell in Table 1 reports the 5th percentile, the median, and the 95th percentile (separated by commas) of the distribution of the performance variable among the 50 institutions in the simulated state in year ten. The uppermost, leftmost cell in Table 1 describes the simulation results when there is no performance funding policy ($\theta = \tau = 0$). The middle 90% of institutions in the state with no performance funding policy have performance scores between -1.61 and 1.61 with a median of -0.01. These results constitute the main benchmark to which we will compare other simulation results since we want to know how a world without performance funding compares to a world with performance funding under various sets of assumptions. If there is a performance funding policy but the policy has no positive effect on institutional behavior ($\tau = 0$) and resources do not effect performance ($\theta = 0$), the results are identical to when there is no performance funding. Thus, the results in the uppermost, leftmost cell can also represent the results of a performance funding policy under these conditions.

Moving across the first row in Table 1, we can see the results of a performance funding policy if the policy has no positive effect on the behavior of institutional actors within colleges and universities ($\tau = 0$) but does have an effect on performance through resources ($\theta > 0$).

⁴ In year 1, the term $\theta p^*_{i,t-1}$ from equation (5) is ignored (set to zero) since there is no prior performance for the institutions.

Here, performance funding policies serve only to widen the distribution of the performance variable, making the top schools better and the bottom schools worse, respectively, in achieving performance than they otherwise would have been in the absence of a performance funding policy. When $\theta = .25$, the distribution is barely affected at all; the 5th percentile moves from -1.61 to -1.65 while the 95th percentile moves from 1.61 to 1.66. The median is unaffected.⁵ When $\theta = .5$ (indicating either a stronger performance-based component of the funding policy or greater effects of resources on performance than when $\theta = .25$), the distributional effects are noticeably larger. Specifically, the middle 90% of the distribution ranges from -1.85 to 1.87. If $\theta = .75$, the middle 90% range goes from -2.43 to 2.45, indicating a much broader distribution than that found in the absence of a performance funding policy.

When performance funding policies induce positive behavioral changes that affect performance ($\tau > 0$), the distribution of performance outcomes is shifted upward. The bottom two cells of the first column show these behavioral effects when we relax our second key assumption and do not allow resources to affect performance ($\theta = 0$ even though performance funding is in place). Relative to the case of no performance funding, the 5th percentile, the median, and the 95th percentile all increase by an amount equal to the value of τ (differences may vary slightly from τ because the number of simulations was not infinite). These results represent the best-case scenario for performance funding; adoption of a performance funding policy causes the low-end, the middle, and the top-end of higher education institutions to improve.

The remaining cells in Table 1 show results for a state with a performance funding policy that does affect institutional behavior ($\tau > 0$) and where resources do affect performance ($\theta >$

⁵ Table 1 shows a .01 move in the median, but this is likely due to sampling error since the number of simulations is only 500.

0). When $\tau = .25$ and $\theta = .5$, the distribution is wider than in the absence of a performance funding policy (the range between the 5th and 95th percentile is 3.71, compared to 3.22 in the absence of performance funding), but the 5th percentile (-1.60) is not any lower than the 5th percentile under no performance funding policy (-1.61) since the distribution has been shifted upward. Meanwhile, the median institution and the institution in the 95th percentile is considerably better when $\tau = .25$ and $\theta = .5$ than in the absence of performance funding. When we consider the higher value of $\tau = .5$, the 5th percentile makes substantial gains relative to scenario without performance funding as long as $\theta = .25$ or $\theta = .5$. If $\theta = .75$, however, the distribution is widened substantially enough that the 5th percentile is worse than under no performance funding policy even if $\tau = .5$. A graphical depiction of how these two processes (positive behavioral changes and resource reallocation) simultaneously affect the distribution of performance is also shown in the online appendix.

The results from these basic simulations show that performance funding policies can theoretically be expected to have two types of effects on institutions. First, the distance between the best- and worst-performing institutions may widen. Second, the average performance of institutions may increase. Existing literature has largely focused on the latter of these two possible effects, but the former also merits attention. Both types of effects would serve to make high-performing institutions even better. Importantly, the two effects act in opposite directions on low-performing institutions, meaning that if both effects are present, the net effect of a performance funding policy on such institutions depends on which of the two effects is stronger.⁶

Empirically Testing the Effects of Performance Funding on Different Types of Institutions

⁶ We can also consider how the existence of fixed institutional characteristics that affect performance might influence the results of our simulations. See the online appendix for additional discussion in which we recognize both fixed and time-varying components of institutions in additional Monte Carlo simulations.

Since many states have adopted and either repealed or changed performance funding policies over the past three decades (Dougherty and Natow 2015), we can implement a fairly strong test of the effects of performance funding policies on U.S. colleges and universities. Given that state policy changes are driven by actors who usually have little involvement with the operation of universities, we expect that changes in performance funding policies will be largely exogenous. Using two-way fixed effects models, we compare the performance of institutions in years during which performance funding policies are in effect with the performance of those same institutions when they are not subject to performance funding, while adjusting for nationwide year-to-year changes. This modeling approach is considered a type of difference-in-differences design that allows for staggered adoption of policy and reversal of policy adoption (Tandberg and Hillman 2014; Angrist and Pischke 2009).

Data & Methodology

We obtained panel data on institutional characteristics and performance by assembling various data files made available by the National Center for Education Statistics through the Integrated Postsecondary Education Data System (IPEDS). This includes data on the population of degree-granting higher education institutions in the U.S. that participate in federal student financial aid programs (although some variables contain missing data). We restrict our dataset to four-year degree-granting public institutions located within the 50 states (excluding D.C. or US territories) that have a Carnegie classification of bachelor's or higher.⁷ Our unbalanced panel of

⁷ Out of 10,704 observations, 182 in our sample belong to schools that report financial data according to the Financial Accounting Standards Board conventions. As a robustness check, we tried dropping these 182 observations and still found substantively similar results to what are reported here.

520 institutions⁸ contains annual data for a 21-year time period (1993-2013). In total, our dataset contains 10,704 observations.

We use three different measures of performance as dependent variables. As noted earlier, many performance funding policies have used the number of degrees awarded, the graduation rate, and/or the retention rate, sometimes in combination with other factors, as performance measures. We utilize these three measures, all of which provide an indication of the degree to which students actually progress through their training after enrolling in the institution. We measure degree production (mean=22.8; s.d.=6.3) as the number of degrees issued in a year per 100 full-time equivalent students. This is the only performance measure that is available for all years included in our dataset. A second measure is the six-year graduation rate (mean=44.9; s.d.=16.4), which indicates the percentage of students who have graduated among those who entered as first-time, full-time undergraduate students six years prior. This measure is generally the most widely-cited measure but is limited to describing experiences for a more narrow set of students than the degree production measure which better accounts for part-time and transfer students. The graduation rate is only available for the years 1997-2013 in our dataset and is correlated at a mere 0.43 with degree production. The retention rate (mean=73.4; s.d.=10.6) indicates the percentage of first-time full-time students who return for a second year of study. This third measure is available starting in 2004, although the first year of data has a relatively large rate of missing data. It correlates fairly strongly with the graduation rate ($r=0.85$) but more weakly with degree production ($r=0.39$). The degree production and retention rate variables contain some extreme outliers where the rate deviates dramatically for one year and then

⁸ By unbalanced, we mean that some years may include data on all 486 institutions, but many do not. Across all years, 486 institutions appear in the dataset at some point.

immediately returns to a level resembling the year prior to the deviation. We suspect that many of these unusual data points are coding errors; as such, they are dropped from the dataset.⁹

Our main independent variable is an indicator compiled from previous research (e.g., Rutherford and Rabovsky 2014, Dougherty and Natow 2015) of whether a performance funding policy is in effect as well as whether the policies exhibit characteristics of the earliest performance funding policies, termed performance funding 1.0, or contain features implemented in many newer performance funding 2.0 policies. The key distinction generally made between 1.0 and 2.0 policies is that 1.0 policies offer bonus funding as a reward for good performance while 2.0 policies make a portion of institution's base funding contingent on prior performance. 33 states utilize a performance funding policy for at least one year during our study, and only one state (Tennessee) has a performance funding policy in effect for the entire timespan (see online appendix for coding of this variable). Since the effects of performance funding policies could differ depending on the content of the policies, we distinguish between the two categories of performance funding policies in our regressions.

We wish to examine whether performance funding policies have heterogeneous effects on institutions and use multiple modeling approaches to estimate these potential effects. In one set of models, we measure how much heterogeneity there is in performance within each state by calculating the standard deviation of performance in each state-year. We then estimate the effect of performance funding policies on the state-level standard deviation in performance (the unit of analysis is the state-year for these models). If performance funding policies create positive

⁹ Specifically, the degree production variable Y was regressed on Y_{t-1} and then on Y_{t+1} in two separate regressions. Observations were dropped any time the residuals from both regressions were greater than 10 or both residuals were less than -10. For the retention rate, which has a larger standard deviation, the same process was followed except that the thresholds were 15 and -15. Observations were also dropped if the residual from the above regression was greater than 40 or if the reported retention rate was 0.

feedback loops leading to a wider spread in performance, the standard deviation of performance within a state should increase when performance funding policies are in effect.

The other way we test for heterogeneous effects is by trying to directly model how the effects of performance funding policies may be moderating by institutional characteristics. We distinguish among institutions in three different ways. First, we use a lagged dependent variable, which indicates where the institution's performance was in the past year. We demonstrated earlier that if state spending levels are held constant (e.g., we control for the state spending level), performance *relative to other institutions in the state* is what drives the allocation of that funding under a performance funding policy. Thus, we adjust the lagged performance variable to indicate relative performance by subtracting the state-year mean of lagged performance from each institution's own performance. The second measure of institutional characteristics we use is Barron's measure of institutional selectivity, which is based on the characteristics of incoming students (class rank, GPA, SAT/ACT score) and the acceptance rate. The Barron's measure is updated annually; this means it is time-varying in our dataset, although most institutions do not see much movement in this measure over time. With our third measure, we consider whether the institution is a Historically Black College or University (HBCU), a variable obtained from the Delta Cost Project Database (this variable does not change at all over time). We try interacting each of these three measures of institution type with the performance funding indicators in order to see whether the effects of performance funding are contingent upon the type of institution being considered.

Because the adoption or repeal of performance funding policies could be related to fluctuations in a state's environment that could also affect higher education outcomes, we control for a set of time-varying state-level factors. First, we control for the overall level of state

spending on higher education, measured in thousands of dollars per student. As noted in the explanation of our theoretical model, we are interested in comparing the case of a state using performance funding with a counterfactual in which funding is not tied to performance outcomes and overall state financial support for higher education is unchanged. Thus, it is important for us to adjust for any changes in state spending levels while estimating the effects of performance funding policies. We constructed our spending measure by aggregating the Delta Cost's institutional-level measures of revenue from state appropriations (does not include grants, contracts, or capital appropriations) and of total enrollments to the state-year level; this level of aggregation should also help to avoid some of the parent-child reporting issues that have been noted in the Delta Cost data (Jaquette and Parra 2016). We also control for the use of a consolidated governing board in the state (McLendon, Hearn, and Deaton 2006), citizen political ideology (as developed by Berry et al. 1998), and the state unemployment rate (obtained from the Federal Reserve Economic Data).

We estimate our models using OLS regression with the following base equation:

$$y_{syt} = \alpha \cdot pf1_{sy} + \beta \cdot pf2_{sy} + \gamma \cdot \mathbf{x}_{sy} + \delta_i + \theta_y + \varepsilon_{syt},$$

where $pf1_{sy}$ and $pf2_{sy}$ are dummy variables indicating whether a 1.0 or 2.0 performance funding policy is in place in state s in year y , \mathbf{x}_{sy} is a vector of control variables, δ_i and θ_y are institution (i) and year (y) fixed effects, and ε_{syt} is an error term, with standard errors estimated allowing for clustering at the state-year level. The use of two-way fixed allows us to directly test whether institutions do better in years when a performance funding policy is in effect versus in years when a performance funding policy is not in effect while also controlling for nation-wide fluctuations in average performance from year-to-year. While the equation shown above does not obviously resemble the difference-in-differences specification most familiar to many researchers,

our specification is in fact a type of difference-in-difference estimator that allows for staggered implementation of a policy that can be turned both on and off (Tandberg and Hillman 2014; Angrist and Pischke 2009). A dummy variable indicating treated units is unnecessary in our specification because unit fixed effects account for any pre-treatment differences between treated and untreated units. Similarly, time fixed effects account for any sample-wide differences evident in a post-treatment period that a post-treatment dummy might otherwise adjust for. Rather than interacting a post-treatment indicator with a treated unit indicator, the policy dummy is simply coded as a 1 only in the years during which the policy is in effect.

Results

Initially, we test whether there appear to be any significant average effects of performance funding policies on performance without separating out different types of institutions.¹⁰ Table 2 shows the results. The first model indicates that institutions appear to grant, on average, 0.2 additional degrees per 100 students (significant at the .10 level) in years in which a performance funding 1.0 policy applies to those institutions (compared to when no performance funding policy is in effect). Further, 2.0 policies are associated with 0.4 additional degrees per 100 students. The second and third models have insignificant coefficients for performance funding, so we cannot conclude that performance funding affects the average graduation or retention rate. We do not control for time-varying student body characteristics since these could function as mediating variables in the relationship between performance funding policies and institutional performance. Including mediating variables as independent variables in a regression can mask the relationship between the independent variable of interest (performance funding) and the dependent variable.

¹⁰ With many of our models, we tried a variety of specifications that included lags of various lengths, but no consistent evidence of lagged effects was obtained.

Our simulation results suggested that performance funding may increase the spread of performance among institutions. Table 3 provides a direct empirical test of whether this spreading out has occurred. The first model indicates that when states adopt 2.0 policies, the standard deviation for degree production among institutions within a state increases, on average, by 0.5 units. This effect size is quite large considering the mean value of the standard deviation for institutional degree production within a state is 4.0 in our sample. Performance funding 1.0 has no significant effect on the spread of degree rates. The second model in Table 3 indicates that 2.0 policies decrease the standard deviation for institutional graduation rates within a state. This result contradicts our expectations, and the magnitude is quite large (0.7 units compared to a mean of 12.8). As in the first model of spread, 1.0 policies have no significant effect here. The final model in Table 3 shows no significant effect of either set of funding policies on the standard deviation in retention rates.

So far, we have found evidence that 1.0 and 2.0 policies are associated with overall gains in degree production.¹¹ Additionally, 2.0 policies may have effects on the spread of performance within a state, but results are inconsistent regarding whether 2.0 policies increase or decrease the spread, depending on which performance measure is used. In the remaining models, we aim to further explore what might be going on at the institutional level by estimating the effects of funding policies *conditional on specific institutional characteristics*.

Our formal model suggests that performance funding policies may introduce positive feedback into the higher education landscape. More specifically, when performance funding policies are in place, prior performance may affect current performance. We test for positive

¹¹ Sources disagree substantially regarding the appropriate coding of the performance funding policy variable for three states: Colorado, North Carolina, and South Dakota. Dropping these three states does not substantially alter the findings presented in Table 2.

feedback by including a lagged dependent variable in our models and interacting this lagged dependent variable with our policy indicators. In Table 4, we can see that past performance (the lagged dependent variable) is a strong predictor of future performance in the absence of a performance funding policy. To determine whether this positive effect of past performance on current performance grows stronger in the presence of performance funding policies, we must look to the interaction terms. Only one interaction terms is significant (at the .10 level) in Table 4. There is a positive and significant (at the .10 level) coefficient for the interaction of 1.0 policy with past retention rates. The positive coefficient indicates that last year's retention rate is a better predictor of the current retention rate when a 1.0 policy is in place. This also indicates that performance funding 1.0 policies have a more positive effect on the retention rates of institutions that are already attaining a high retention rate relative to other institutions in their state.¹²

Next, we consider whether the effects of performance funding are moderated by institutional characteristics picked up in Barron's selectivity rating. We interact dummy variables for each selectivity level with our funding policy indicators. Because of the large number of coefficients, the results of these models are shown in Figure 1 (coefficients are reported in the online appendix). The first column of graphs in Figure 1 shows the marginal effect of performance funding on degree production. We want to see whether more selective schools accrue larger benefits from the adoption of performance funding policies. The upper-left graph in Figure 1 indicates that 1.0 policies have a positive and significant effect on some relatively non-selective schools (1s and 2s). Overall, the trend for 1.0 policies is a bit ambiguous but may point downward, indicating less positive effects on degree production in more selective schools, which

¹² While there is a theoretical ceiling in performance (a 100% retention rate), even high-performing institutions in our sample rarely come anywhere close to hitting this ceiling (the 99th percentile is 96% retention).

cuts against our expectations. The bottom-left graph in Figure 1 shows the estimated marginal effects of 2.0 policies. Here, the effect size appears to bounce around a bit without any clear upward or downward trajectory as selectivity increases.

The second column of graphs in Figure 1 shows us the marginal effects of funding policies on graduation rates. Here, 1.0 policies appear to have positive effects on some of the least selective schools (1s) and negative effects on the most selective schools (6s). There appears to be an inconsistent but downward trend, cutting against expectations. The estimated effects of 2.0 policies on graduation rates conform more neatly to expectations. As selectivity increases, the marginal effects of performance funding on the graduation rate generally become more positive. Effects for retention rates also appear to mostly conform to expectations. In the top-right corner of Figure 1, we see evidence of an upward trend, although the very most selective schools (6s) depart somewhat from this general pattern. The graph for 2.0 policies looks fairly similar to the one for 1.0 policies, except that schools with a selectivity of 6 are omitted from the bottom-right graph since only one institution with a selectivity of 6 was ever subject to a 2.0 performance funding policy. Overall, the results for selectivity serving as a moderator of performance funding effects are mixed, although there does appear to be some evidence that performance funding 2.0 policies tend to more positively affect the more selective schools for graduation and retention. In terms of our theoretical model, this suggests that θ and τ are both positive for 2.0 policies. Where significant effects exist, their size is generally estimated to be between a half-point and two-point change in the number of degrees or percentage of students graduating or retained.

Finally, given that recent research has suggested a number of disadvantages for minority-serving institutions subject to performance funding policies (Jones 2015, Hillman and Corral 2017), we examine whether performance funding policies have different effects on HBCUs and

non-HBCUs in Table 5.¹³ Since we have included institutional fixed effects in our models and there is no variation over time in whether an institution is designated an HBCU, there is no linear HBCU term included in the model. The linear performance funding terms (labeled “Perf. Fund. 1.0” and “Perf. Fund. 2.0”) show the estimated effects of performance funding for non-HBCUs. In the first column, we see no significant result for performance funding 1.0 on non-HBCUs, but there is a significant, positive effect of performance funding 2.0 on non-HBCUs (an estimated 0.4 additional degrees/100 students). The first interaction term in the model for degree production is significant, meaning that the effect of 1.0 policies differs for HBCUs versus non-HBCUs. Since the interaction term is positive, performance funding 1.0 policies are estimated to have a more positive effect on HBCUs: by adding the linear term to the 1.0 interaction term, we can see that when 1.0 policies are in effect, HBCUs produce an additional 1.2 degrees per 100 students. Looking to graduation rates, we find that while there are no significant effects of performance funding on non-HBCUs, 2.0 policies appears to have a negative effect on graduation rates at HBCU institutions. Specifically, HBCUs have a graduation rate that is 3.7 percentage-points lower (on average) when 2.0 policies are in place. We do not find any evidence of performance funding policies affecting retention rates for either HBCUs or non-HBCUs.

In summary, we find mixed evidence regarding whether performance funding policies affect already high-performing institutions more positively than low-performing institutions. There is not much evidence that 1.0 performance funding policies generally create additional gaps between low and high performers. However, 2.0 policies, which are now in vogue more

¹³ As a robustness check, we tried restricting the sample to include only institutions with a Barron’s selectivity rating of 3 or lower since almost all HBCUs have a selectivity rating of 3 or lower. The results with this restricted sample (not shown here but available upon request) are very similar to what we found with the full sample.

than older 1.0 policies, appear to have less positive or even negative effects on traditionally low-performing institutions, as compared to their higher performing counterparts.

Conclusion

This study highlights the need for scholars and policymakers to rigorously consider the distributional effects of performance funding policies across institutions. While performance funding may attempt to better align the incentives of college and university administrators with the needs of students, the fact that such policies result in allocating extra resources to some institutions—those that perform well—and not others has important implications for students attending institutions that are “rewarded” or “punished” for institutional performance.

Of course, it should be noted that the intent of performance funding policies for institutions of higher education may have little to do with equity. Policymakers and other stakeholder groups might place higher value on efficiency or merit rewards, or they may simply want to justify budget cuts. Even so, there should still be recognition that these policies have the potential to reinforce existing inequalities across campuses and can have important consequences for institutions, institutional actors, and students. Further, multiple actors can attempt to influence the policymaking process to favor some institutions over others. This can stack the deck in favor of institutions with the capacity, social capital, and saliency to have a voice at the table as performance policies are considered and formulated. Likewise, the deck may be easily (intentionally or unintentionally) stacked against those institutions that are already less well-off in a way that further stratifies institutions and the groups of students attending them. It might serve state policymakers to ensure that all institutions are empowered to participate in the policymaking process.

The formal model and computer simulations we presented clearly indicate that performance funding policies can, theoretically, produce two simultaneous effects: one pushing overall performance of colleges and universities upward because of improved performance and the other widening the distribution of performance outcomes among college and universities because of positive feedback resulting from resources being disproportionately allocated to institutions that are performing well. Depending on the relative strength of these two effects, schools at the bottom of the distribution may perform better or worse under a performance funding policy than they would under a funding policy that distributed resources equally without regard to performance. Schools on the high end of the distribution would benefit from either of the two effects. If the amount of resources allocated to a school does not affect its performance, then the distribution of performance effects is not expected to widen, so no school will be harmed by a performance funding policy (if we consider only the factors included in our theoretical model).

The extent to which performance funding policies improve overall performance and/or widen the distribution of performance outcomes among institutions is ultimately an empirical question. We conducted an initial empirical test of the effects of performance funding policies on different types of institutions using a large panel dataset of public U.S. higher education institutions. Overall, our results were mixed. While our full dataset includes over 10,000 observations, several aspects of the dataset were limiting. First, two of our performance measures (graduation rate and retention rate) were only available for more recent years in the dataset. Second, our ability to distinguish among various performance funding policies based on the content of those policies was limited by a single binary distinction (1.0 policies versus 2.0 policies), though policies may vary across multiple dimensions such as the share of total funding,

whether funding influences base or bonus funding, the number of performance indicators for which institutions are accountable, and length of policy implementation. Additionally, there exists conflicting coding schemes for performance funding policies in states across existing studies that do not help to provide clarity in whether some states actually implemented such policies.

Despite these limitations, the results that emerged are fascinating and suggest several avenues for further investigation. We find that performance funding policies are capable of improving overall degree production but have no statistically significant association with graduation or retention rates. Perhaps even more interesting, examining the standard deviation of these performance measures suggests that 2.0 policies widen the spread for degrees, per our initial expectations, while actually lowering the spread for graduation rates, which was not expected. Further, a feedback loop is potentially detected for retention rates when 1.0 policies are in effect, such that those already doing well experience a more positive effect of performance funding policies.

It is not clear from this empirical assessment why performance funding policies would lead in some cases to opposite effects for different sets of performance metrics. One possibility could be related to institutional characteristics. Indeed, our analyses show that more selective schools get larger benefits in the case of 2.0 performance funding policies for graduation and retention rates. A comparison of HBCU and non-HBCU institutions tells a similar story—2.0 policies adversely affect HBCUs for graduation rates. Of course, other reasons are also possible. For instance, performance funding is often expected to induce gaming behaviors, and some performance metrics are more vulnerable to gaming than others. It is also the case that some institutions may be better able to participate in gaming strategies than others. Institutions that are

able to attract more applicants and become more selective or raise tuition to offset declines in other revenue streams come to mind. Alternatively, there could be ceiling effects for the highest performing institutions for some measures, or the availability of resources could affect one performance metric more than the other. Future work should seek to replicate these findings of heterogeneous effects across different measures and attempt to investigate alternative explanations for such differences.

Attention to the differences observed for 1.0 and 2.0 policies is also critical for considering the continued development of performance funding for institutions of higher education, particularly due to the fact that many of the most recently updated or implemented policies would likely be categorized as 2.0 policies. Though policies have continued to evolve in the last few years (and these changes are not captured in our data), it seems that 2.0 policies may introduce more complications with regards to equity of outcomes across institutions. Though this study does not attempt to uncover all of the reasons underlying these differences, they may be due to differences in bonus versus base funding decisions, perceptions of how long the policy may be in place, or more recent capacities at collecting large swaths of data.

If performance funding policies do in fact have negative consequences for some schools by depriving them of resources as some of our results suggest, what can be done to mitigate these concerns? If the factors that make it more difficult for a school to perform well can be identified, performance funding formulae could theoretically adjust for these factors. Some performance funding policies already reward schools for enrolling low-socioeconomic status or minority students (Gandara and Rutherford 2017, Kelchen 2018). To the extent that these types of enrollments indicate that an institution is serving students who—for a number of reasons—may have fewer external supports and thus be at greater risk of dropping out, funding policies may be

able to create a more level playing field for universities (allowing for more equitable distribution of performance-based funds) by giving performance credit for minority and low-socioeconomic status student enrollments. Of course, it would be impossible to fully measure all of the factors that make a college more or less able to hit performance benchmarks, but adjustments for environmental or immutable institutional factors may reduce the extent to which performance-based funding potentially widens the distribution of performance outcomes.

Another possibility for reducing this effect is to make performance-based funding contingent on changes in performance rather than absolute levels of performance. Some policies already incorporate changes in performance to some extent. In Florida, for example, colleges can earn points for either exceeding system-wide averages or by improving their own performance from one year to the next (Florida College System 2018). While this approach could be considered unfair because already-high-performing institutions will only be rewarded if they push their performance to even higher levels, the use of changes in performance should not diminish the constant incentive for high performance. After all, under a policy that attaches resources to changes in performance (rather than absolute levels of performance), any decrease in performance will result in fewer resources while any improvement in performance will result in greater resources. As we develop a better understanding of how performance funding policies affect various types of institutions, we will be able to make more informed statements of how policies can be crafted to minimize any adverse distributional effects of the policies.

One additional limitation of this study is the restricted consideration of potential gaming behaviors. We assume in our interpretation of empirical results that changes in performance metrics reflect true changes in performance rather than the manipulation of metrics. If gaming behaviors are widespread, positive changes in performance metrics might actually indicate

institutional gaming in response to heightened performance incentives. It is harder to envision how gaming behavior would lead to negative effects of performance funding policies on performance metrics, so greater confidence in such results might be warranted. With regards to our theory, if gaming behaviors are widespread, we might expect that the institutions that are most successful at gaming performance metrics will tend to receive more resources, allowing them to improve actual performance. In other words, illusory performance improvements might lead to actual performance improvements. Allowance for gaming behavior, however, should not alter the central insight of the model—namely, institutions that fail to produce good performance numbers (whether real or illusory) may struggle to improve without access to the additional resources that are being handed out to top-performers.

References

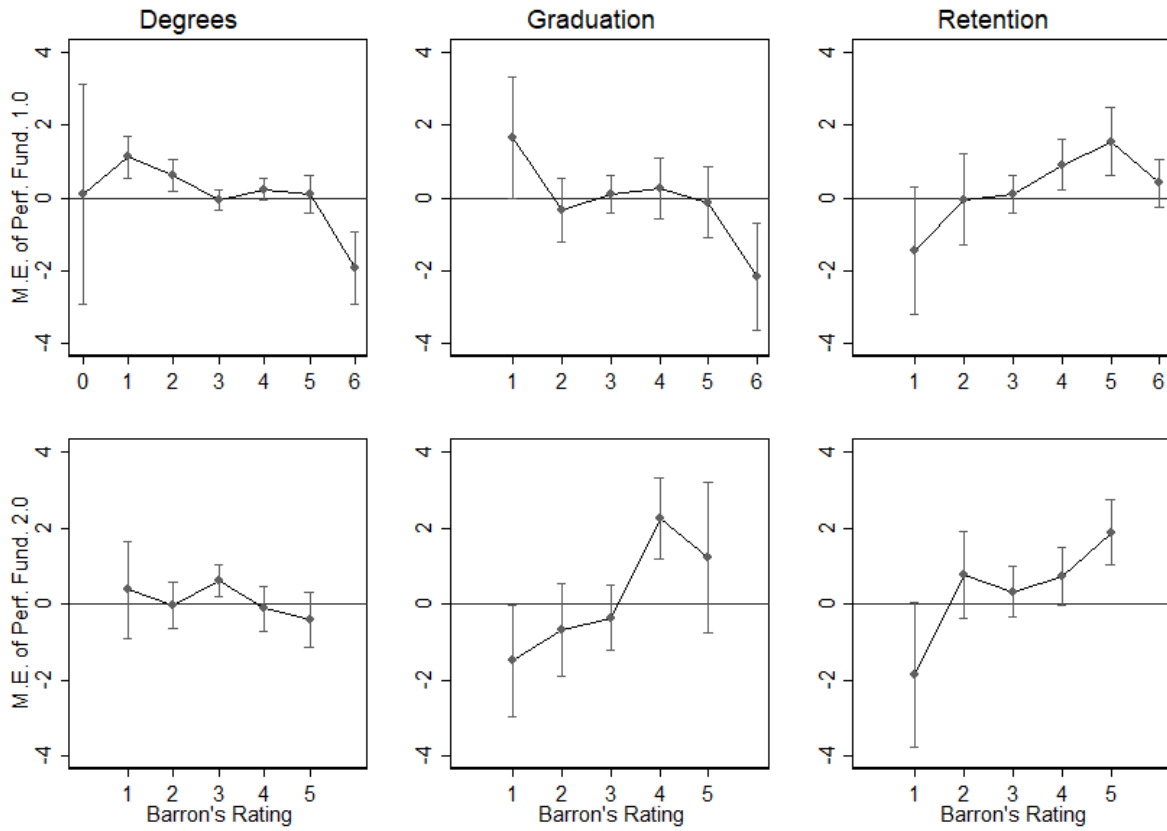
- Angrist, J. D., and J. S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.
- Abernathy, S. 2007. *No child left behind and the public schools*. Ann Arbor, MI: University of Michigan Press.
- Alexander, F. K. 2000. The changing face of accountability: Monitoring and assessing institutional performance in higher education. *The Journal of Higher Education* 71(4): 411-431.
- Balla, S. J. 1998. Administrative procedures and political control of the bureaucracy. *American Political Science Review* 92(3): 663-673.
- Bell, E., A. H. Fryar, and N. Hillman. 2018. When intuition misfires: a meta-analysis of research on performance-based funding in higher education. *Research Handbook on Quality, Performance and Accountability in Higher Education*. Edward Elgar Publishing.
- Behn, R. D. 2002. The psychological barriers to performance management: Or why isn't everyone jumping on the performance-management bandwagon? *Public Performance & Management Review* 26(1): 5-25.
- Bendor, J, S. Taylor, and R. Van Gaalen. 1987. Stacking the deck: Bureaucratic missions and policy design. *American Political Science Review* 81(3): 873-896.
- Berry, W. D., E. J. Ringquist, R. C. Fording, and R. L. Hanson. 1998. Measuring citizen and government ideology in the American states, 1960–93. *American Journal of Political Science* 42(1): 327-348.
- Cantwell, B., and B. J. Taylor. 2013. Global status, intra-institutional stratification and organizational segmentation: A time-dynamic tobit analysis of ARWU position among US universities. *Minerva* 51(2): 195-223.
- Carnevale, A. P., and S. J. Rose. 2003. Socioeconomic Status, Race/Ethnicity, and Selective College Admissions. A Century Foundation Paper.
- Complete College America. 2016. "Performance Funding" Accessed September 21, 2016. Available at <http://completecollege.org/the-game-changers/#clickBoxBlue>.
- Dadgar, M., and M. J. Trimble. 2015. Labor market returns to sub-baccalaureate credentials: how much does a community college degree or certificate pay? *Educational Evaluation and Policy Analysis*, 37(4), 399–418.
- Darling-Hammond, L. 1994. Performance-based assessment and educational equity. *Harvard Educational Review* 64(1): 5-31.
- Deming, D. J., and C. R. Walters. 2017. *The impact of price caps and spending cuts on US postsecondary attainment*. No. w23736. National Bureau of Economic Research.

- Diamond, J. B., and J. P. Spillane. 2004. High stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record* 106(6): 1145-1176.
- Dougherty, K. J., and R.S. Natow. 2015. *The politics of performance funding for higher education: Origins, discontinuations, and transformations*. JHU Press.
- Dougherty, K. J., and V. Reddy. 2011. The impacts of state performance funding systems on higher education institutions: Research literature review and policy recommendations. CCRC working paper No. 37. *Community College Research Center, Columbia University*.
- Dougherty, K. J., R. S. Natow, and B. E. Vega. 2012. Popular but unstable: explaining why state performance funding systems in the United States often do not persist. *Teachers College Record* 114(3): n3.
- Dougherty, K. J., S. M. Jones, H. Lahr, R. S. Natow, L. Pheatt, and V. Reddy. 2014. Implementing performance funding in three leading states: Instruments, outcomes, obstacles, and unintended impacts. *New York, NY: Community College Research Center Working Paper (74)*.
- Dougherty, K. J., S. M. Jones, H. Lahr, R. S. Natow, L. Pheatt, and V. Reddy. 2016. Looking Inside the Black Box of Performance Funding for Higher Education: Policy Instruments, Organizational Obstacles, and Intended and Unintended Impacts. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(1): 147-173.
- Eagan, K., E. B. Stolzenberg, J. J. Ramirez, M. C. Aragon, M. R. Suchard, and S. Hurtado. 2014. The American freshman: National norms fall 2014. *Los Angeles: Higher Education Research Institute, UCLA*.
- Gándara, D. and A. Rutherford. 2017. Mitigating Unintended Impacts? The Effects of Premiums for Underserved Populations in Performance-Funding Policies for Higher Education. *Research in Higher Education*, forthcoming.
- Florida College System. 2018. "2016-17 Performance Funding Model." Available at https://www.floridacollegesystem.com/resources/performance_funding_2016-17.aspx.
- Fryar, A. H. 2011. The disparate impacts of accountability—Searching for causal mechanisms. *Public Management Research Conference, Syracuse, NY*.
- Hagood, L. P. 2017. The financial benefits and burdens of performance funding. Doctoral dissertation. Athens, GA: University of Georgia.
- Hambrick, D. C., and P. A. Mason. 1984. Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review* 9(2): 193-206.
- Hicklin, A., and K. J. Meier. 2008. Race, structure, and state governments: The politics of higher education diversity. *The Journal of Politics* 70(3): 851-860.
- Hillman, N., and D. Corral. 2017. The Equity Implications of Paying for Performance in Higher Education. *American Behavioral Scientist* 61 (14): 1757-1772.

- Hillman, N. W., A. H. Fryar, and V. Crespín-Trujillo. 2018. Evaluating the impact of performance funding in Ohio and Tennessee. *American Educational Research Journal* 55 (1): 144-170.
- Hillman, N. W., D. A. Tandberg, and A. H. Fryar. 2015. Evaluating the impacts of “new” performance funding in higher education. *Educational Evaluation and Policy Analysis* 37 (4): 501-519.
- Hillman, N. W., D. A. Tandberg, and J. P. Gross. 2014. Performance funding in higher education: Do financial incentives impact college completions? *The Journal of Higher Education* 85(6): 826-857.
- Hillman, N. and T. Weichman. 2016. *Education deserts: The continued significance of “place” in the Twenty-First Century*. Viewpoints: Voices from the Field. Washington, DC: American Council on Education.
- Jackson, C. K., R. C. Johnson, and C. Persico 2016. The effects of school spending on educational and economic outcomes: evidence from school finance reforms. *Quarterly Journal of Economics* 131: 157–218.
- Jaquette, O., and E. Parra. 2016. The problem with the Delta Cost Project database. *Research in Higher Education* 57(5): 630-651.
- Jones, T. 2016. A historical mission in the accountability era: A public HBCU and state performance funding. *Educational Policy* 30 (7): 999-1041.
- Kelchen, R. 2018. Do performance-based funding policies affect underrepresented student enrollment? *The Journal of Higher Education* forthcoming.
- Kelchen, R., and L. J. Stedrak. 2016. Does Performance-Based Funding Affect Colleges' Financial Priorities? *Journal of Education Finance* 41(3): 302-321.
- Lafortune, J., J. Rothstein, and D. W. Schanzenbach. 2018. School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics* 10 (2): 1-26.
- Lahr, H. E., L. E. Pheatt, K. J. Dougherty, S. M. Jones, R. S. Natow, and V. T. Reddy. 2014. Unintended impacts of performance funding on community colleges and universities in three states (CCRC Working Paper No. 78). New York, NY: Community College Research Center. Retrieved from <http://ccrc.tc.columbia.edu/Performance-Funding.html>
- Long, B. T. 2016. State Support for Higher Education: How Changing the Distribution of Funds Could Improve College Completion Rates. The Miller Center. http://web1.millercenter.org/commissions/higher-ed/Long_No9.pdf
- Lumina Foundation. 2016. “Frequently Asked Questions: Outcomes-Based Funding” Accessed September 21, 2016. Available at <https://www.luminafoundation.org/outcomes-based-funding-faq>.

- McCubbins, M. D., R. G. Noll, and B. R. Weingast. 1987. Administrative procedures as instruments of political control. *Journal of Law, Economics, and Organization* 3(2): 243-277.
- McLendon, M. K., J. C. Hearn, and R. Deaton. 2006. Called to account: Analyzing the origins and spread of state performance-accountability policies for higher education. *Educational Evaluation and Policy Analysis* 28(1): 1-24.
- Moynihan, D. P. 2008. *The dynamics of performance management: Constructing information and reform*. Washington, D.C.: Georgetown University Press.
- National Conference of State Legislatures. 2015. "Performance-Based Funding for Higher Education." Accessed June 5, 2016. Available at <http://www.ncsl.org/research/education/performance-funding.aspx>
- Paulsen, M. B., and E. P. St. John. 2002. Social class and college costs: Examining the financial nexus between college choice and persistence. *The Journal of Higher Education* 73(2): 189-236.
- Posselt, J. R., O. Jaquette, R. Bielby, and M. N. Bastedo. 2012. Access without equity longitudinal analyses of institutional stratification by race and ethnicity, 1972–2004. *American Educational Research Journal* 49(6): 1074-1111.
- Roscigno, V. J. 2000. Family/school inequality and African-American/Hispanic achievement. *Social Problems* 47(2): 266-290.
- Rutherford, A., and K. J. Meier. 2015. Managerial goals in a performance-driven system: theory and empirical tests in higher education. *Public Administration* 93(1): 17-33.
- Rutherford, A., and T. Rabovsky. 2014. Evaluating impacts of performance funding policies on student outcomes in higher education. *The ANNALS of the American Academy of Political and Social Science* 655(1): 185-208.
- Smallwood, S., and A. Richards. 2011. How educational are state legislators? *Chronicle of Higher Education*. Accessed April 10, 2018. Available at <https://www.chronicle.com/article/How-Educated-Is-Your/127845>.
- Spence, D. B. 1997. Agency policy making and political control: Modeling away the delegation problem. *Journal of Public Administration Research and Theory* 7(2): 199-219.
- Tam, M. 2001. Measuring quality and performance in higher education. *Quality in higher Education* 7(1): 47-54.
- Tandberg, D. A., and N. W. Hillman. 2014. State higher education performance funding: Data, outcomes, and policy implications. *Journal of Education Finance* 39(3): 222-243.
- Umbricht, M. R., F. Fernandez, and J. C. Ortagus. 2015. An examination of the (un)intended consequences of performance funding in higher education. *Educational Policy* 31(5): 643-673.

Figure 1. Marginal Effects of Performance Funding Policies on Performance Measures



Note: Bands show 95% confidence intervals.

Table 1. Monte Carlo Simulations of Institutional Performance (5th percentile, median, 95th percentile)

	$\theta = 0$	$\theta = .25$	$\theta = .5$	$\theta = .75$
$\tau = 0$	(-1.61, -0.01, 1.61)	(-1.65, -0.01, 1.66)	(-1.85, -0.02, 1.87)	(-2.43, -0.01, 2.45)
$\tau = 0.25$	(-1.36, 0.24, 1.86)	(-1.40, 0.24, 1.91)	(-1.60, 0.23, 2.11)	(-2.17, 0.24, 2.71)
$\tau = 0.5$	(-1.10, 0.49, 2.11)	(-1.16, 0.49, 2.16)	(-1.35, 0.48, 2.36)	(-1.93, 0.49, 2.96)

Note: Bolded numbers indicate lower values (worse performance) than those derived in the case of no performance funding policy ($\theta = \tau = 0$). Values that differ by only 0.01 are not bolded (since such slight differences are probably the result of sampling error).

Table 2. Direct Effects of Performance Funding Policies

	Degrees b/se	Graduation b/se	Retention b/se
Perf. Fund. 1.0	0.226+ (0.115)	0.023 (0.215)	0.011 (0.203)
Perf. Fund. 2.0	0.383* (0.169)	-0.115 (0.360)	0.411 (0.297)
State-level Controls:			
State Spend./Pupil	0.073+ (0.040)	-0.155+ (0.092)	-0.158 (0.101)
Citizen Ideology	-0.004 (0.008)	-0.006 (0.020)	0.020 (0.017)
Unemployment	0.056 (0.049)	0.256* (0.095)	0.027 (0.084)
Cons. Gov. Board	0.199 (0.475)		
Adj R-sqr	0.811	0.923	0.917
N	10649	8333	4830

Unit of analysis is the institution-year.

Institution- and year- fixed effects not shown.

Standard errors are clustered by state-year and shown in parentheses.

+ p<0.10, * p<0.05

Table 3. Effects of Performance Funding Policies on State-level Variance in Performance

	SD(Degrees)	SD(Graduation)	SD(Retention)
	b/se	b/se	b/se
Perf. Fund. 1.0	0.045 (0.115)	-0.281 (0.192)	0.002 (0.266)
Perf. Fund. 2.0	0.465* (0.190)	-0.695* (0.267)	-0.060 (0.334)
State-level Controls:			
State Spend./Pupil	0.072+ (0.042)	0.064 (0.079)	0.204* (0.096)
Citizen Ideology	0.003 (0.008)	0.006 (0.015)	0.003 (0.018)
Unemployment	0.103* (0.043)	0.039 (0.087)	0.035 (0.087)
Cons. Gov. Board	0.609* (0.235)		
Adj R-sqr	0.648	0.877	0.828
N	1028	815	441

Unit of analysis is the state-year.

State- and year- fixed effects not shown.

Robust standard errors are shown in parentheses.

+ p<0.10, * p<0.05

Table 4. Interactive Effects of Performance Funding Policies and Prior Performance

	Degrees b/se	Graduation b/se	Retention b/se
Lagged DV	0.650*	0.435*	0.217*
	(0.021)	(0.028)	(0.029)
Perf. Fund. 1.0	0.260*	0.023	0.069
	(0.122)	(0.215)	(0.214)
PF 1.0 X Lagged DV	-0.052	-0.014	0.042+
	(0.038)	(0.019)	(0.024)
Perf. Fund. 2.0	0.468*	0.137	0.374
	(0.186)	(0.353)	(0.322)
PF 2.0 X Lagged DV	0.003	0.000	0.040
	(0.041)	(0.017)	(0.026)
State-level Controls:			
State Spend./Pupil	0.068	-0.145	-0.184
	(0.042)	(0.095)	(0.114)
Citizen Ideology	-0.005	-0.002	0.013
	(0.008)	(0.020)	(0.018)
Unemployment	-0.009	0.276*	0.040
	(0.051)	(0.090)	(0.094)
Cons. Gov. Board	0.450		
	(0.578)		
Adj R-sqr	0.878	0.944	0.927
N	10107	7797	4312

Unit of analysis is the institution-year.

Institution- and year- fixed effects not shown.

Standard errors are clustered by state-year and shown in parentheses.

+ p<0.10, * p<0.05

Table 5. Interactive Effects of Performance Funding Policies and HBCU Status

	Degrees b/se	Graduation b/se	Retention b/se
Perf. Fund. 1.0	0.156 (0.118)	-0.028 (0.208)	-0.035 (0.194)
PF 1.0 X HBCU	1.007* (0.474)	0.694 (1.550)	0.585 (1.021)
Perf. Fund. 2.0	0.390* (0.169)	0.183 (0.349)	0.459 (0.299)
PF 2.0 X HBCU	-0.030 (0.508)	-3.918* (1.088)	-0.660 (1.119)
State-level Controls:			
State Spend./Pupil	0.069+ (0.040)	-0.160+ (0.092)	-0.163 (0.101)
Citizen Ideology	-0.003 (0.008)	-0.004 (0.020)	0.021 (0.017)
Unemployment	0.055 (0.049)	0.244* (0.095)	0.025 (0.085)
Cons. Gov. Board	0.203 (0.475)		
Adj R-sqr	0.811	0.923	0.917
N	10649	8333	4830

Unit of analysis is the institution-year.

Institution- and year- fixed effects not shown.

Standard errors are clustered by state-year and shown in parentheses.

+ p<0.10, * p<0.05