Can We Simply Raise the Bar on Teacher Quality?

James Shuls, PhD

Spring 2016

About the Author

James Shuls is the director of education policy at the Show-Me Institute in St. Louis, Missouri. He earned his PhD in education policy from the University of Arkansas. Prior to attending the U of A, he earned bachelor's and master's degrees in elementary education and was a public school teacher in southwest Missouri.

About ACRE

The Arkansas Center for Research in Economics (ACRE) is a university-based, nonprofit research and education organization that focuses on improving the lives of Arkansans by training students, by evaluating the state's policies, and by offering solutions in order to promote free markets and limited, transparent government. ACRE does not promote, advocate, support, or campaign for specific political parties or political candidates.

If you have questions or comments, or if you would like more information about this study or about ACRE and its endeavors, please contact ACRE's director, David Mitchell (dmitchell@uca.edu), or the author, James Shuls (james.shuls@showmeinstitute.org).

Contents

Can We Simply Raise the Bar on Teacher Quality?

James Shuls, PhD

Executive Summary

Improving the classroom performance of Arkansas's teachers could be the key to sustaining the state's recent educational improvements. What is the best way to accomplish this goal? Counterintuitively, studies show that allocating more money to public education per student has weak and inconsistent effects on boosting student achievement. The quality of public education will not be greatly improved simply by increasing school budgets. However, studies consistently show that teacher effectiveness plays a critical role in how much a student learns. The difference between an effective teacher and an ineffective teacher can be as much as a year's worth of learning.

Policymakers regularly seek to improve the quality of the teacher workforce. They often try to achieve this by constructing barriers to entering the profession. Many policy makers wish to increase the minimum score, though Arkansas recently reduced the minimum score on the licensure exams. The logic behind raising the bar is that by screening prospective teachers on the front end, state officials prevent ineffective teachers from entering the classroom. The logic behind lowering the bar is that licensing scores are a mediocre predictor of teacher success. Teaching is a doing profession, and teachers need to be assessed on their actual teaching not their exam grades. It is harder to improve quality within the system than previously believed.

Ideally, for teacher licensure to work, the licensure exam should be highly correlated with teacher performance in the classroom. Since licensing scores and teaching performance are imperfectly correlated, raising the passing score on licensure exams prevents ineffective teachers from becoming licensed, but also prevents many effective teachers from becoming licensed. Furthermore, in the absence of perfect correlation, raising minimum passing licensure exam scores may have a disproportionately negative impact on the most disadvantaged schools. In Arkansas, there is a positive relationship between performance on licensure exams and teacher performance in the classroom, but that correlation is relatively low, and neither of the primary exams for teacher licensure is a perfect licensure screen. At all possible exam scores, there are

both effective and ineffective teachers. Therefore, simply raising the score needed to pass the licensure exams in Arkansas does not appear to be an effective strategy.

Rather than consume time, effort, and political will to pursue policy changes that are unlikely to improve teacher effectiveness—such as expanding budgets and requiring higher licensure scores—Arkansas's leaders and citizens would be better served by searching for innovative approaches to improving teacher quality such as loosening entry restrictions and introducing performance incentives for teachers. A well designed pay-for-performance incentive can improve teachers' performance both by clarifying teaching goals and by attracting and retaining effective teachers (Lavy, 2007).

Teacher Quality in Arkansas: What Do We Know? How Much Does It Matter?

In every profession, the difference between an effective employee and an ineffective one can be substantial. This is especially true in education, where the quality of a teacher can have a significant impact on a child's life. Hanushek and Rivkin (2006) find that students in the average teacher's classroom learn a year's worth of material, while students in an ineffective teacher's classroom learn only half a year's worth of material. Students in a highly effective teacher's classroom learn a year and a half's worth of material. As a result, being in an ineffective teacher's classroom for two years could put a student a full year behind their peers who had average teachers and even further behind students who had highly effective teachers.

President Obama noted the importance of teachers in his 2012 State of the Union address when he cited a study that linked the academic records of 2.5 million students to adult outcomes (Chetty, Friedman, and Rockoff 2011). The study found that students who had highly effective teachers, as measured by increased student achievement (value-added student achievement), were more likely to go to college, earn higher salaries, and save more for retirement. Teachers who produced the greatest student learning gains also contributed to better outcomes for those students later in life.

Arkansas policymakers recognize the importance of teacher quality and have enacted many policies to improve students' education. These efforts have earned the state high marks from the notable publication *Education Week,* which releases an annual state-by-state report card called Quality Counts on the status of education. Quality Counts assesses the condition of each state's education system in three categories: chance for success; K–12 achievement; and school finance. In 2015, Arkansas scored poorly in most of the categories of standards, assessments, and

accountability; transitions and alignment; and teaching profession, giving the state a 36th place ranking overall (Education Week 2015a).

Previously *Education Week* included six categories instead of three. The previous ranking showed that Arkansas had improved over the past two decades. However, the state's high scores came primarily from categories that focus on inputs rather than outputs. This scoring clouds the picture of how Arkansas students are faring. In what is arguably the most important area of Quality Counts, K–12 student achievement, Arkansas received a C- and was below the national average of C (See table 1).[1]

Table 1. Quality Counts 2015 rankings on K–12 achievement

State	Letter Grade	Rank	State	Letter Grade	Rank	State	Letter Grade	Rank
Massachusetts	B	1	Colorado	C-	18	Dakota	D+	35
New	B-	2	Nebraska	C-	19	Missouri	D+	36
Hampshire	C+	3	Illinois	C-	20	Oregon	D	37
Vermont	C+	4	Wyoming	C-	21	Nevada	D	38
Maryland	C+	5	Georgia	C-	22	Oklahoma	D	39
Minnesota	C	6	Rhode	C-	23	Kansas	D	40
Virginia	C	7	Texas	C-	24	Arkansas	D	41
Indiana	C	8	Ohio	C-	25	Alaska	D	42
Wisconsin	C	9	Arizona	C-	26	Michigan	D	43
Pennsylvania	C	10	New	C-	27	South Dakota	D	44
Florida	C	11	Montana	C-	28	Alabama	D	45
Connecticut	C	12	Iowa	C-	29	South Carolina	D	46
Washington	C	13	California	D+	30	District of Columbia	D	47
Utah	C	14	Idaho	D+	31	West Virginia	D	48
Maine	C-	15	Hawaii	D+	32	Louisiana	D	49
Kentucky	C-	16	North C	D+	33	New	D-	50
Tennessee	C-	17	North	D+	34	Mississippi	D-	51

Source: Education Week

The Quality Counts report uses three measures to create its grade for K–12 student achievement: high school graduation rates, performance on Advanced Placement (AP) exams, and performance on the National Assessment of Educational Progress (NAEP). Arkansas's graduation rate is 78 percent, ranking the state 34th in the country. The state was also well below the national average in the percentage of students passing their AP exams. Nationwide, 29.3 percent of test-takers passed, while just 21.5 percent of Arkansas students passed.

Arkansas also falls below the national average on the NAEP. The NAEP, known as the nation's report card, "is the largest nationally representative and continuing assessment of what America's students know and can do in various subjects" (National Center for Education Statistics 2015). It is the best measure for comparing one state to another and for comparing a state to itself over time. Random samples of fourth- and eighth-graders take the test in math and reading every two years. In 2015 Arkansas eighth-graders scored significantly below the national average in both subjects (National Center for Education Statistics 2015a).

While Arkansas students improved between 1990 and 2011, by 2015 of its neighboring states, Arkansas outperformed only Louisiana and Mississippi in K–12 achievement, Hanushek, Woessmann, and Peterson (2012) examined each state's performance on the NAEP from 1992 through 2011 and found that Arkansas's gains were among the highest in the nation. But *Education Week's* analysis of more recent achievement gains between 2003 and 2015 showed that much work is still needed. On 8th grade math achievement, our gains placed us 9th which is good..But on 4th grade reading achievement our gains placed us 25th and on the. 4th Grade Math 28th, and on 8th Grade Reading 34th.

To return to the previous positive trend, Arkansas should continue to seek out policies that will improve teacher quality. On the surface, an obvious answer seems to be simply to increase government spending on K–12 schooling. However, the evidence from academic research reveals that increasing spending has very little, if any, direct impact on student achievement. Preeminent education researcher Hanushek (1997) reviewed nearly 400 studies that related spending on education to student outcomes and concluded that the relationship between education spending and student outcomes is neither strong nor consistent. Since then, no evidence has overturned Hanushek's findings. Increased inputs do not seem to lead to improved outputs in public education (Hanusheck 2003). Somewhat gloomily, he concluded that "simple resource policies hold little hope for improving outcomes" (Hanushek 1997, p. 141).

If spending more money on K–12 education will not work, then what can be done to improve the quality of Arkansas's teacher workforce? Generally, there are three broad strategies that any state might attempt: screening out ineffective teachers on the front end, helping current teachers improve through professional development, and removing ineffective teachers from the classroom. There are benefits and shortcomings to each of these strategies, and each deserves careful study.

The goal of this research is to examine one front-end strategy that raises the barrier to entry: increasing licensure exam "cut scores." The cut score is the minimum score that a teacher must attain on statewide licensure exams in order to earn a teaching license. Nearly all—but not all—teachers practicing in Arkansas scored at or above this threshold in order to obtain their licenses. Increasing cut scores is a strategy that some states, including Arkansas, have used in the past in an effort to improve teacher quality. This paper explores whether increasing cut scores will improve teacher quality in two ways. First, it examines whether the students of teachers who

have passed licensing exams score significantly higher on standardized exams than do the students of teachers whose score is not high enough to pass the licensure exam. Second, it examines what would happen if Arkansas were to raise the score required to pass its licensure exams.

Teacher Licensure Exams in Arkansas

Nationwide, the most commonly used licensure exams for teachers were the Praxis I and Praxis II exams. (They have since been renamed.) Aspiring teachers must pass both. On September 1, 2010, Arkansas began requiring prospective teachers to take the Praxis I pre-professional skills assessments. The Praxis I assessments required specific scores in reading, writing, and math. Though the Praxis series is the most widely used licensure exam, passing scores are set at the state level and vary across states. During this research period, Arkansas, teachers needed to score a 172, 173, and 171 on the respective Praxis I exams. However, in the neighboring state of Louisiana, for example, teachers must score a 176, 175, and 175. It is this state-to-state variation in Praxis cut scores that allows us to estimate the effects of raising the cut score in Arkansas.[1]

Licensing exams are used to measure both a teacher's intelligence and a teacher's grasp of content and pedagogical knowledge. Fergusson and Ladd (1996) and Boyd et al. (2008) have established that individuals scoring higher on the ACT and SAT—tests that arguably measure intelligence—are more effective teachers, on average. Consistent with these results, researchers have found a positive relationship between a teacher's performance on licensure exams and a teacher's performance in the classroom (Clotfelter et al. 2006, 2007; Goldhaber 2007). Using Arkansas data, Shuls and Trivitt (2013) found both the Praxis I and Praxis II licensure exams to

[1] The current Praxis exams are called Praxis Core and Praxis Core II which has two sections. The first section is content and the second is pedagogy.

be positively correlated to teacher effectiveness in math and language arts.[2] However, this is far from a perfect screen.

While generally true, these results may not present the whole picture. Imposing cut scores on standardized exams such as the ACT, Praxis I, and Praxis II will have some benefits as a screening method, but it is unlikely to be a perfect screen. Allowing teachers to be certified on the basis of how they perform on the Praxis exams will still allow some low-quality teachers to become certified while preventing some high-quality teachers from becoming certified. Similarly, increasing the cut score might weed out some additional applicants who would turn out to be poor teachers, but doing so would also likely block additional applicants who would turn out to be good teachers. Increasing cut scores on the Praxis exams does not seem to be a perfect method for ensuring that every teacher in Arkansas is a good teacher.

To study these issues, I examine the differences between potential teachers who pass the Praxis exams—those who score at or above the cut scores—and those who fail the exams. Researchers have examined this question infrequently, and the results have been inconclusive. Hanushek et al. (2005) found no difference between the effectiveness of teachers who had *not* passed a state's licensure exam and those who had. In North Carolina, Goldhaber (2007) examined the impact of raising licensure exam cut scores. He found no relationship between passing the exam and teacher effectiveness in reading, but he found a positive relationship in math. He concluded, "If states are seeking criteria to ensure a basic level of quality, then licensure tests appear to have some student achievement validity" (Goldhaber 2005, p. 788). Despite this conclusion, Goldhaber noted that licensure exams are not a perfect measure of quality. He found that by raising its cut score, North Carolina would not dramatically improve

the teacher workforce. It would, however, eliminate many potentially effective teachers from the labor force.

Data and Methods

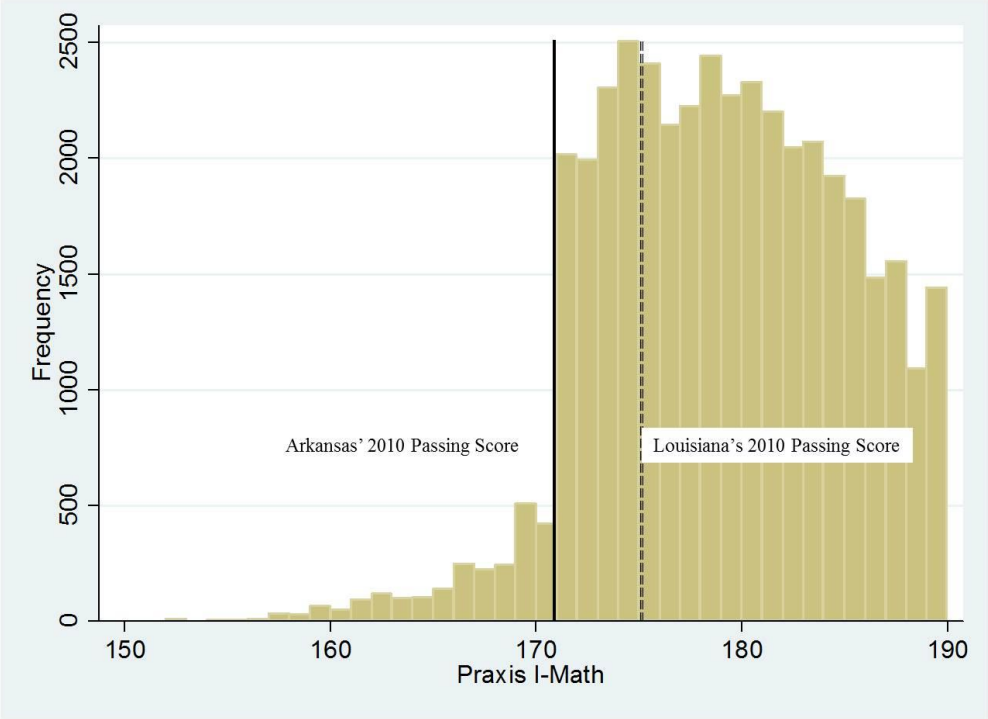
I used a number of data sets to examine the impact that raising the minimum score on teacher licensure exams would have on the teacher workforce. Both the teacher data and the student data are across schools and across time. The data were provided by the Arkansas Department of Education (ADE).

The Teacher Data

The teacher data indicate where a teacher is employed and the subject he or she teaches in a given year. The data also include descriptive characteristics, including race, gender, and whether the teacher has an advanced degree. In addition, the data indicate how well a teacher performed on various licensure examinations. Scores are provided for the three sections of the Praxis I: reading, writing, and mathematics. As noted, to pass the licensure exams in Arkansas, teachers needed score a 172, 173, and 171 on the respective Praxis I exams.[2] In the neighboring state of Louisiana, teachers needed to score a 176, 175, and 175. Praxis I exams, by design, do not generate a normal bell curve distribution. Teachers often achieve the maximum possible score. This outcome results in a negatively skewed distribution, as figure 1 shows.

Figure 1. Histogram of Arkansas' teacher scores on the Praxis I mathematics exam

[2] The current scores are 156, 162, and 150.



Source: Arkansas Department of Education

The ADE also provides scores on a variety of Praxis II examinations. These tests are in content areas and professional knowledge or pedagogy. Arkansas required teachers to pass both Praxis I and Praxis II. The Praxis II exam scores are similarly distributed, but, unlike the Praxis I, no state sufficiently matched the specifics of Arkansas's required content exams so no comparison for the cut score exists. More to the point, teachers of various subjects take a variety of different Praxis II tests. Thus, the same strategy cannot be employed for analyzing both the Praxis I and Praxis II scores. Therefore, I standardized all exam scores for the Praxis II and estimated the impact of raising the cut score by 0.25 standard deviations (slightly less than the difference between Arkansas's and Louisiana's cut scores on the Praxis I tests).

On the Praxis I mathematics exam, there is a large spike in the number of test takers at 171—Arkansas's passing score. This spike occurs primarily because people who do not score at least 171 typically will not become teachers; they are screened out. Since Arkansas raised its

minimum score to match Louisiana's, nearly everyone who falls below the dashed line during the study years in figure 1 would be weeded out

Some teachers whose scores fall below Arkansas's cut score still get a teaching license for various reasons. These individuals could be teachers who were needed to fill a vacant position and were awarded a temporary license, teachers who passed the test during a time when the cut scores were lower, or teachers who moved to Arkansas from a state with teacher reciprocity that required lower scores.

The Student Data

The student data assigns each student a unique ten-digit identifier that indicates the grade and school a student is enrolled in during a given year. The data also indicate whether the student is eligible to receive free or reduced-price lunches under the National School Lunch Act. Additionally, the data indicate whether the student has an Individualized Education Program to provide for a special need or is an English-language learner. Student demographics are relatively consistent. For each year, there are more than 200,000 student records. Table 2 displays relevant demographic statistics for the 2007–08 school year. As the table shows, the majority of students were white and over half were eligible to receive free or reduced-price lunches.

Table 2. Demographics for Arkansas students in grades 3–8, 2007–08 *(Study Years)*

Student demographics	2007–08
Total students (grades 3–8)	209,844
Free or reduced-price lunch	56.0%
English language learner	5.5%
Individualized Education Program	10.8%
Female	49.1%
White	67.4%
Black	22.1%
Hispanic	8.2%

Source: Arkansas Department of Education.

Students in Arkansas used to take Benchmark achievement exams in grades three through eight in language arts and mathematics. They currently take ACT Aspire. This student data from ADE include student records of these exam results from 2005 through 2008. Although the Benchmark—in part—is designed to reflect how student learning progresses sequentially through the years, for analytical purposes I standardized student test scores within each grade and year with a mean of zero and a standard deviation of one.

Overview of Methods

I ask two related research questions in this paper: whether teachers who pass the licensure exams outperform teachers who do not pass the licensure exams and what impact raising the required score to pass the licensure exams would have on the effectiveness of the teacher workforce.

This analysis measures teacher performance in terms of a teacher's impact on student achievement. The process of linking student achievement gains to having been a particular teacher's student is called "value-added modeling," or VAM. In an ideal model, students would be randomly assigned to teachers and the data would clearly link students to teachers. Unfortunately, Arkansas's data are not suited for this type of analysis. Students are linked to teachers geographically and by grade, so a true random sampling is not possible. Furthermore, the data do not link students to specific teachers; in Arkansas, the data are only linked by student to school and grade. Therefore, in this analysis, I utilize a two-step strategy for estimating a teacher's impact on increasing student achievement.

In the first stage of the analysis, individual students' Benchmark test scores are regressed on one- and two-year lagged student test scores;[3] that is, I use Benchmark scores from the previous two years to predict a student's Benchmark score the current year. I interpret the

difference between the actual test score and the test score predicted by the model—the residual—as the value added by the student's teacher for that school year. A residual value greater than zero indicates that a student's current performance is better than what was predicted from the same student's previous exam scores. This might be because in the current year, the student was taught by a high-quality teacher. Alternatively, a residual value less than zero indicates that a student's current performance is worse than what was predicted from previous exam scores. This might be because in the current year, the student was taught by a low-quality teacher.

Using changes in individual student's test scores helps control for differences in students' backgrounds, prior performance, and personal characteristics. As expected, a student's performance last year in English language arts (ELA) and math is the strongest predictor of performance this year on the corresponding exam (see appendices A and B).

Because the data do not allow researchers to directly match teachers to their students, I calculate the value added at each grade level by school. What this means is that all teachers of a particular grade at the same school will be treated as being equally effective at adding educational value. This adds imprecision to the measure, but it should not statistically bias the outcomes.

In the second stage of the research strategy, I use the measure of performance by the teachers in a grade level—the residual from the first stage regression—as the variable to be "explained" by various teacher characteristics, including teachers' performance on licensure exams. A positive relationship between the residual and licensure exam scores would indicate that teachers with *higher* licensure exam scores tend to produce *larger* student learning gains.[4] A negative relationship would indicate that teachers with *lower* licensure exam scores tend to produce *smaller* student learning gains. The third possibility is that there is no detectable

relationship between a teacher's licensure exam score and students' learning gains. Notice that the unit of observation here is the individual teacher, since the value-added measure from the first regression has been assigned for each teacher within a grade at a school. (For a technical explanation of this strategy, see Shuls and Trivitt [2013].)

I then use this analysis to determine the effectiveness of the Praxis I as a licensure screen and to estimate, for the purposes of the example in this paper, the impact of raising the cut scores from Arkansas's levels to Louisiana's levels.

The Relationship between Failing the Praxis and Teacher Effectiveness

As noted previously, there are teachers currently working in Arkansas who have scored below Arkansas's cut score on licensure exams. Whatever the cause for allowing these teachers into the classroom, their presence provides an opportunity to estimate whether teachers who pass the exam are more effective, on average, than those who do not pass the exam.

Recall that Louisiana and Arkansas require teachers to take the same Praxis I exams, but Louisiana's cut scores are several points higher than Arkansas's. Also recall that there are dozens of Praxis II exams, and states have varying requirements. Because I cannot compare Arkansas's Praxis II scores to other states' scores, I estimate the impact of raising the various Praxis II licensure exam cut scores by 0.25 standard deviations—slightly less than the difference between Arkansas's and Louisiana's cut scores on the Praxis I.

Using a technique similar to Goldhaber's (2007), I use Louisiana's scores as a guideline to estimate the impact of raising Arkansas's cut scores to Louisiana's mark. As expected, the number of teachers who would fail the exam increases on both Praxis I and Praxis II when I use the Louisiana cut scores for Praxis I and the raised (by 0.25 standard deviations) score for Praxis

II. Using these higher scores, the number of teachers who would fail climbs from 291 on Arkansas's Praxis I exam to 1,150 (see appendix C). The failure rate for the Praxis II exam nearly doubles, from 889 to 1,604.[5]

Under existing Arkansas standards, teachers who failed the licensure exam (both the Praxis I and Praxis II) were more likely to be minorities, more likely to have an alternative teaching license that does not require an education degree, and more likely to have fewer years of experience (see appendix C). Teachers who failed under existing standards were also more likely to work in schools serving higher percentages of disadvantaged and minority students. Therefore, it is likely that raising Arkansas's cut scores to equal Louisiana's would potentially exclude minority candidates who came into teaching through alternative methods and who would go on to teach in poorer districts that have more minority students.

Table 3 (mathematics) and table 4 (ELA) analyze whether passing the Praxis I in Arkansas is a significant indicator of teacher quality. The results in these tables indicate that passing the Praxis I in Arkansas is *not* a significant indicator of teacher quality. In both math and ELA, the difference between those who pass and those who fail is negative, but the difference is not statistically significant. This means that the Praxis I exam is screening out some individuals who may be ineffective teachers, but it is also screening out some potentially effective teachers. When the Praxis I cut scores are raised to those used in Louisiana, the teacher performance difference between those who passed and those who failed is still not statistically significant, but the total number of failed examinees would increase.

Table 3. Relationship between failing a licensure exam and teacher effectiveness in mathematics (robust standard errors in brackets), sample size 17,627

	Praxis I (Arkansas cut score)	Praxis I (Louisiana cut score)	Praxis II (Arkansas cut score)	Praxis II (0.25 sd higher)
Teacher effectiveness	–0.00421	–0.00387	–0.0147**	–0.0192***
	[0.0126]	[0.00676]	[0.00721]	[0.00560]
R-squared	0.004	0.004	0.004	0.005

Note: *** p<0.01, ** p<0.05, * p<0.1
Source: Author's calculations based on Arkansas Department of Education Data.

Table 3 indicates that the Praxis II exam is a more effective licensure screen for mathematics as compared to the Praxis I. Teachers who fail a Praxis II exam at the Arkansas cut score tend to be significantly less effective in teaching math. If Arkansas were to raise the various Praxis II cut scores by 0.25 standard deviations, the higher barrier to entry would continue to screen out some ineffective math teachers; teachers who failed the Praxis II at the higher cut score level were, on average, significantly less effective than those who passed.

Table 4 shows that Praxis II may be less effective in screening ELA teachers. Teachers who fail the Praxis II at the actual Arkansas cut score tend to be no less effective in teaching ELA than the teachers who pass the Praxis II. However, raising the Praxis II cut score would screen out some ineffective ELA teachers. Teachers who would fail a Praxis II at the higher cut score are significantly less effective ELA teachers.

Table 4. Relationship between failing a licensure exam and teacher effectiveness in English language arts (robust standard errors in brackets), sample size 17,627

	Praxis I (Arkansas cut score)	Praxis I (Louisiana cut score)	Praxis II (Arkansas cut score)	Praxis II (0.25 sd higher)
Teacher effectiveness	–0.00539	–0.00117	–0.00392	–0.00851**
	[0.00974]	[0.00482]	[0.00499]	[0.00410]
R-squared	0.001	0.001	0.001	0.002

Note: *** p<0.01, ** p<0.05, * p<0.1
Source: Author's calculations based on data from Arkansas Department of Education and Louisiana Department of Education data.

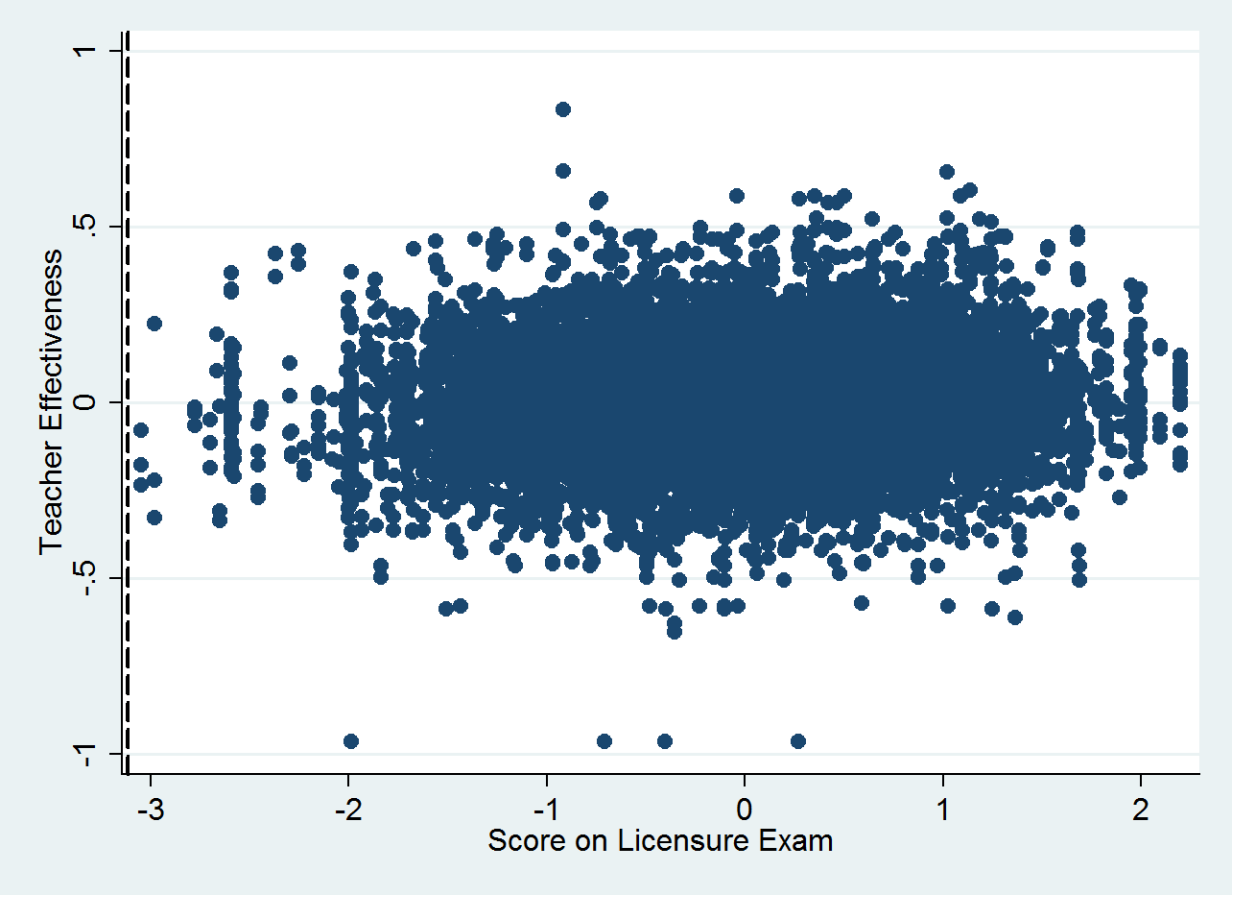
Although the differences between teachers who passed the Praxis II and teachers who failed were statistically significant, there is a question as to whether these differences are practically significant. In other words, are the differences large enough to warrant action? Figure 2 presents a scatterplot that displays the relationship between a teacher's effectiveness and his or her performance on the Praxis II professional knowledge exam. Both variables were adjusted so that a value of zero is the average Praxis II score on the horizontal axis and the average effectiveness of an Arkansas teacher.[6]

Figure 2 demonstrates that teachers vary widely in their performance over the entire range of Praxis II scores. This means that regardless of the Praxis II score, some teachers are of above average effectiveness and some teachers are of below average effectiveness. Applying cut scores to the Praxis II exams is a partially effective screening tool, but not a perfectly effective one. If the Praxis II were a perfectly effective screening method, then all teachers above the cut score on the horizontal axis would be at zero or above in teacher effectiveness; meanwhile, all teachers below the cut score would be below zero in teacher effectiveness. In reality, though, some teachers with low Praxis II scores—including scores below the cut point—are effective teachers, and some teachers with high Praxis II scores are ineffective teachers.

Imagine drawing a vertical line in figure 2 between the (–)3 and (–)2 across the horizontal axis, roughly consistent with current Arkansas cut scores. The individuals to the left of the line would fail the licensure exam, and the individuals to the right would pass the exam. Now imagine sliding that vertical line to the right, increasing the cut score to a level similar to Louisiana's. As the required passing score increases, Arkansas would remove more low-performing teachers, but would also remove many high-performing teachers.

This imperfect correlation between improving students' scores and performing well on the various Praxis exams makes intuitive sense. Subject knowledge matters, but teaching is a doing profession. It is not a test taking profession. The two are related but not closely.

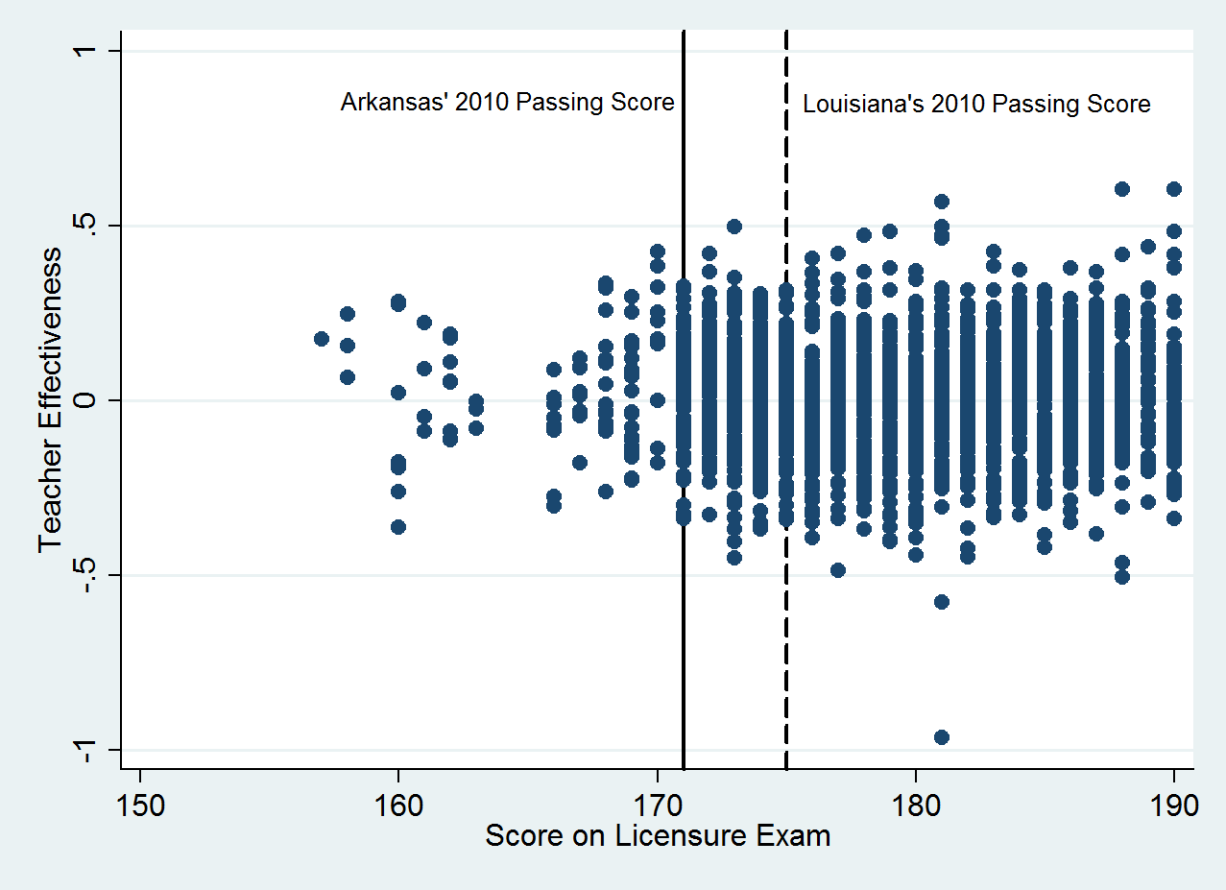
Figure 2. Scatterplot of teacher effectiveness in mathematics and performance on the Praxis II professional knowledge exam



Source: Author's calculations based on Arkansas Department of Education data.

Figure 3 presents a more concrete illustration of this imperfect relationship. I plot the relationship between teacher effectiveness in math and a teacher's score on the Praxis I mathematics exam. I include a vertical line at 171, Arkansas's cut score in 2010. The dashed line represents Louisiana's cut score in 2010. The individuals between the two lines represent those who passed Arkansas's current licensure exam but who would fail under Louisiana's requirements.

Figure 3. Scatterplot of teacher effectiveness in mathematics and performance on the Praxis I math exam



Source: Author's calculations based on Arkansas Department of Education data.

As figure 3 shows, these individuals vary in terms of effectiveness from a half of a standard deviation above the mean to a half below. Thus, raising the cut score would remove

many ineffective (and some exceptionally ineffective) teachers, but it would remove nearly as many effective (and some exceptionally effective) teachers.

Policy Implications and Conclusion

What can be done in the area of public policy to keep Arkansas's public education on its current upward trajectory? Hanushek (1997) demonstrated that simply devoting more public spending to education is not a solution. Increasing government spending on education in Arkansas would likely maintain the current quality, but at a higher price. Hanushek and Rivkin (2006) found that teacher quality is an important determining factor of student outcomes; high-quality teachers should yield higher student performance. Therefore, it seems sensible to study the state's policies in its role as the gatekeeper that determines who will become a licensed teacher. The implicit purpose of licensure exams is to ensure a minimum quality among teachers by preventing inadequate teachers from entering the profession. For the system to work, the licensure exam should be highly correlated with the desired outcome—a teacher's performance in the classroom. If the two are perfectly correlated, the licensure exam would be a perfect screen. If, however, they are not perfectly correlated, the licensure exam will let some ineffective teachers into the classroom and keep out some highly effective teachers. In other words, licensure exams are unreliable screens if they are not highly correlated to a teacher's performance.

In Arkansas, there *is* a positive relationship between performance on licensure exams and performance in the classroom, but that correlation is relatively low, meaning that neither the Praxis I nor the Praxis II is a particularly effective licensure screen. In fact, at all points of the testing distribution, there are effective and ineffective teachers. Therefore, the practical use of raising licensure exam cut scores is called into question when we consider the trade-offs.

One important factor to consider is that low-scoring teachers are not randomly distributed among Arkansas schools. Teachers who did not perform well on the licensure exams are more likely to find employment in a more disadvantaged school. Presumably, these schools did not intentionally hire low-scoring teachers. More likely, they hired the teachers who were available. Although some of these low-scoring teachers will prove to be ineffective, some—luckily for these schools—will prove to be excellent in the classroom.

If Arkansas could raise the cut score on the licensure exams and replace the low-performing teachers who would have failed the exam under the new score with average-performing teachers, the state's schools overall and the disadvantaged schools in particular would be better off. However, it is more likely that increasing the cut scores would lead to teacher shortages in hard-to-staff subjects and regions of the state. Consequently, raising licensure exam cut scores may have a disproportionately negative impact on the most disadvantaged schools.

Though the current licensure exams are correlated to performance, there is tremendous variation in teacher quality among those who pass and those who fail the exams. It seems apparent that policies that attempt to determine who will be an effective teacher before the teacher enters the classroom are a less than ideal method of judging teacher quality. Therefore, if Arkansas wants to improve teacher quality through a front-end policy, simply raising the score needed to pass the licensure exams does not appear to be an effective strategy.

Hanushek (2009) suggests an alternative method for improving teacher quality: teacher deselection. Rather than attempt to prevent ineffective teachers from entering the classroom, Hanushek demonstrates how removing the teachers we know are ineffective would lead to improved teacher quality. As a simple illustration, look back at the distributions presented in

figures 2 and 3. This time, instead of drawing a vertical line, as is done with licensure exams and other front-end screens, draw a horizontal line across the bottom of the distribution. This represents the policy of "deselecting" the most ineffective teachers. Currently removing ineffective teachers is very difficult. But targeted removal of ineffective teachers would not only produce the desired result, but also improve the teacher workforce based on classroom data rather than potential.

There are no public policy magic bullets or simple panaceas that will ensure that all of Arkansas's teachers are effective in the classroom. However, improving the classroom performance of our teachers, on average, could be the key to sustaining the recent trend of improved education in the state. Rather than consume time, effort, and political will to pursue policy changes unlikely to increase teacher effectiveness, Arkansas's leaders and citizens would be better served by looking in new directions and searching for something more elusive: an innovative approach to improving teacher quality. Baltimore is experimenting with additional pay for productivity instead of using only credentials and seniority (Sawchuck 2014). Rhode Island is firing teachers who are ineffective for more than two years in a row, even if they have tenure. Colorado and Nevada passed laws removing tenure after multiple ineffective ratings (Bonner, 2012)Innovative approaches are already taking place in K-12 public education aimed at increasing competition among schools to improve outcomes. More than half of the states are already offering school choice programs (Friedman Foundation). Augmenting this with competition among teachers would improve the education outcomes. (Friedman Foundation).

References

Bonner, J.L. (2012, January 25). States Weaken Tenure Rights for Teachers. *The Associated Press*. Retrieved from http://usatoday30.usatoday.com/news/education/story/2012-01-25/teacher-tenure-rights-firings/52772354/1

Boyd, D., H. Lankford, S. Loeb, J. Rockoff, and J. Wychoff. 2008. "The Narrowing Gap in New York City Teacher Qualifications and Its Implications for Student Achievement in High Poverty Schools." *Journal of Policy Analysis and Management*, 27(4): 793–818.

Chetty, R., J. N. Friedman, and J. E. Rockoff. 2011. *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood*. National Bureau of Economic Research Working Paper No. 17699.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources*, 41(4): 778–820.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor. 2007. "How and Why Do Teacher Credentials Matter for Student Achievement?" National Center for Analysis of Longitudinal Data in Education Research Working Paper No. 2, http://www.caldercenter.org/sites/default/files/1001058_Teacher_Credentials.pdf.

Education Week. 2015. Quality Counts: National Summary Report. http://www.edweek.org/ew/qc/2015/state-highlights/2015/01/08/nation-earns-c-on-quality-counts-new-state-report-card.html Accessed May 1, 2016

Educaton Week. 2015a. State Report Cards: Arkansas. http://www.edweek.org/ew/toc/2015/01/08/index.html Acccessed April 14, 2016.

Ferguson, R. F., and H. F. Ladd. 1996. "How and Why Money Matters: An Analysis of Alabama Schools." In *Holding Schools Accountable: Performance-Based Reform in Education*, ed. H. F. Ladd. Washington, DC: The Brookings Institution.

Goldhaber, D. 2007. "Everyone's Doing It, but What Does Teacher Testing Tell Us about Teacher Effectiveness?" *Journal of Human Resources*, 42(4): 765–94.

Hanushek, E. A. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis*, 9(2): 141–64.

Hanushek, E.A., 2003. The Failure of Input-based Schooling Policies*. *The economic journal*, *113*(485), pp.F64-F98.Hanushek, E. A. 2009. "Teacher Deselection." In *Creating a New Teaching Profession*, ed. D. Goldhaber and J. Hannaway. Washington, DC: Urban Institute Press.

Hanushek, E. A., and S. G. Rivkin. 2006. "Teacher Quality." *Handbook of the Economics of Education*, 2, 1051–78.

Hanushek, E. A., L. Woessmann, and P. E. Peterson. 2012. "Is the US catching Up?" *Education Next*, 12(4): n4.

Hanushek, E. A., J. F. Kain, D. M. O'Brien, and S. G. Rivkin. 2005. "The Market for Teacher Quality." National Bureau of Economic Research Working Paper No. 11154, http://www.nber.org/papers/w11154.

Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *The future of children*, *17*(1), 87-109.

National Center for Education Statistics. 2013a. "How Are States Performing?" http://www.nationsreportcard.gov/reading_math_2013/#/state-performance.

National Center for Education Statistics. 2013b. "National Assessment of Educational Progress," http://nces.ed.gov/nationsreportcard/.

Sawchuck, S. 2014 “Baltimore's Teacher-Pay Experiment Gains Foothold” *Education Week* (Aug 25). Online: http://www.edweek.org/ew/articles/2014/08/27/02baltimore.h34.html?qs=achievement+gains.

Shuls, J. V., and J. R. Trivitt. 2013. “Teacher Effectiveness: An Analysis of Licensure Screens.” *Educational Policy*, December 30. http://epx.sagepub.com/content/early/2013/12/18/0895904813510777.abstract.

Appendix A. School-by-grade value-added coefficients for the Benchmark ELA exam

Student demographic	2007	2008
Math lag 1	0.093***	0.093***
	(0.002)	(0.002)
Math lag 2	0.064***	0.046***
	(0.002)	(0.002)
ELA lag 1	0.448***	0.455***
	(0.002)	(0.002)
ELA lag 2	0.301***	0.304***
	(0.002)	(0.002)
Standard deviation: random effect	0.111	0.181
	(0.003)	(0.005)
Number of observations	299,413	362,031

Source: Author's calculations with data from ADE

Appendix B. School-by-grade value-added coefficients for the Benchmark math exam

Student demographic	2007	2008
Math lag 1	0.409***	0.415***
	(0.002)	(0.002)
Math lag 2	0.290***	0.301***
	(0.002)	(0.002)
ELA lag 1	0.136***	0.136***
	(0.002)	(0.002)
ELA lag 2	0.071***	0.060***
	(0.002)	(0.002)
Standard deviation: random effect	0.154	0.268
	(0.004)	(0.007)
Number of observations	299,413	362,031

Source: Author's calculations with data from ADE

Appendix C. Praxis I and Praxis II pass/fail descriptive statistics (standard deviations in parentheses)

Variable	Praxis I (Arkansas cut score)		Praxis I (Louisiana cut score)		Praxis II (Arkansas cut score)		Praxis II (0.25 sd higher)	
	Pass	Fail	Pass	Fail	Pass	Fail	Pass	Fail
Percent with a master's degree	.375	.371	.378	.338***	.375	.376	.376	.364
	(.484)	(.484)	(.485)	(.473)	(.484)	(.484)	(.484)	(.481)
Percent with nontraditional licensure	.051	.027*	.040	.210***	.049	.088***	.041	.155***
	(.221)	(.164)	(.196)	(.407)	(.216)	(.283)	(.198)	(.362)
Percent white	.885	.701***	.887	.808***	.890	.725***	.890	.797***
	(.320)	(.459)	(.317)	(.394)	(.313)	(.446)	(.313)	(.403)
Percent female	.806	.814	.809	.768***	.808	.772***	.814	.735***
	(.395)	(.390)	(.393)	(.422)	(.394)	(.420)	(.389)	(.441)
Years of experience	12.4	4.0***	12.9	2.7***	12.6	5.9***	12.9	5.0***
	(9.8)	(2.3)	(9.7)	(2.9)	(9.8)	(6.5)	(9.8)	(6.1)
Percentage of students eligible for free or reduced-price lunches	.581	.617***	.580	.602***	.580	.615***	.579	.609***
	(.143)	(.151)	(.143)	(.146)	(.143)	(.153)	(.143)	(.148)
Percentage of minority students	.320	.391***	.319	.360***	.316	.417***	.316	.381***
	(.280)	(.299)	(.279)	(.294)	(.278)	(.306)	(.277)	(.303)
District enrollment	5,335	5,301	5,323	5,501	5,297	6,043	5,299	5,705
	(6,877)	(6,995)	(6,855)	(7,226)	(6,832)	(7,675)	(6,801)	(7,630)
Sample size	17,336	291	16,477	1,150	16,738	889	16,023	1,604

Note: *** p<0.01, ** p<0.05, * p<0.1
Source: Data from Arkansas Department of Education

[1] The K–12 Achievement rankings were not updated from the 2012 Quality Counts report.
[2] These analyses examine performance on a licensure exam as a continuous measure.
[3] Regression analysis is one of the most commonly used analytical techniques in education, social sciences, economics, and business research.
[4] That is, a positive and statistically significant coefficient.
[5] Keep in mind that this is just for the sample of teachers included in the current analyses; the actual number of teachers that would be removed from the entire workforce would be much higher.
[6] Both of these measures are standardized with a mean of zero and a standard deviation of one.