B | Economic Studies at BROOKINGS

Evidence Speaks Reports, Vol 2, #52
May 17, 2018

Did Tenure Reform in Florida Affect Student Test Scores?

Celeste Carruthers, David Figlio, & Tim Sass

Executive Summary

We examine the elimination of teacher tenure in Florida to better understand how job protections for teachers affect students. We focus on identifying the effect of removing tenure as part of Florida's 2011 Student Success Act on student outcomes, and in particular the effect if this legislation on student achievement.

We take a quasi-experimental research approach by comparing the effect of teachers' relative exposure to tenure reform on changes in individual students' test scores after versus before the policy. We find limited and circumstantial evidence that Florida's tenure reform *slightly* increased student test achievement in math and reading, and that the gains were more prominent for the lowest-performing students. Specifically, we look across schools where faculties were most versus least vulnerable to Student Success Act provisions – measured in two different ways – and find that students in the most vulnerable schools show gains that compared favorably to students in the least vulnerable schools, but by a very small degree.

Introduction

Employment protections for teachers, or "tenure," were first introduced in New Jersey in 1909 and by the mid-1940s, 70 percent of public school teachers in the U.S. enjoyed tenure protections. The coverage continued to grow over time, with 80 percent of teachers having tenure protections by the mid-1950s. By the 1960s, some degree of job protection was nearly universal, with 37 states and the District of Columbia providing tenure for teachers and the other 13 states either providing some form of long-term contract or allowing districts to automatically renew contracts with teachers.[1]

Under the tenure system in the U.S., new teachers go through a probationary period in which they work under annual contracts and can be terminated or simply not have their contract renewed at the discretion of the school district employing them. The degree of latitude given to districts during this probationary period, however, may be constrained by provisions in collective bargaining agreements. The probationary period can last from one to five years, with a three-year period being most common.[2] After the probationary period is completed, teachers receive "due process" rights that require employers to show "just cause" for dismissal and give teachers the right to have a dismissal dispute decided by an impartial body. Exactly what constitutes just cause can vary across states. While tenure laws do not prohibit the firing of teachers, the costs of demonstrating just cause and going through due process procedures result in very few teachers being dismissed after the initial probationary period.

The original rationale for teacher tenure was to curb political influence in teacher employment decisions and limit political patronage. It was later viewed as a means of protecting the civil rights of teachers, including opponents to U.S. involvement in World War I and supporters of school desegregation in the 1960s.[3]

From a labor market standpoint, the increased job security associated with teacher tenure can have two opposing effects. First, by reducing the risk of termination, it will make the job of teaching more attractive and should increase the quality of candidates who are willing to become teachers at any given wage. Second, by lessening the chance of dismissal, it reduces the incentive of teachers to maximize their effort, which in turn would reduce their productivity. Similarly, if the likelihood of being dismissed for poor performance is reduced by tenure protections, less qualified candidates may be attracted to the teaching profession.

As part of a broader school reform movement, states began to re-assess their tenure laws around the start of the 21st century. Georgia dropped due-process rights for teachers hired after July 1, 2000, but reversed course three

years later. Idaho also had a short-lived departure from tenure, with the legislature abolishing teacher tenure in 2011 only to have the decision reversed by voters the next year. Three other states have had more enduring repeals of tenure. Florida abolished tenure protections for new teachers beginning in July 2011 (more on this below). In 2013, North Carolina passed a law which phases out tenure over a five-year period. In 2014, Kansas effectively repealed teacher tenure by narrowing the definition a teacher, thereby removing due process procedures for classroom teachers.[4,5]

While outright repeals of teacher tenure have been limited to a handful of states, many states have limited teacher tenure protections in recent years by expanding the length of the probationary period or introducing performance requirements for the receipt of tenure. Between 2009 and 2012, the number of states using student achievement as a criterion for tenure rose five-fold, from four states to 20.[6] Similarly, from 2011-2014 eight states increased the length of the probationary period before teachers can receive tenure.[7]

Previous Research

Despite the recent legislative activity around teacher tenure, there is relatively little empirical evidence on the effects of teacher tenure, either on the labor market decisions of prospective teachers, the productivity of teachers in the classroom or most importantly, the impact on educational outcomes for students.[8] What little evidence currently exists is largely based on modifications to existing tenure systems rather than instances where tenure has been eliminated.

Jacob studied the impact of loosening the constraints for dismissing probationary teachers in Chicago.[9] In 2004 the Chicago Public Schools agreed to a new collective bargaining contract with the Chicago Teachers Union that allowed principals to dismiss probationary teachers for any reason and without the documentation and due-process hearings typically required for teacher dismissals. Jacob found the elimination of job protections for early-career teachers reduced teacher absences by about 10 percent and lowered the incidence of frequent absences by 25 percent. Most of the observed change in teacher absenteeism was a result of changes in the composition of teachers, but there was also evidence of modest incentive effects for probationary teachers.

Similar to Jacob, Loeb, Miller and Wyckoff analyzed a change in an existing tenure system, rather than a wholesale elimination of tenure. In New York City, principals are required to make tenure recommendations to the superintendent, and historically, nearly all eligible teachers (94 percent) were approved for tenure. Beginning in 2009–2010, New York City increased the information available to principals on the

performance of their probationary teachers and simultaneously required them to provide justification for granting tenure, extending a teacher's probationary period, or denying tenure. The district also gave principals explicit guidance for teachers whose measured performance was particularly strong or weak. While the changes did little to increase the proportion of teachers who were denied tenure, they did lead to a dramatic increase in the fraction of teachers who had their probationary periods extended, particularly among low-performing and less qualified teachers. The policy change also led to increases in voluntary attrition for teachers whose probationary periods were extended and for the small share of teachers who were denied tenure. Among extended teachers, those with lower principal ratings were more likely to leave. Extended teachers who chose to leave their schools were less effective, as measured by principal ratings and value-added estimates, than new teachers who were likely to replace them.

While the evidence presented by Loeb, Miller, and Wyckoff suggests that with the right incentives and information principals could use pre-tenure flexibility to filter or counsel out low-performing teachers that may not occur in all circumstances. In North Carolina, Chingos found that "principals are not using the four-year [pre-tenure] period to identify and remove their lowest performers."[10] Following Louisiana's tenure reform of 2012, which required that teachers be rated "highly effective" in five out of six years to gain tenure, Strunk, Barrett, and Lincove found that departures rose among teachers eligible for retirement and among teachers in the lowest-performing schools, but it is not clear if or how these exits changed teacher quality across the state.[11]

We contribute to this literature by studying the short-run consequences of Florida's experience with tenure reform in 2011 -- the "Student Success Act" (SB 736, or henceforth, SSA), focusing on whether and where student achievement changed in the years immediately after 2011.

How we determine whether SSA affected student outcomes

It is difficult to identify how a policy like SSA shapes individual student achievement. Ideally, researchers would like to compare students whose teachers were affected by SSA to unaffected students who were otherwise very similar, and who made it into the "unaffected" group for reasons unrelated to student achievement. But SSA was a statewide policy, and it affected all students and all teachers at the same time. Looking at statewide achievement trends is not necessarily the right approach either. Average achievement across a state moves slowly. It reflects the accumulation of several years' teaching, as well as student populations that change from year to year, as

students move in and out of the state and the public schools. So, we take a quasi-experimental approach to see if students whose teachers were *more* affected by SSA had higher, lower, or about the same achievement growth as students whose teachers were *less* affected by SSA.

Which students were *more* affected by SSA? Certainly those with teachers hired after July 1, 2011. But it is problematic to compare students with new teachers to those with more experienced teachers because experience itself is a factor in student achievement, and because student-teacher assignments are not random. Principals hoping to protect new teachers from the law's effects could have assigned them more high-achieving students. Another group of students who were more directly affected are those who were in subjects and grades with a history of standardized testing. It was easier to implement SSA provisions for teachers in these tested subjects.[12] But school leaders can and do change teacher assignments across grades and subjects from year to year, and it's possible that they did so in response to SSA.

That leaves us with comparing test scores of students in *schools* where teachers faced different levels of exposure to the policy change. Schools and districts have much less flexibility in strategizing teacher-school assignments than teacher-student or teacher-class assignments.

We consider two different ways of measuring this differential exposure across schools. One approach is to compare schools that historically – prior to SSA – had more first-year, rookie teachers to schools that historically had fewer rookie teachers. We expect that teachers in schools with more new faculty hired under the SSA regime would be more sensitive to the policy than teachers in schools with fewer rookies. Our second approach is to compare schools where relatively more versus relatively fewer teachers were evaluated – regardless of when they were hired – in the first year of the tenure reform policy. Schools varied in the fraction of teachers they could formally evaluate. In 2011-12, half of the state's schools evaluated somewhere between 88 and 98% of teachers, and one in four evaluated less than 88%. We expect that teachers in schools where more were evaluated would be more sensitive to a policy based on teacher evaluations.

In order to observe what happened to student test scores following the introduction of the tenure reform policy, we analyzed achievement data from every student attending Florida public schools between 2007-08 and 2012-13. This includes three school years before SSA and two after. We can track achievement for the same student over time as he or she transitions from one tenure policy regime to the next. This longitudinal data structure allows us to control for what education researchers

call “student fixed effects,” that is, factors that affect student learning every year regardless of whose class they attend. We combine each student’s reading and math performance on the Florida Comprehensive Assessment Test (FCAT), which was Florida’s annual test administered to third through tenth graders during this window of time. To help us interpret the size of results, we transform FCAT scores into standardized units, where the average score is zero and the standard deviation is equal to one. To fix ideas about how big a standard deviation is, about 68% of students will be within one standard deviation of the average, and 95% will be within two. Or for a policy context, note that when schools face the threat of receiving an “F” on a state report card, researchers find that student achievement rises by 6-7% of a standard deviation compared to schools not facing such a threat.[13]

Estimated average effects of the policy change

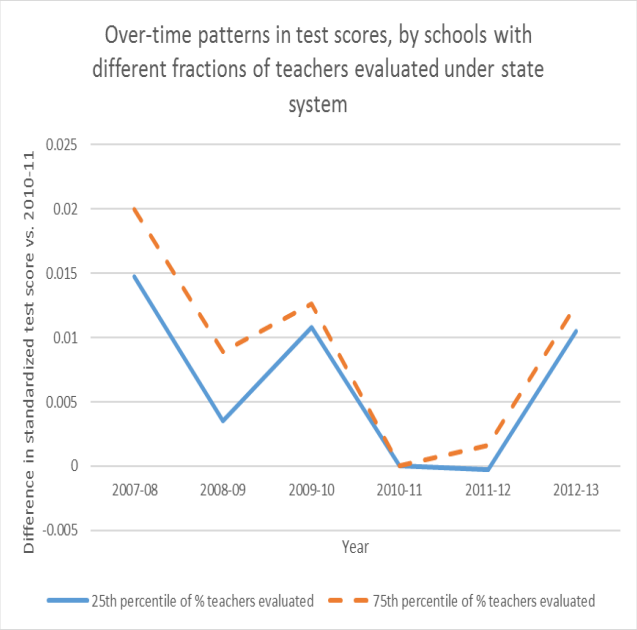
The easiest way to observe the potential effects of the policy change is to compare the over-time achievement growth of individual students attending more affected versus less affected schools. For each of the two different exposure measures, we compare student performance at schools above the 75th percentile of exposure to those at the 25th percentile of exposure. We present graphs that take 2010-11 – the year immediately prior to SSA – as our base year, allowing us to compare the years before versus after tenure reform.

First, let’s compare student performance in schools with relatively high versus relatively low rates of teacher evaluation in the first year of the SSA regime. The 25th percentile school had 88% evaluation rates in 2011-12, while the 75th percentile school had 98% evaluation rates.

As can be seen in the graph below, overall student performance – conditional on student fixed effects – was declining in the years prior to the policy change and rebounded in the years following the policy change. Note that the pre-post difference measured just 1-2% of a standard deviation, and this would be hard to detect from trends in statewide average achievement. The rebound in test scores following the policy change may be due to SSA but may also be due to unobserved factors from around the same time, or to “regression to the mean,” the tendency for extreme values to shift toward the long-run average with time.

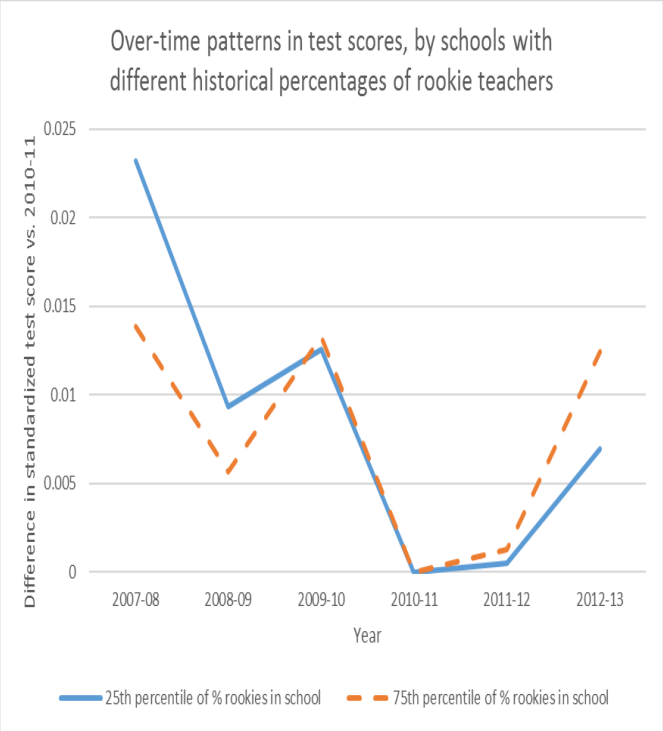
Our quasi-experimental approach is to infer the effect of SSA from changes in the difference between students in more versus less exposed schools. In the years following the policy change, the dashed line is above the solid line, meaning that students in schools with more evaluated teachers did relatively better on the state tests than students

in schools with fewer evaluated teachers. These differences are statistically distinct from zero at conventional levels in both 2011-12 and 2012-13[14], but the differences are very small – less than one-half of one percent of a standard deviation increase in test scores.[15] And the pre-reform gaps tended to be at least as large as the post-reform gaps. Students in schools that were more exposed to SSA in terms of teacher evaluation rates were already outperforming less exposed students prior to 2011, and they did not extend that lead afterward.



As an alternative, let's compare student performance in schools with historically high versus low rates of rookie teachers. We measure this based on the average percentage of first-year teachers in a school from 2007-08 through 2009-10. The 25th percentile school had 3.2 percent rookie teachers and the 75th percentile school had 7.6 percent rookie teachers during this time period.

Results are in the graph below. As with the first measure of exposure, student test scores in both the more exposed schools and the less exposed schools had been declining prior to SSA and rebounded in the two years following SSA. But unlike the fraction-evaluated measure, more exposed schools went from lagging to leading the less exposed schools. That is, students in schools with relatively fewer first-year teachers in 2007-08 and 2008-09 outperformed their peers in schools with less experienced teachers, but the opposite was true after SSA. While the post-SSA difference is statistically distinct from zero at conventional levels in 2012-13 (though not in 2011-12), the difference between the two types of schools is also quite small – around one-half of one percent of a standard deviation differential improvement for the relatively affected set of schools. And we can't rule out the possibility that the two types of schools were just trending differently over this time period for reasons unrelated to tenure reform.



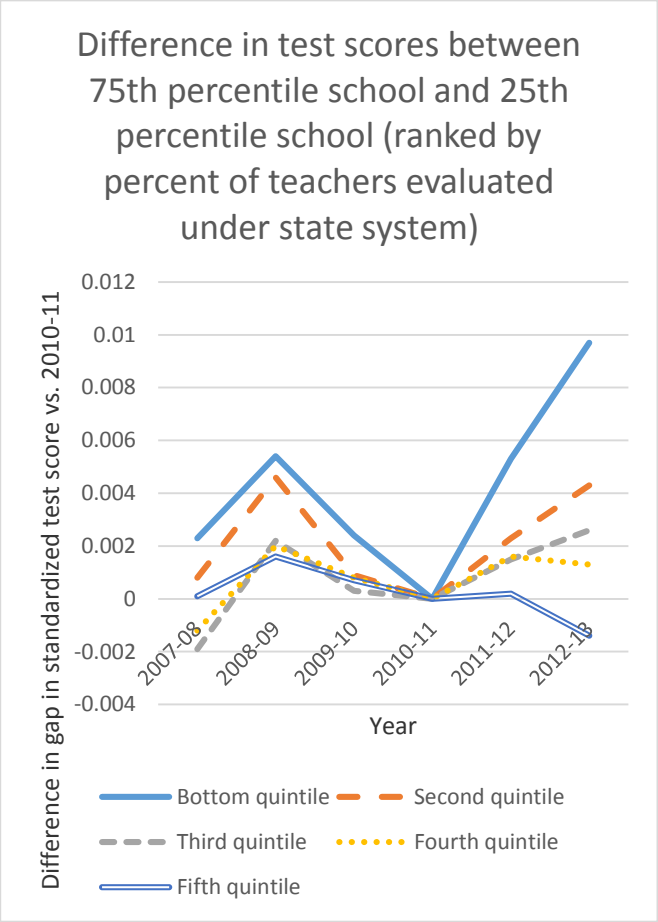
Differences by Student Performance Level

Results depicted in the first two figures are average estimated effects of tenure reform in Florida, and they tend to suggest that students in more affected schools developed at about the same or a slightly better rate after the SSA transition than students in less affected schools. But these reflect trends for the average student in each type of school, and tenure reform might have had different effects on different sets of students. For example, even though the state's evaluation metric for teachers of courses with statewide assessments is based on their contribution to student achievement gains, or value-added,[16] school and district leaders might focus their efforts on raising the achievement of relatively low-performing students – and so may teachers. In testimony and debate leading up to Governor Scott's signature, SSA opponents argued that the law would penalize teachers whose students face myriad challenges outside of school.[17] Also relevant is Florida's school accountability system, which grades schools on average performance levels, average performance growth, and also on the performance growth of the lowest-scoring students in the school.

With this in mind, we conduct the same analysis within five equally-sized groups of students (quintiles), divided according to their first, pre-SSA test score. For each student quintile, we start by assessing the difference in student achievement growth across schools with a high versus a low fraction of teachers evaluated under SSA. The graph below shows the gap in test scores between 75th percentile and 25th percentile schools, now broken out into five quintiles of initial student achievement.

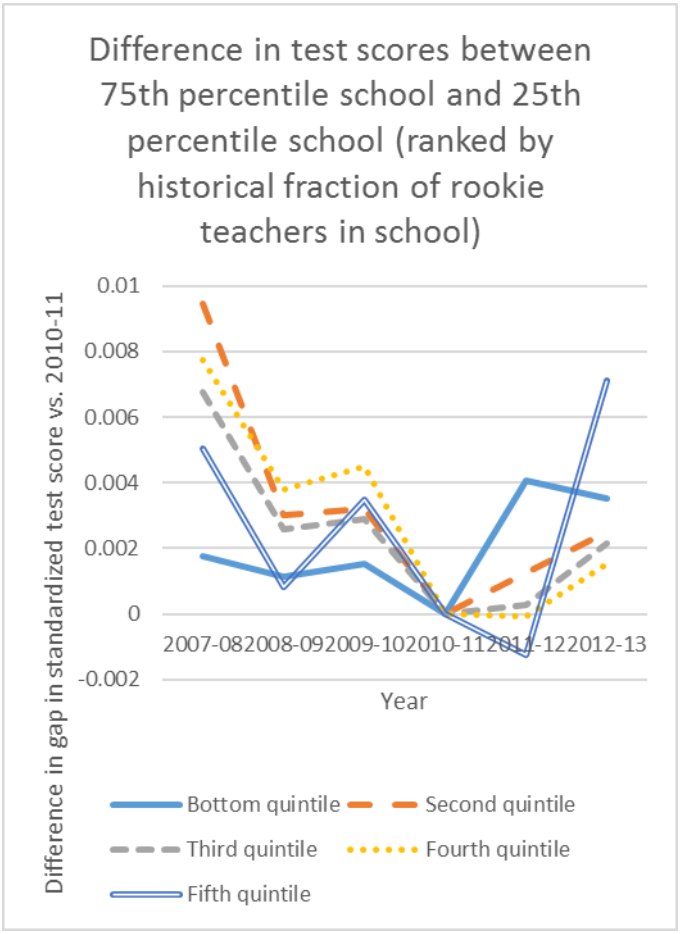
Prior to SSA, the lowest-achieving students (i.e., the bottom quintile of students based on initial test performance) differed by about 0.2-0.6% of a standard deviation across schools with a high versus a low share of teachers subject to evaluation. Second-quintile students differed by at most 0.4% of a standard deviation, and higher-achieving students differed by no more than 0.2% of a standard deviation. Following tenure reform, these test performance gaps fan out, with the increase in test scores between most-evaluated and least-evaluated schools

being the largest for the bottom quintile of students, followed by the second quintile, followed by the third quintile, and so on. These relative changes in achievement are significantly different from zero at conventional levels for the bottom two quintiles, and not for higher-achieving students. Note, however, that even for the bottom quintile of students, the evident gain in achievement is not large – less than one percent of a standard deviation. That said, this pattern of findings makes us somewhat more comfortable that the post-tenure reform results may be genuinely due to tenure reform and not either "regression to the mean" or pre-existing trends, as may (reasonably) be suspected given the first pair of results we describe above.



The same analysis based on historical shares of rookie teachers yields many similar findings with two notable differences. First, the similarities: The gap in test scores between schools with a high share of rookies (75th percentile) and a low share of rookies (25th percentile) trended similarly across all five groups of students in the pre-SSA period. And following tenure reform, test score gaps for the bottom four quintiles groups fanned out in such a way suggesting that lower-achieving students gained ground in schools where SSA would have had more bite. Now, the differences. Prior to SSA, the group of lowest-achieving students were the most – not the least

– similar across schools that were more versus less exposed to SSA requirements for novice teachers. Visually, the pre-SSA pattern of gaps for low-performing students was flatter than for the other groups of students (rather than steeper, as in the other analysis). And, even more noteworthy, the largest gains after tenure reform accrued to the *highest*-performing students (in 2012-13 only). For 2012-13, our estimate of the dosage effect of SSA is positive and statistically distinct from zero for the lowest-performing group of students as well as the highest-performing group of students. So, while the general patterns of findings are consistent across the two approaches, the two sets of findings are not identical.



Summary and Implications

Florida's "game changing" tenure reform law of 2011 did not precede a large change in statewide student achievement,[18] but under a quasi-experimental microscope, we find limited and circumstantial evidence that SSA slightly increased student test achievement in math and reading, and that the gains were more prominent for the lowest-performing students. Our approach of comparing relatively vulnerable schools to relatively less vulnerable schools means that we aren't able to investigate the overall effects of SSA, which might have been larger. And our being limited to

studying only the first two years following SSA means that we aren't able to observe any changes – positive or negative – that happened to teacher recruitment in the wake of SSA.

These findings must be interpreted with substantial caution. The pattern of test scores across more and less exposed schools in the years prior to tenure reform suggest that these schools were following different trajectories, so the less exposed schools might not be an ideal control group in our research design. To the point, this reduces our confidence that changes – even small changes – in student achievement where SSA was more potent are attributable to that policy. On the other hand, because SSA applied to the whole state at once, we rely on shallow gradients of tenure reform exposure (88 versus 98% of teachers evaluated; 3.2 versus 9.6 rookie teachers on the faculty) to identify its effects, so the full consequences of tenure reform could be larger and more robust than the very modest estimates we uncover herein. Note also that we examine just one likely outcome of tenure reform. Returning to the language of the statute, the intent was to raise student achievement *by improving the quality of instructional, administrative, and supervisory services in the public schools*. Whether SSA or policies like it succeed in attracting and retaining high quality teachers remains an open question.

The question of whether SSA worked as intended may ultimately be a political one. "Regression to the mean" is a statistical concept that we worry about when interpreting results, but it is also a fair description of what has happened to SSA since 2011. The lofty goals of tying teacher evaluation, retention, and pay to student achievement growth have since faded to a considerable degree. Requirements to test all subjects and grades were dropped in 2015, student performance took on a minority role in teacher evaluations the same year, and requirements to use VAM in district evaluation systems were also short-lived. Teacher evaluations went from 99.7% satisfactory prior to SSA to 97.2% effective afterwards. If effects from the strongest and earliest years of tenure reform in Florida were modest and limited to particular groups of students, we might expect effects to narrow further as the scope of the law diminishes.

That said, we end with a note of optimism regarding the policy. Florida's performance on the NAEP was very impressive during the time period following SSA. Especially since we are only able to compare between schools with different degrees of exposure to the policy, it may be the case that SSA was still a substantial contributor to that performance, even though the effects we are able to uncover were relatively modest.

Appendix: Background on Tenure Reform in Florida

Florida first passed legislation to eliminate tenure for newly hired teachers during the 2010 legislative session. The legislation, Senate Bill 6, was vetoed by Governor Crist on April 15, 2010. After significant revision, a new version of the bill was introduced in the 2011 legislative session. On March 24, 2011 Florida Governor Scott signed into law the "Student Success Act" (SB 736, or henceforth, SSA). In addition to overhauling teacher evaluation and pay systems, the law eliminated tenure for teachers hired on or after July 1, 2011.

Prior to 2011, new teachers were given probationary one-year contracts. During their first 97 days on the job, they could be fired without cause. After that, districts had to show cause to terminate a teacher. If a teacher received satisfactory evaluations in his or her first three years, they could be given a "professional services contract" that granted them extensive job protections. Since unsatisfactory evaluations were extremely rare (99.7% of teachers received a "satisfactory" evaluation in 2009), this system effectively granted tenure to almost all teachers after three years of service.

Enter SSA. New teachers hired on or after July 1, 2011 are given a probationary one-year contract. Teachers are evaluated at least twice during their first year and may be dismissed without cause during that time. Based on these evaluations, superintendents can recommend that a teacher be given an annual contract for the next year. During the term of an annual contract the teacher can only be dismissed for cause. Renewal of a teacher's annual contract is at the discretion of the school district and is dependent on annual performance evaluations. Critically, at least half of a teacher's evaluation must be based on three years of standardized test outcomes for that teacher's students. Beginning in 2012 (and required in 2015), student outcomes were derived from "value added modeling" (VAM), a complicated statistical technique designed to identify the rate at which a student's knowledge grows in a year, and how much his or her teacher contributed to that growth. The other half of a teacher's evaluation is based on "instructional practice" measures that typically include classroom observations.

Teacher evaluations have five possible conclusions: "highly effective," "effective," "needs improvement", "unsatisfactory," and "developing." The last category is reserved for teachers without three prior years of student outcomes. A teacher's annual contract is at risk if the teacher (1) receives two consecutive "unsatisfactory" ratings, (2) receives two "unsatisfactory" ratings within three years, or (3) receives three consecutive annual

ratings of "needs improvement" or "unsatisfactory."

SSA changed how Florida teachers are paid, shifting from a traditional "step-and-lane" pay scale that rewards experience and advanced degrees to a performance-based scale with bonuses based on the new evaluation system. Hard-to-staff subjects like math and science pay more. Teachers hired before July 1, 2011 could remain under the old contract system, or they could switch to a performance-based pay system with renewable annual contracts.

SSA also dramatically changed the rules regarding teacher layoffs. Previously, layoffs were governed by collective bargaining agreements which typically required that the "last hired" were the "first fired" when shrinking enrollments or budgets required cutbacks. Under the new statute, however, districts are required to retain teachers based on educational program needs and the teachers' performance evaluations. Teachers with low performance evaluations would be the first to be laid off. Retention based on seniority is specifically prohibited.

[1] "A timeline of teacher tenure." 2016. *The Phi Delta Kappan*, vol. 97, no. 6, p. 12.

[2] Goldhaber, Dan and Joe Walch. 2016. "Teacher Tenure: Fog Warning," *The Phi Delta Kappan*, vol. 97, no. 6, pp. 8-15.

[3] Kahlenberg, Richard D. 2016. "Teacher Tenure Has a Long History and, Hopefully, a Future", *The Phi Delta Kappan*, vol. 97, no. 6, p. 16-21.

[4] Loeb, Susanna, Luke C. Miller and James Wyckoff. 2015. "Performance Screens for School Improvement: The Case of Teacher Tenure Reform in New York City," *Educational Researcher*, vol. 44, no. 4, pp. 199-212.

[5] Thomsen, Jennifer. 2014. *50 State Comparison: Teacher Tenure/Continuing Contract Policies*. Denver, CO: Education Commission of the States.

[6] Loeb, Susanna, Luke C. Miller and James Wyckoff. 2015. "Performance Screens for School Improvement: The Case of Teacher Tenure Reform in New York City," *Educational Researcher*, vol. 44, no. 4, pp. 199-212.

[7] Goldhaber, Dan and Joe Walch. 2016. "Teacher Tenure: Fog Warning," *The Phi Delta Kappan*, vol. 97, no. 6, pp. 8-15.

[8] There are a number of recent studies that indirectly estimate the potential impact of tenure reform by simulating the effects on the distribution of teacher quality and student achievement if a given proportion of low-quality teachers could be dismissed. These studies are hypothetical and do not measure the actual effects of changes in tenure laws.

[9] Jacob, Brian A. 2013. "The Effect of Employment Protection on Teacher Effort," *Journal of Labor Economics*, vol. 31, no. 4, pp. 727-761.

[10] Chingos, Matthew M. 2014. "Ending Teacher Tenure Would Have Little Impact on its Own."

[11] Strunk, Katherine O., Nathan Barrett, and Jane Arnold Lincove. 2017. "When Tenure Ends: Teacher Turnover in Response to Policy Changes in Louisiana." Education Research Alliance for New Orleans.

[12] Teachers in non-tested subjects (such as art or kindergarten) were evaluated according to student achievement in other classes.

[13] Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David Figlio. 2013. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure," *American Economic Journal: Economic Policy*, vol. 5, no. 2, pp. 251–281.

[14] As a technical point, while we make use of individual student data, we recognize that the variation in the policy effects that we are measuring comes at the school level. Therefore, we adjust the standard errors – which are interpreted in conjunction with the point estimates presented in the graph – for clustering at the school level, to account for this important point.

[15] To gauge the magnitude of this, note that the standard deviation of each 800-point section of the SAT is 100 points, so a one-half of one percent of a standard deviation change in an SAT score is equivalent to one-half of one point, out of 800 points, on each section of the SAT. Or in terms of other aspects of Florida education policy, one-half of one percent of a standard deviation is about 6-8% of the estimated effect of a school being threatened with an "F" grade.

[16] http://www.fldoe.org/teaching/performance-evaluation/instructional.stml

[17] Harrison, C. & Cohen-Vogel, L. (2012) "The politics of teacher reform in Florida: Analyzing causal narratives surrounding state adoption of performance-based evaluations, performance pay, and tenure elimination". *Leadership and Policy in Schools*, *12* (2), 122-145.

[18] http://www.orlandosentinel.com/features/education/school-zone/os-orange-teachers-pay-students-evaluations-20170213-story.html