

The School Leaders Licensure Assessment in Tennessee: Distribution, Passage Rates, and Usefulness as a Predictor of Principal and School Outcomes

*A Policy Brief on Strengthening Tennessee's
Education Labor Market*

Jason A. Grissom

About this Brief

This policy brief is an examination of Tennessee's usage of the School Leaders Licensure Assessment (SLLA) as a part of the principal licensure process. Tennessee is among 18 U.S. states and territories that rely on the SLLA when licensing principals. Researchers analyzed data for 10 years of Tennessee SLLA test takers, including their performance evaluations, student achievement in their schools, and their teachers' survey ratings of school leadership.

MARCH 2017

Executive Summary

Since 2010, Tennessee has required principals seeking administrative licensure to achieve a score of 160 on the School Leaders Licensure Assessment (SLLA), a standardized assessment administered by the Educational Testing Service (from 2004 to 2009, the cut score was 156 on a prior form of the test). We matched SLLA scores to state administrative records to examine the distribution of scores and passage rates by principal and location characteristics and to test whether the SLLA predicts policy-relevant outcomes (e.g., school composite TVAAS scores) for novice principals.

This analysis uncovers two important findings:

First, failure rates are very different by the race/ethnicity of the candidate. Under the current form of the test and cut score of 160, the failure rate for licensure candidates in Tennessee is **14.2%**. The failure rate for whites is **10%**, whereas the failure rate for nonwhites is **29%**. Controlling for other factors, a nonwhite test taker's odds of failure are **4.3 times greater** than those of a white test taker. Given the current distribution of scores, raising the cut score to 163 (i.e., on par with Missouri) would raise the overall failure rate to approximately 21%; for whites, the new failure rate would be 15%, compared to 40% for nonwhites. In short, the SLLA appears to be a significantly higher barrier to administrator licensure for nonwhites.

Second, there is little evidence that principals' pre-service SLLA scores predict available measures of principal performance for novice principals (i.e., those in their first 3 years as a principal). SLLA scores show small and statistically insignificant associations with principals' average scores on the qualitative (i.e., non-test score-based) portions of TEAM for either 2012 or 2013 once school and other characteristics are controlled for. SLLA scores are similarly statistically uncorrelated with average ratings from the School Leadership module provided by teachers on the TELL-TN survey in 2013. Single-year TVAAS school composite scores (currently only available for 2013) generally also are uncorrelated once other factors are accounted for; if anything, the association between SLLA scores and TVAAS school composites for novice principals is negative. As an additional check, we ran student-level growth models using TCAP scores in math and reading, controlling for student and school characteristics; these models also showed no meaningful relationship between novice principals' SLLA scores and student achievement.

Taken together, these results suggest few clear benefits of the SLLA as a condition of principal licensure but potentially large drawbacks if Tennessee has a goal of increasing racial diversity in its administrator workforce.

Background

Principals in 18 states, including Tennessee, take the SLLA as part of the administrator licensure process. Tennessee has required the SLLA since 2004. From 2004 to 2009, principal candidates were required to achieve a score of 156. In 2010, the Educational Testing Service introduced a new version of the test (form 1011). This introduction coincided with a raising of the required cut score to 160. In 2012, the SLLA was migrated to an online test administration (form 6011). Note that only one other state we identified (Kentucky) has a cut score as low as Tennessee's.

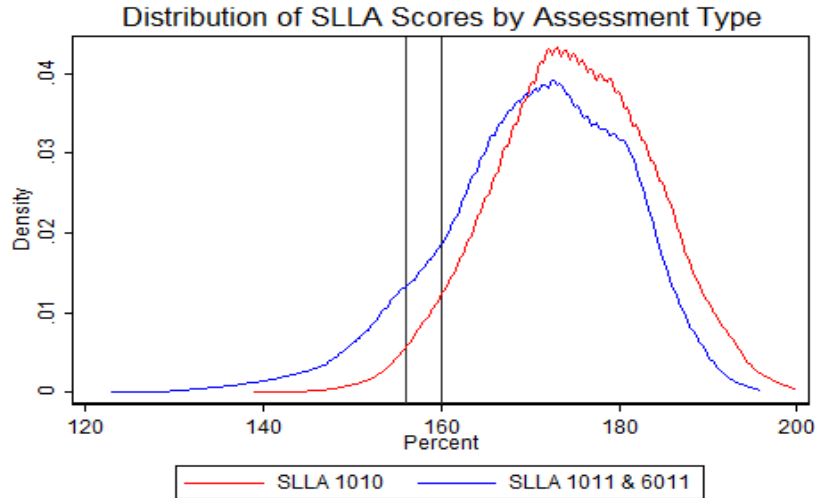
Data

With assistance from TDOE, TNCRED requested complete SLLA score histories for all SLLA test takers with a Tennessee affiliation from 2004 to the spring of 2014. Complete score histories means all scores, not just the passing scores. Score history files were matched by TNCRED research assistants to restricted personnel files (PIRS/EIS/TEAM) and school information for each year. Once scores were matched to a candidate/principal and school, they were then linked to school mean scores from the 2013 TELL-TN teacher survey, a school composite TVAAS file for 2013 provided by TDOE, and student-level TCAP math and reading files for 2008 to 2013.

Research Questions

Our analysis covers 3 research questions:

1. What is the distribution of SLLA scores for all prospective principals who take the test in TN?
2. What characteristics are associated with higher SLLA scores or differential passage rates?
3. What evidence do we see for the predictive validity of the SLLA? That is, how do principals' scores on SLLA correlate with other outcomes we care about, particularly early in the principal career?



The Distribution of Scores

The pre-2010 version of the SLLA (form 1010) was substantially easier than the version used from 2010 to the present. The mean score under the prior version was 175 (SD = 9) but only 170 (SD = 10) under the new version (form 1011/6011). This graph shows the distribution of Tennessee SLLA scores for all test takers by test version. The two vertical lines show the cut scores for the two versions of the test (156 for form 1010 and 160 for form 1011/6011).

The increased difficulty for the new version of the test and the higher cut score have resulted in substantially higher SLLA failure rates since 2010. Prior to 2010, only 1.4% of test takers failed to achieve the mandated cut score, compared to 14.2% from 2010 to 2014.

Factors that Predict SLLA Scores and Passage Rates

The distribution of scores varies by test taker characteristics. In particular, across the three test forms, females score significantly higher than males (173.7 to 170.6, or about 3 points), on average, and younger test takers score higher than older ones. Perhaps most noticeably, white test takers have much higher average scores than nonwhites (174.0 to 168.2, approximately a 6-point differential). This difference translates into a much higher probability of failure for

nonwhites (15%) than for whites (5%);¹ on the more recent forms of the test with the higher cut score, the failure probability for nonwhites is 29%, compared to only 10% for whites. When we run a regression that predicts the probability of failure as a function of test taker characteristics and characteristics of the school he or she worked in (usually as a teacher) at the time the test was taken, we find that the odds of failure for nonwhites are approximately four times higher than for whites with otherwise similar characteristics. The SLLA appears to be a much higher barrier to licensure for nonwhite principals than for white ones, a potentially important finding in light of Tennessee's 82% white principal workforce.

We also find differences in failure rates by region of the state. Looking across test forms, failure rates are highest in Southwest/Memphis-Shelby CORE region (12%) and Upper Cumberland (11%) and lowest in First Tennessee (5%), East Tennessee (6%), and Mid Cumberland (6%). In a regression format that controls for candidate and school characteristics, however, most regions look similar to one another, with the exception of the Mid Cumberland region, which still has significantly lower failure rates.

One possible explanation for these regional differences is differences in the administrator preparation programs that principals can access around the state. Indeed, some preparation programs have very low failure rates, and

¹ Given the very small numbers of nonwhite, nonblack test takers, we combined all nonwhite test takers into a single group.

others have very high ones. Limiting ourselves to programs with at least 30 test takers over the years of the data, we find failure rates on the new form of the test (1011/6011) as low as 3% for programs such as Lipscomb or Cumberland University but as high as 36% for Freed-Hardeman or 55% for Cambridge College–Memphis.²

Does SLLA Predict Principal or School Performance Outcomes?

Our final set of analyses test whether principals' SLLA scores predict job performance outcomes. Operating under the assumption that SLLA scores are likely to be most useful as a predictor of principal performance early in the principal's career, we focus on first-year principals and those with three years of experience or less, though results are similar if we combine all principals with SLLA scores.

We take two approaches. The first we call a *screening analysis*, which looks for evidence that test-takers who do not pass the SLLA cut score have lower job performance than those who do pass. Of course, we do not observe future job performance of candidates who fail and never are licensed to be school leaders. Instead, we take advantage of the fact that Tennessee has low failure rates (due to its low cut score) and compare the performance of test-takers who passed through the SLLA screen in Tennessee but would have failed at higher cut scores used in neighboring states. The second is a *signaling analysis*, which tests whether higher scorers have more positive outcomes. In both cases, we run regressions of different outcome measures on SLLA scores (or indicators for passage) with different sets of control variables, including characteristics of the schools the principals led (e.g., student demographics, percent free/reduced lunch eligible, school size, school level), other principal characteristics (e.g., race, gender, education level, years of experience as an educator), and either district characteristics or district fixed effects to control for district context. Extensive control variables are important for avoiding the mistake of attributing performance associated

with a high SLLA score to other factors, such as the sorting of principals with high SLLA scores into high-performing schools.

TEAM. First, we examine whether SLLA scores are associated with higher ratings of principals on the qualitative domains of the TEAM administrator evaluation. We examined ratings from 2011-12 to 2013-14 separately and pooled together. In other analysis, we show that principals' scores across domains on the TEAM rubric in each year are so highly inter-correlated that they can be reduced to one summative scale score, so we use the average rating in the analysis.

Across models, we find that, once we condition on school and principal characteristics, there is no clear statistical association between SLLA passage or score and principals' TEAM ratings. The point estimates are very small and sometimes even negative. We also estimated related models for TEAM ratings of assistant principals (APs), and in this case did find some positive correlations in some years.

TELL-TN. The 2013 TELL-TN survey of teachers included a school leadership module which asked teachers to rate various characteristics of the leadership in their schools. Finding teachers' responses across items in this module to be very highly inter-correlated, we reduced the responses to a single scale score that we averaged across all responding teachers in the school. We then tested for associations between SLLA score and this summative measure.

Regardless of the statistical model employed, we find no evidence of a meaningful or statistically significant correlation between SLLA passage or score and TELL-TN summative scores. This null result holds from models with first-year principals only, principals in years 1-3, or all principals with SLLA scores. We also examined APs' ratings of leadership and found no correlation with SLLA score.

TVAAS. In 2013 and 2014, TDOE calculated school-level composite TVAAS measures that aggregated school performance across available tests in each school. We used this one-year composite growth measure to test whether schools' TVAAS performance is associated with student test score growth.

² The 7 highest-scoring programs (in terms of average SLLA score) for the more recent test form (1011/6011), in descending order, were Lipscomb, UT-Knoxville, East Tennessee State University, UT-Chattanooga, Austin Peay, Cumberland, and Middle Tennessee State University. All had failure rates of 9% or lower. The 7 lowest-scoring, in ascending order, were Cambridge College, Freed-Hardeman, Tennessee State University, Union, University of Memphis, Tennessee Technological University, and Trevecca Nazarene.

As with the prior measures, we found no evidence that TVAAS scores were associated with SLLA scores once other factors were taken into account.

TCAP Adjusted Growth. As an additional look at student achievement, we also estimated student-level growth models for math and reading for elementary and middle schools, using data from 2008 to 2014. These models estimate a student's test score in a given subject in year t as a function of the scores in both math and reading in year $t-1$, plus a large number of controls for student and school characteristics, and controls for grade and year. We estimated these models with and without other principal characteristics, and again for first-year principals, principals in years 1–3, and all principals with SLLA scores. Even the models using only first-year principals make use of data from more than 200,000 student observations, providing substantial power for detecting associations with principal SLLA score.

As with TVAAS, however, we find no evidence of an association between SLLA score and student achievement growth in any model for any subject. The SLLA coefficients are uniformly small and never statistically distinguishable from zero.³

We also tested for associations between the SLLA and outcomes only for principals taking the newer form of the test (form 1011/6011). Although the recency of this change means small sample sizes, which will make differences harder to find statistically, we did not see any clear evidence that the newer form is a superior predictor.

³ We also ran models using the full sample of principals with SLLA scores that interacted SLLA score with a categorical measure of principal experience (0, 1, 2, 3, 4, 5+). This model would allow for the possibility that any principal performance differences identified by the SLLA could take a few years as a principal to show up. These interactions were small and never statistically significant, however; SLLA score is not a predictor of TCAP score growth in either subject at any of the tested experience levels.

Conclusion

Tennessee principal candidates' scores on the SLLA vary markedly by their individual characteristics, including race and gender. Gender differences in scores likely are less relevant given that the group traditionally underrepresented in school leadership positions—women—in fact score higher. Racial differences, however, are more concerning, given that the significantly lower scores achieved by nonwhite principals translate into much higher failure rates. The SLLA appears to be a much more significant barrier to administrator licensure for nonwhite candidates than for white candidates.

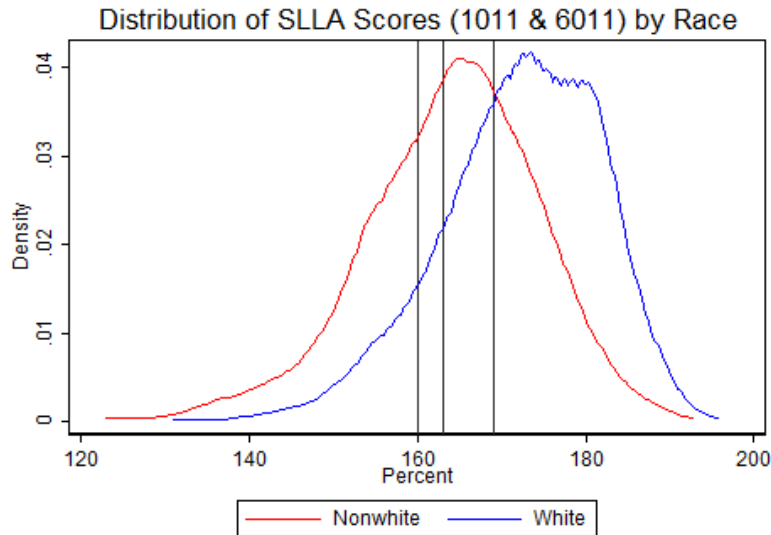
This difference is even more concerning given the absence of evidence that a principal candidate's performance on the SLLA is predictive of his or her performance after moving into the principalship. Across TEAM ratings, TELL-TN school leadership ratings, TVAAS scores, and TCAP growth in math and reading, we uncovered minimal evidence that SLLA scores correlate with other outcomes.

An important caveat to these results is that, in other analysis, we do find that higher scorers on the SLLA are significantly more likely to be hired into principal positions, which suggests that the SLLA captures something that districts find attractive in hiring school leaders.⁴ Note that we do not find, however, that high scorers are any more or less likely to remain in their jobs. We also find some evidence that SLLA score may predict ratings of AP job performance.

Our findings permit a useful thought experiment. Suppose that Tennessee was to raise its licensure cut score from 160 to 163 (equal to the score required in Missouri or Arkansas) or 169 (the score required in Mississippi, the highest in the U.S.). Our results show no evidence that either of these moves would result in a “higher quality” principal workforce. In fact, we explicitly tested whether principals

⁴ If higher scorers are more likely to be hired, one may wonder whether the weak correlations between the SLLA and other outcomes could be driven by the fact that low scorers, who would have been less effective as principals, simply tend not to be hired into those positions. However, we still observe substantial variation in SLLA score among principals—i.e., many low scorers go on to be hired—so this concern does not appear important. Besides, unless districts are examining SLLA scores when making principal hiring decisions, it is unlikely that any “weeding out” of ineffective principals is a result of the SLLA itself.

scoring at least a 163 or at least a 169 had higher performance than principals scoring lower across any of the outcome measures, and in no cases could we reject the null hypothesis that lower-scoring principals had the same outcome as those meeting the higher cutoffs.



A higher cutoff would, however, reduce the number of personnel licensed to be administrators, and this reduction would come disproportionately from nonwhite aspirants, as shown by the figure above.

The red line shows the existing distribution of scores of nonwhites on the newer form of the SLLA; the blue line shows the distribution for nonwhites. Given the current distributions, raising the cut score to 163 would result in an overall failure rate of 21%, but it would be 15% for white candidates and 39% for candidates of color. Further raising it to 169 would mean an overall failure rate of 40%, including 33% for whites and 65% for nonwhites. It does not appear that raising the cut score is consistent with the goals of increasing racial diversity in the Tennessee school administrator workforce.