



A FINAL REPORT ON PENNSYLVANIA'S FRAMEWORK FOR LEADERSHIP: FREQUENTLY ASKED QUESTIONS

A new study from REL Mid-Atlantic finds that Pennsylvania's tool for evaluating school leaders, the Framework for Leadership (FFL), is a promising principal evaluation tool, producing scores that appear to be both reliable and related to principals' contributions to student achievement growth. The study analyzed data from 517 principals and 123 assistant principals participating in Pennsylvania's pilot of the FFL in 2013/14.

The report finds that the FFL produces reliable scores, as measured by both internal consistency and year-to-year stability. The FFL had good internal consistency for principals and acceptable internal consistency for assistant principals, meaning that school leaders identified as effective or ineffective on one domain of the FFL tended to be identified in a similar way on the other domains. The FFL also exhibited a moderate level of year-to-year score stability that is comparable to the stability of widely used teacher observation instruments.

Most importantly, the study provides the first tentative evidence of the FFL's validity: principals with higher scores on the FFL, on average, make larger estimated contributions to student achievement growth. No other measures of principals' professional practice have been shown to be related to principals' effects on student achievement (Grissom et al., 2015). Variation in FFL scores was nonetheless limited: most school leaders scored in the upper third of the rating scale. As the FFL is implemented statewide, it will be important to examine whether there continues to be enough variation in scores to provide meaningful information on performance differences.

The full report, *Measuring school leaders' effectiveness: Final report from a multiyear pilot of Pennsylvania's Framework for Leadership*, is available at <http://ies.ed.gov/ncee/edlabs>.

PURPOSE OF THE STUDY

1. Why was this study conducted?

Legislation enacted in Pennsylvania (Act 82 of 2012) created a new system for evaluating principals and assistant principals. Starting in school year 2014/15, half of a school leader's annual evaluation must be based on a supervisor's assessment of the quality of leadership practices. (The other half must be based on measures of student achievement.)

As mandated by the legislation, the Pennsylvania Department of Education created the Framework for Leadership, a tool for supervisors to rate the quality of school leaders' practices. The department piloted the tool in school years 2012/13 and 2013/14 and requested an independent study of the pilot data to assess how well the tool measures the effectiveness of school leaders. An interim report based on data from 2012/13 was released in December 2014. This final report is primarily based on data from 2013/14 but also uses data from 2012/13 for some analyses.

2. What is the Framework for Leadership?

The Framework for Leadership (FFL) was developed by the Pennsylvania Department of Education to measure the quality of school leaders' practices. The FFL specifies 20 leadership practices on which each school leader is rated by an administrator—such as a superintendent or assistant superintendent—who has supervisory authority over the school leader. On each practice, a school leader can receive a score of distinguished (3 points), proficient (2 points), needs improvement (1 point), or failing (0 points) on each component. School leaders also receive a summary score (with the same possible 3, 2, 1, or 0 points) for each domain. The ratings supervisors assign are based on direct observation and on evidence submitted by the school leaders.

The 20 practices are grouped into four categories: (1) strategic/cultural leadership, (2) systems leadership, (3) leadership for learning, and (4) professional and community leadership. In the 2012/13 pilot year, the FFL included 19 practices; one additional practice in the Systems Leadership domain was added and implemented in the 2013/14 pilot year. Analysis for this study is specific to the 19 practices that were consistent across the 2012/13 and 2013/14 pilot years.

3. Why was it important to gather evidence on the accuracy of the Framework for Leadership?

This analysis of data from the pilot of the Framework for Leadership (FFL) can help education officials in Pennsylvania refine and improve this evaluation tool before it is used in formal evaluations of school leaders. Evidence on the FFL's strengths and weaknesses can also be used by other states and districts to inform the development or refinement of their own tools for measuring school leaders' effectiveness.

KEY FINDINGS AND LESSONS FROM THE STUDY

4. What strengths of the Framework for Leadership did this study find?

The study identified three strengths of the Framework for Leadership (FFL):

- i. *In both pilot years, different parts of the tool were consistent with each other in producing conclusions about a school leader's effectiveness.* School leaders who earned higher scores in one category of leadership practices also tended to earn higher scores in the other categories. This finding suggests that different parts of the FFL are connected by a coherent definition of leadership quality.
- ii. *School leaders' total and domain scores in one year were moderately consistent with their scores in the next year.* Scores of individual principals varied somewhat across the two pilot years, as might be expected, particularly if principals improve their practice over time. But the study did not find evidence that scores have large random fluctuations. The scores on the FFL are approximately as stable as those on several commonly used teacher observation tools, including the Framework for Teaching.
- iii. *Principals with higher FFL scores tended to have larger contributions to student achievement growth.* While no relationships were found between principals' estimated contributions to student achievement and FFL scores in the 2012/13 pilot year, the 2013/14 pilot year provided the first tentative evidence to validate the FFL's accuracy in measuring performance.* On average, principals with higher value-added in all subjects combined and higher value-added in math had higher total FFL scores and higher scores in two of the four domains. This finding sets the FFL apart from other principal evaluation tools, for which no evidence of a relationship with principal effects on student achievement currently exists.

5. What shortcomings of the Framework for Leadership did this study find?

The primary shortcoming of the Framework for Leadership (FFL) is that, in both the 2012/13 and 2013/14 school years, FFL scores did not differ much among school leaders. Most ratings were high. On specific leadership practices, supervisors rated principals and assistant principals as proficient or distinguished 95 percent of the time and as needing improvement in the remaining 5 percent. Failing ratings were extremely rare. In contrast, prior research has shown that principals differ considerably in their effectiveness at raising student achievement. Moreover, the average estimated contribution to student achievement growth for school leaders participating in the pilot was not statistically distinguishable from the state average. A possible explanation of this finding is that supervisors in this pilot were being lenient in their assignment of ratings.

During the pilot years, FFL scores were not part of principals' formal evaluations. The variation in scores may become even more limited when FFL scores become part of school leader evaluations if the high stakes incentivize lenient ratings.

One additional shortcoming of the FFL is that, in both pilot years, the internal consistency of the FFL for assistant principals was substantially lower than the internal consistency for principals. Only one of the four domains meets an acceptable level of internal consistency for assistant principals, and one of the three remaining domains does not meet the threshold even for a marginally acceptable level.

6. Is the Framework for Leadership any more or less accurate than other tools for rating school leaders?

The study provides some limited evidence that the Framework for Leadership (FFL) accurately measures school leader performance. The relationship between principals' FFL scores and their estimated contributions to student achievement growth corroborates the assessment of principal performance provided by the FFL. This corroborating evidence currently sets the FFL apart from other tools used to measure principal performance.

For nearly all school leader evaluation tools, there have been no studies as yet documenting how well those tools indicate school leaders' contributions to student achievement growth. The few tools that have been examined are district-specific evaluation instruments and have not been shown to be consistently associated with principals' contributions to student achievement growth.

* *Why was there a relationship in the 2013/14, but not in the 2012/13 pilot year? There are two possible factors. First, a larger sample of school leaders participated in the 2013/14 pilot year relative to the 2012/13 pilot year. Second, the supervisors assigning ratings in the 2013/14 pilot year may have been better equipped to apply the standards for assigning ratings accurately and reliably, perhaps due to an additional year of experience with the FFL and to improved training materials from the Pennsylvania Department of Education.*

7. What lessons can be drawn from the study findings to improve the Framework for Leadership?

Continuing to gather and examine evidence on the Framework for Leadership (FFL) will be especially important during the early years of implementation of the FFL as supervisors are becoming familiar with the tool. Based on the prevalence of high scores across two pilot years, the Pennsylvania Department of Education might consider the use of corroborative evidence to check that supervisors are applying the standard for FFL scoring correctly. One approach is to gather ratings of school leaders by other knowledgeable individuals, such as teachers. Evidence from ratings by teachers could be used to compare average scores based on teachers' ratings with average scores based on supervisors' ratings to assess whether supervisors are being too lenient or too strict.

In addition, the more limited evidence of the extent to which the FFL defines a coherent measure of leadership quality for assistant principals suggests the need to continue to gather evidence on the statistical properties of the FFL for assistant principals specifically. The current role of the assistant principal in practice may not fit the construct of school leadership defined by some components of the FFL. The Pennsylvania Department of Education might consider either tailoring the FFL more specifically to the assistant principal position or redefining the role of the assistant principal in line with the school leader role conceptualized by the FFL.

STUDY DESIGN, PARTICIPANTS, AND STAKEHOLDERS

8. Who participated in this study?

A total of 517 principals and 123 assistant principals were evaluated in the 2013/14 pilot. These school leaders came from 541 schools spread across 193 school districts and other local education agencies. Most school districts in the study were required to participate because they received funds from the federal Race to the Top program. In addition, schools receiving federal School Improvement Grants to implement a transformation model of improvement were required to participate. Some school districts also voluntarily selected schools to participate.

9. Who sponsored this study?

This study was sponsored by the Regional Educational Laboratory Mid-Atlantic, an organization established by the U.S. Department of Education to help states and districts in the mid-Atlantic region use data and research to improve students' academic outcomes.

10. Who conducted this study?

Mathematica Policy Research conducted this study as a subcontractor to ICF International, which operates the 2012–2017 Regional Educational Laboratory Mid-Atlantic.



11. What was the Pennsylvania Department of Education's role in this study?

The study was conducted in collaboration with the Pennsylvania Department of Education to help it refine and enhance the Framework for Leadership (FFL). The department also provided the data (FFL scores and other administrative data on students and school leaders) used in this study. The department did not fund the study, conduct the analysis, or contribute to the report.

12. What were the limitations of this study?

Although this study developed a new analytic method for measuring school leaders' contributions to student achievement growth (as a point of comparison for Framework for Leadership (FFL) scores), there is not yet a clear consensus among researchers on the best approach to measuring these contributions, particularly for assistant principals.

A valid measure for estimating the contributions to student achievement growth for longer-serving principals in the face of limited longitudinal data remains outstanding. This study was able to estimate contributions to student achievement growth only for principals with six or fewer years of tenure at a school. It is unclear whether the estimated contributions to student growth of longer-serving principals are related to their FFL scores.

FUTURE STEPS

13. What comes next?

The Pennsylvania Department of Education has started implementing the Framework for Leadership (FFL) statewide during the 2014/15 school year. It will be important for the Pennsylvania Department of Education to continue to closely monitor implementation of the FFL across the state to confirm that the FFL is a valid and reliable measure of performance across all school leaders in the state, not just among the sample of pilot participants. The Pennsylvania Department of Education might also consider gathering more evidence on the accuracy of the FFL as an evaluation tool; for example, by examining the relationship between FFL scores and other measures of school leader performance, such as principals' effects on graduation and dropout rates based on enrollment data or principals' effects on student safety and engagement based on student surveys.

Reference

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, 37(1), 3–28. <http://eric.ed.gov/?id=EJ1050959>