

Analysis of the Stability of Teacher-Level Growth Scores from the Student Growth Percentile Model

ANDREA LASH, REINO MAKKONEN, LOAN TRAN, MIN HUANG

<http://relwest.WestEd.org> | alash@WestEd.org

Why this study?

The Nevada Department of Education wanted to test an implicit assumption of high-stakes teacher evaluation systems that use student learning to measure teacher effectiveness: that the learning of a teacher's students in one year will predict the learning of the teacher's future students. Evaluation systems that identify low-scoring teachers for remediation assume that if the teachers are not retrained, their future teaching will also be relatively ineffective. Systems that award tenure to teachers who score higher assume that those teachers will remain effective. Examining the stability of teacher-level growth scores over time examines the validity of using such scores for teacher evaluation. In this study, growth, or student learning, was measured by the student growth percentile model.¹

What the Nevada Research Alliance examined

The alliance examined three years of teacher-level growth scores in reading and in mathematics for about 370 teachers of grades 4 through 8. The scores that were derived from student data provided by Washoe County School District, Nevada's second largest district. As defined by the state, a teacher's growth score is the median of the growth scores of the teacher's students.

Through a generalizability study,² the alliance derived

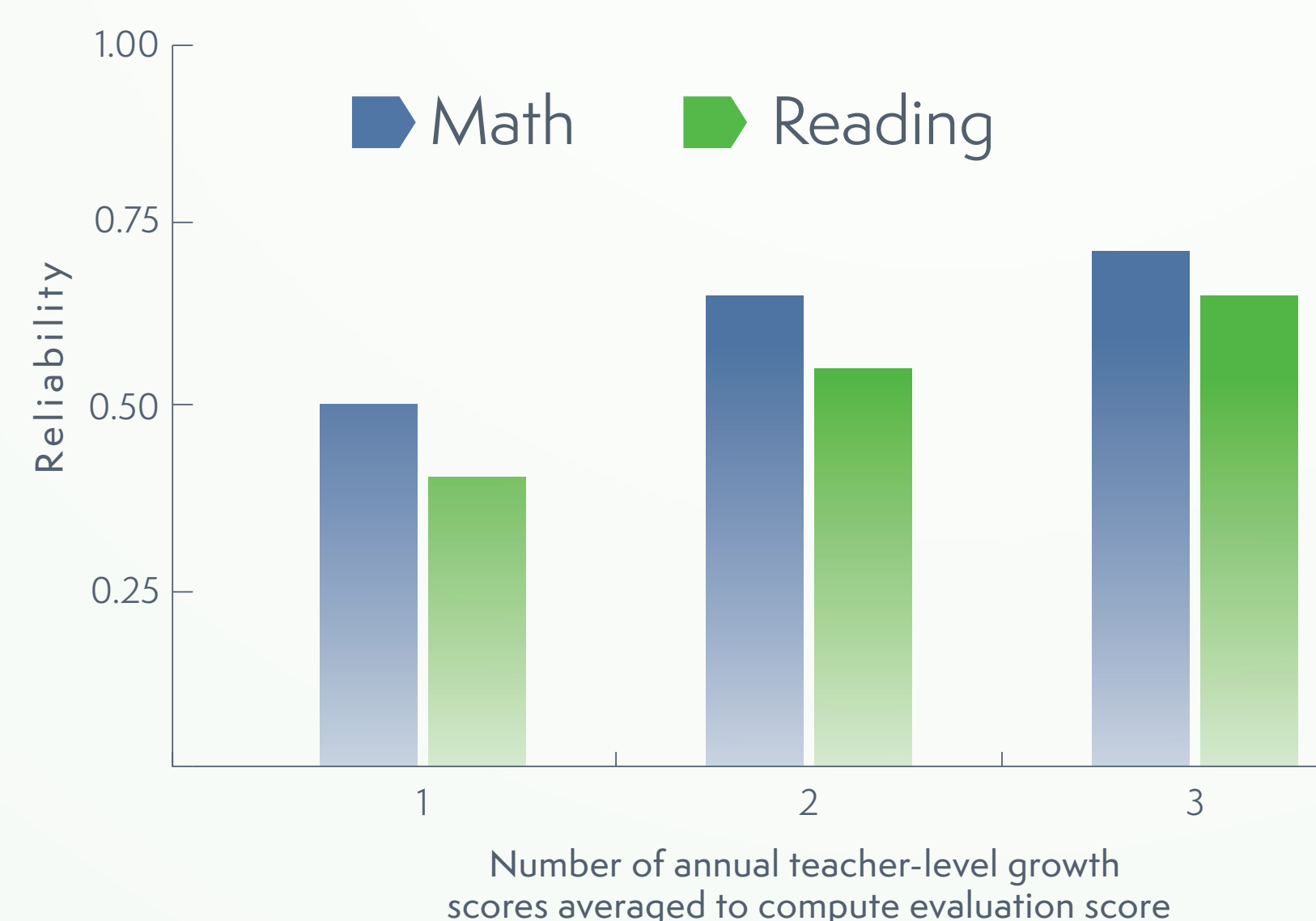
- » the reliability, over time, of the annual teacher-level score along with estimates of how reliability would change if teachers were evaluated based on the average of two or three years of annual scores (see Figure 1);
- » a confidence interval useful for examining the likelihood of making errors when judging teachers (see Figure 2); and
- » a method to estimate error rates when specific rules are applied to evaluate teachers (see Figure 3).

How the findings may be used

In Nevada, any changes to the current design of the teacher evaluation system will require that new legislation be passed in the 2017 legislative session. Results of the study have been shared with the Nevada Department of Education and the Teacher and Leader Council, a committee of stakeholders, appointed by the governor, to recommend how to implement current law. Stakeholders are now considering what action, if any, should be taken.

Although designed to inform Nevada policymakers, the research may have value to policymakers in other states. Utah State Representative Poulson read a summary of the study when she introduced a new bill (HB0201) calling for removal of teacher-level growth scores from the teacher evaluation system. Administrators at Gwinnett County Schools in Georgia plan to replicate the study by applying the methods to their teacher-level growth data.

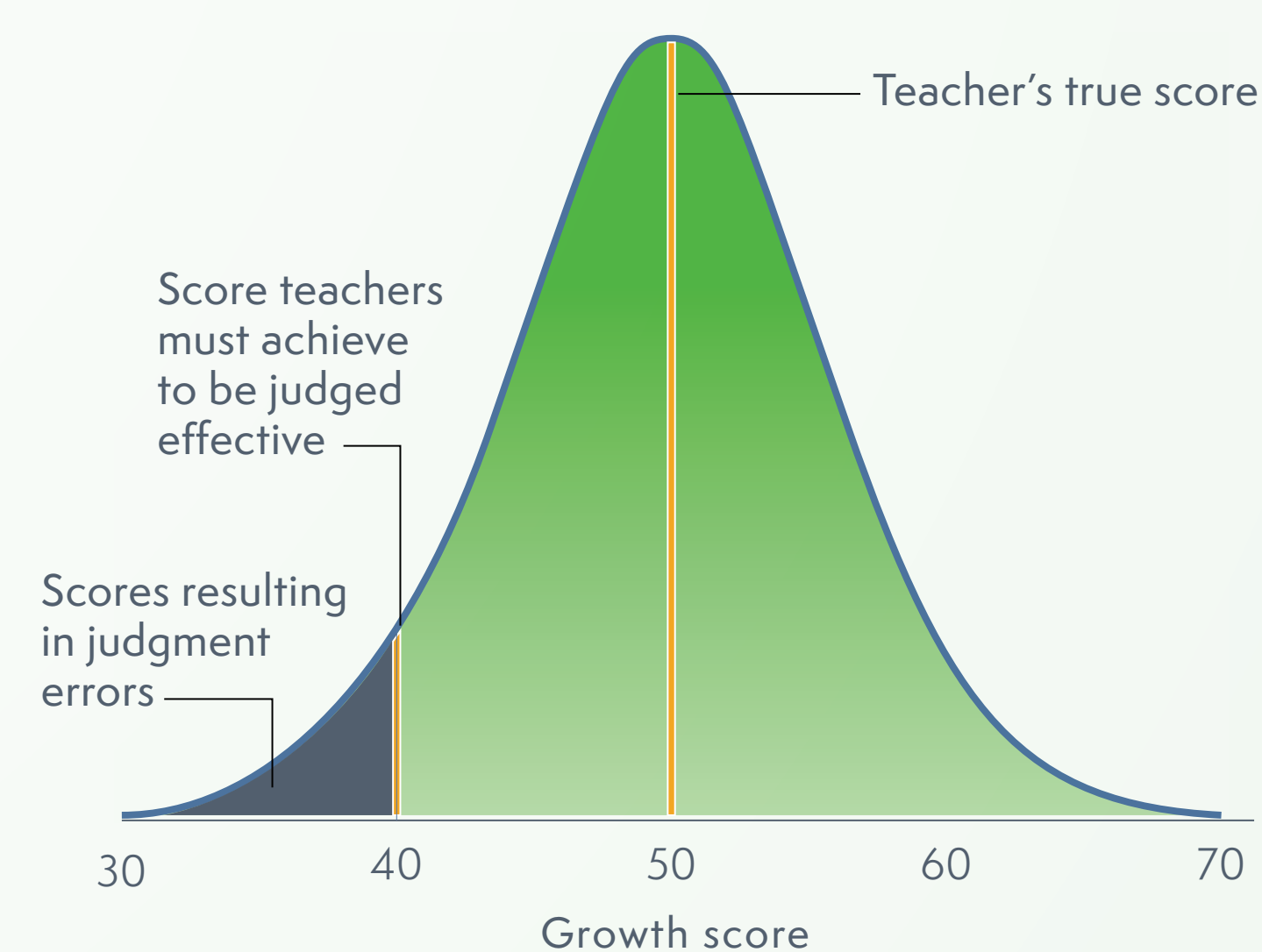
Figure 1. Reliability of measures of teacher effectiveness derived from teacher-level growth scores



What the study found

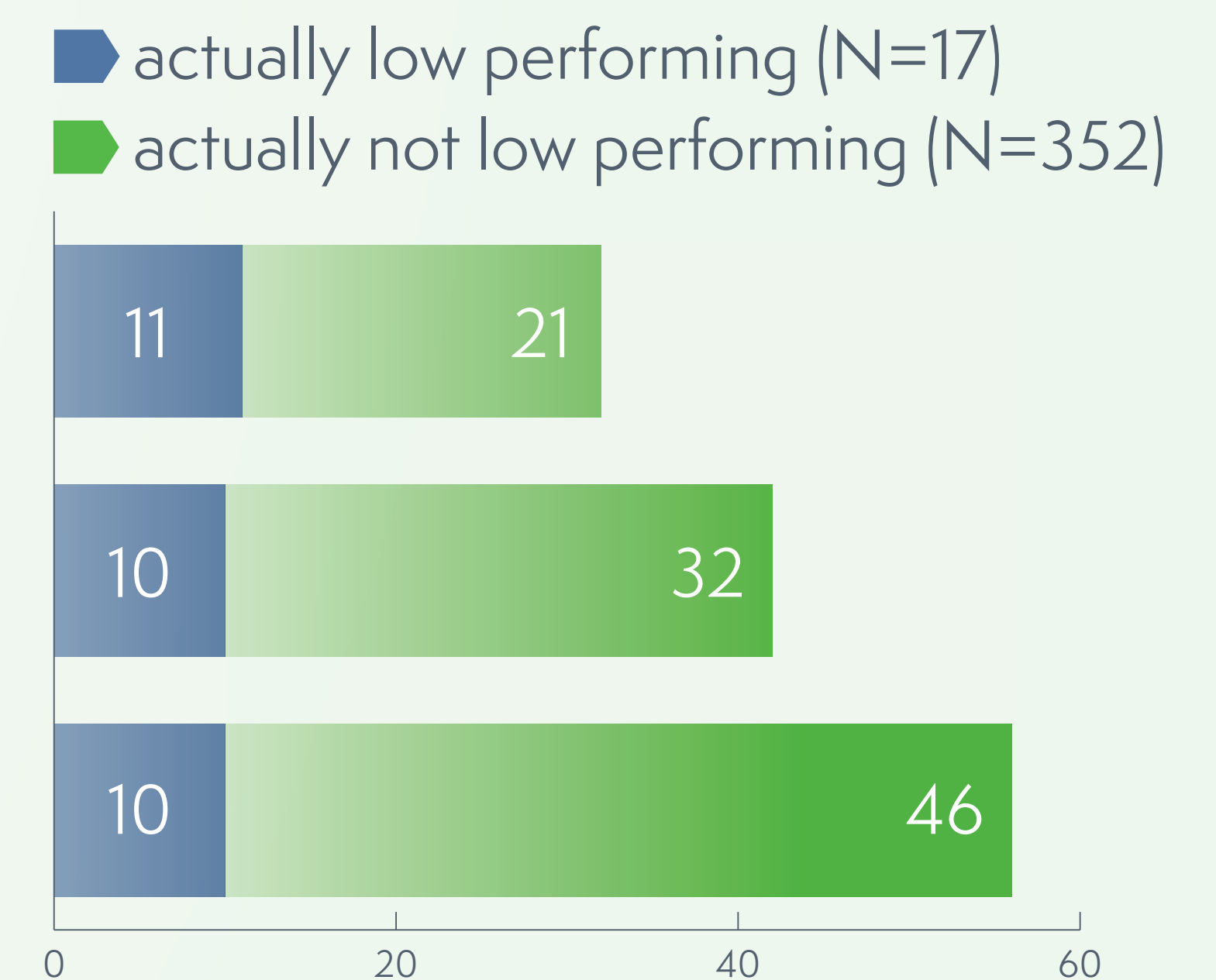
- » Half or more of the variance in annual teacher-level growth scores was due to random or otherwise unstable fluctuations.
- » Even when computed as an average of annual teacher-level growth scores over three years, estimates of teacher effectiveness would not be expected to meet the level of stability that some argue is needed for high-stakes decisions about individuals, which is a reliability of .85 or higher.³

Figure 2. Hypothetical distribution of possible growth scores for a typical teacher with true growth score at 50



- » The range of scores likely to include a teacher's true score would span close to half the 100-point scale. For the annual score, the 95 percent confidence interval would span 48 points for mathematics and 44 points for reading.
- » A "typical" math teacher, with a growth score of 50, would be expected to be misclassified as low performing 20 percent of the time, when low performance is defined as an annual growth score less than 40.

Figure 3. Teachers classified as low performing based on an annual growth score and on averages of 2 or 3 annual scores



REFERENCES

1. Betebenner, D. W. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories*. Dover, NH: National Center for the Improvement of Educational Assessment.
2. Brennan, R.L. (2001). *Generalizability theory*. New York, NY: Springer.
3. Haertel, E. H. (2013). *Reliability and validity of inferences about teachers based on student test scores* (14th William H. Angoff Memorial Lecture). Princeton, NJ: ETS.