Estimating the Amount of Noise in Longitudinal Proficiency Data among Knox County Elementary Schools

Technical Report

Clint Sattler
Supervisor, Research and Evaluation
Knox County Schools
Department of Research, Evaluation, and Assessment

June 2018

Overview

The reauthorization of the Elementary and Secondary Education Act (ESEA) in 2001 (known at the time as the No Child Left Behind [NCLB] Act) placed increased emphasis on standardized test results for school, district, and state accountability (Jorgensen, 2003). As a result of NCLB and later revisions to ESEA, district, and school improvement plans became increasingly focused on state assessment outcomes (Linn, 2002). Additionally, the Tennessee Department of Education (TDOE) requires schools and districts to reflect upon trends in their state assessment data as part of thier needs assessment (TDOE, 2017). After the 2018-2019 academic year (SY1819) schools will receive A through F grades related to student performance on state assessments (TDOE 2018). Accordingly, schools and districts make considerable monetary investments in programs and activities that aim to significantly impact student achievement (Goertz, 2005).

Virtually all schools in Knox County use state assessment data as part of their continuous improvement cycles and progress monitoring activities. Often, administrators are uncertain if incremental changes in student achievement are due to specific programmatic inputs or are related to random "noise" in the state assessment data (Bernhardt, 2013). Things like cohort refresh or uncertainty in state assessment scaled scores may lead to random fluctuations in student proficiency data (Brookhart, 2015). The Department of Research, Evaluation, and Assessment (REA) conducted this analysis in an attempt to provide estimates of the magnitude of "noise" in longitudinal state assessment data. The scope of this current study is limited to English/Language Arts (ELA) and Math results in grades 3 through 5.

Methodology

School goals are generally related to the percentage of students who meet grade level expectations on the state assessments. These students are placed in the top two performance categories on the Tennessee Comprehensive Assessment Program (TCAP). Prior to SY1516, the top two performance categories were "Proficient" and "Advanced" (P and A). Starting in SY1617, the top two performance categories were renamed "On-Grade Level" and "Mastered". For consistency's sake, this analysis uses the pre-SY1617 nomenclature. Further references to student "achievement" and "proficiency" refer to the percentage of students in the top two performance categories on the state assessment (see Equation 1).

Equation 1: Calculation of the Percent of Students in the Top Two Performance Categories

$$p\ or\ a_{ijk} = \frac{n\ Proficient\ (On\ Grade\ Level)_{ijk} + n\ Advanced\ (Mastered)_{ijk}}{n\ tested_{ijk}}$$

Where p or a_{ijk} is the raw percentage of students who were proficient or advanced in school i in year j and in subject k.

Longitudinal achievement data were aggregated by school for the academic years between 2009-2010 (SY0910) and 2016-2017 (SY1617). Exploratory analysis of the data included earlier data but the data collected prior to SY0910 shows insufficient school-to-school variation in order to be useful for this study. The dataset used in this analysis included 47 elementary schools with complete achievement data from SY0910 through SY1617. Data from SY1516 could not be included in the analysis because the state halted testing that year. All data was standardized by subtracting the arithmetic mean achievement of the 47 schools in any given year (see Equation 2).

Equation 2: Calculation of the Standardized Achievement Measure

$$P\ or\ A_{ijk} = p\ or\ a_{ijk} - \frac{\sum_{ik}^{n_{jk}} p\ or\ a}{n_{jk}}$$

Where P or A_{ijk} is the standardized percent of students who were proficient or advanced in school i in year j and in subject k, p or a_{ijk} is defined in Equation 1, the summation term is the sum all school-level proficiency data in year j and in subject k, and n_{jk} is the total number of schools with proficiency data in year j and in subject k (47).

The amount of "noise" in each schools' longitudinal achievement data was estimated using additive time series seasonal decomposition. Decomposition deconstructs the data into an overall trend (a centered moving average) and an irregular component (error, or noise).

Seasonal variations in the data were not modeled as no seasonality was expected in the data. Figure 1 shows a decomposed dataset for RLA at a specific Knox County elementary school. Each blue point represents a standardized achievement measure (as % P or A). The orange line is the centered moving average (calculated using three data points in this example) of % P or A. The irregular component for each year is represented by the grey lines. The estimate of the amount of "noise" in the longitudinal state assessment was determined as the median of the absolute values of each irregular component from all years and schools for a given subject (see Equations 3 and 4). The 95% confidence interval for each median was calculated using bootstrapping with the bias corrected and accelerated (BCa) methodology.

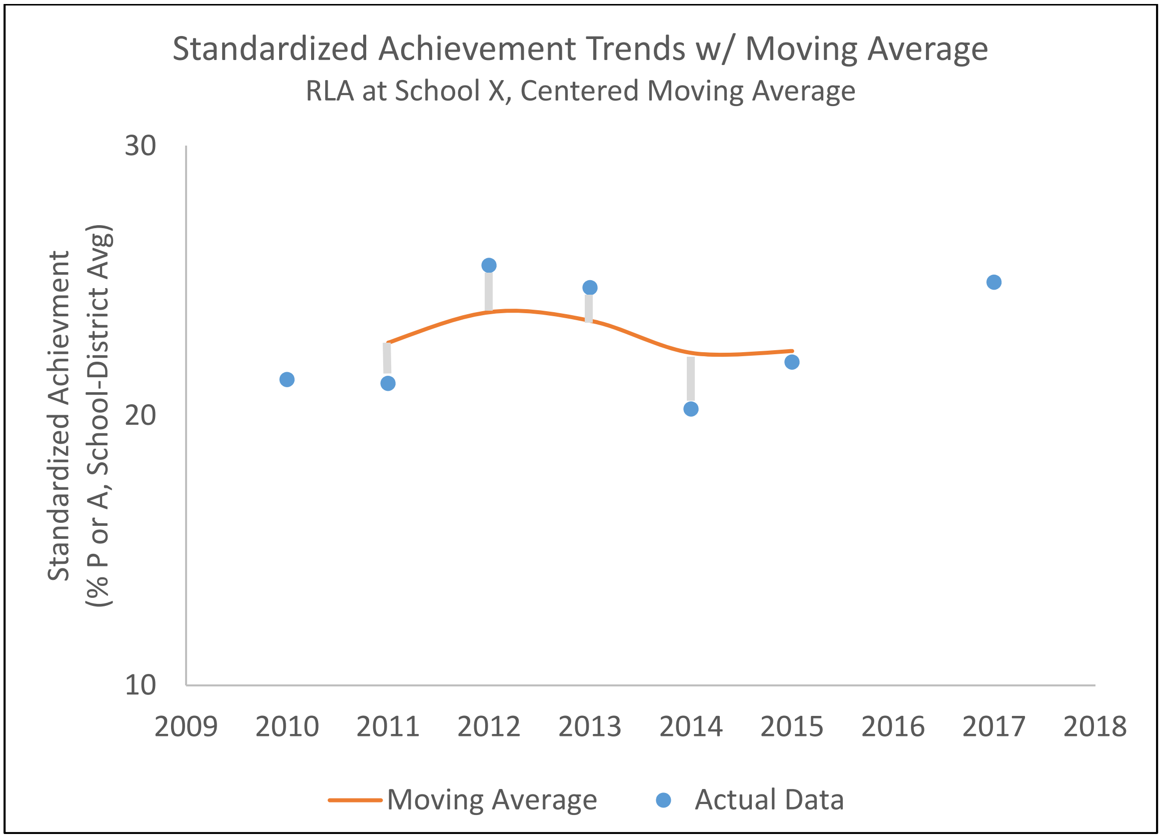

Figure 1: Example Time Series Decomposition

Equation 3: Calculation of Irregular Component of Time Series Decomposition

$$ic_{ijk} = \left|P\ or\ A_{\ ijk} - ma_{ijk}\right|$$

Where ic_{ijk} is the irregular component in school i in year j and in subject k, P or A_{ijk} is defined in Equation 2 and ma_{ijk} is the centered moving average of the achievement data in school i in year j and in subject k.

Equation 4: Calculation of Noise Estimate

$$noise_{est,k} = Median(ic_{ijk})$$

Where $noise_{est,k}$ is the estimated noise in the longitudinal data for subject k and ic_{ijk} is defined in Equation 3.

Time series decomposition requires one to specify the period of time over which to calculate the centered moving average. Each schools' data was analyzed in isolation using the Time Series Analysis (TSA) package in R in order to identify any periodicity in the schools' longitudinal trends. Each school-level periodicity in both math and RLA were analyzed in order to choose the most appropriate periodicity for the entire dataset.

The R packages used in this study include TSA (version 1.01), forecast (version 8.3), fpp (version 0.5), and boot (version 1.3-20). All calculations were completed using R version 3.4.3 running on RStudio version 1.0.143.

Results: Periodicity

The interquartile range of the 47 school-level periodicities can be found in Table 1. The median periodicity in both subjects was 4 years and was selected for the period over which to calculate centered moving averages in the achievement data for both math and RLA.

Table 1: Subject Specific Periodicity (n=47 for Each Subject)

Subject	0%	25%	50%	75%	100%
RLA	2 years	3 years	4 years	4 years	8 years
Math	2 years	3 years	4 years	4 years	8 years

Results: RLA

The centered moving average for each year was calculated using a period of 4 years. The irregular component was extracted per Equation 3. This resulted in a total of 141 individual estimates of irregular components (3 irregular components per school * 47 schools). The distribution of the irregular components for RLA are visible in the violin plot in Figure 2.

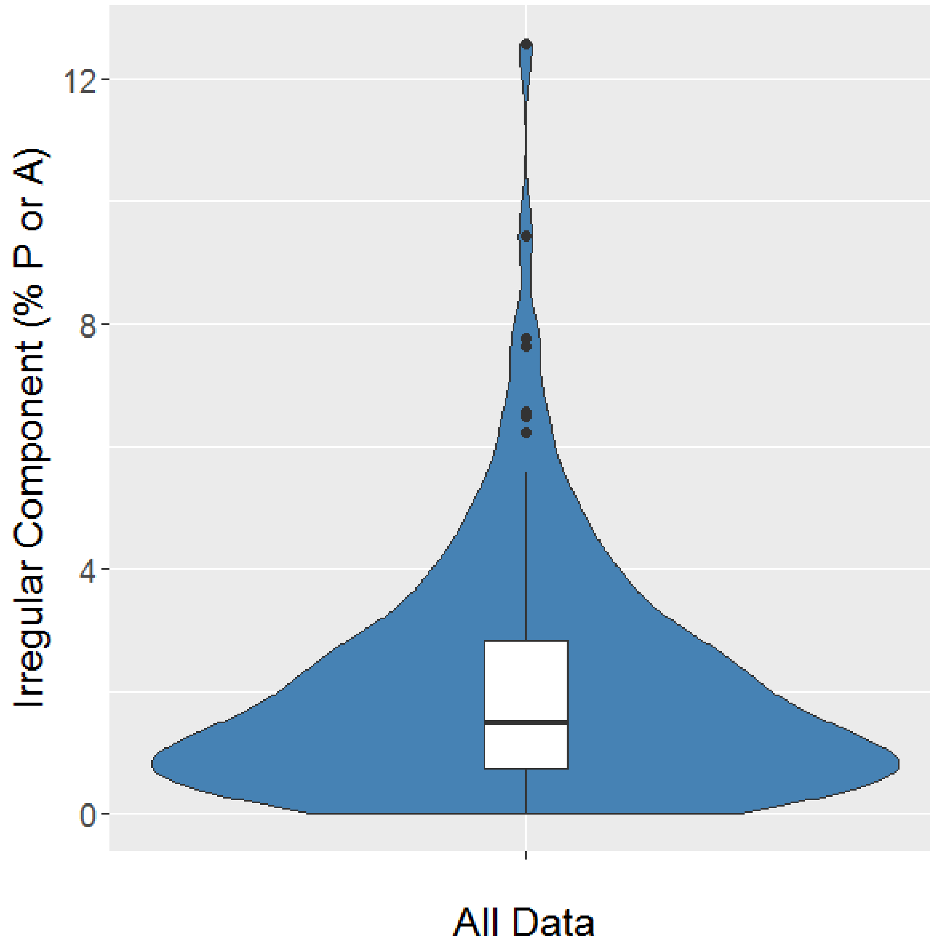

Figure 2: Distribution of Irregular Components: RLA

The median of the 141 irregular components was 1.50 percent proficient or advanced. The bootstrapped 95% confidence interval for the median was 1.16 percent proficient or advanced (lower) to 1.92 percent proficient or advanced (upper).

Results: Math

The centered moving average for each year was calculated using a period of 4 years. The irregular component was extracted per Equation 3. This resulted in a total of 141 individual estimates of irregular components (3 irregular components per school * 47 schools). The distribution of the irregular components for math are visible in the violin plot in Figure 3.

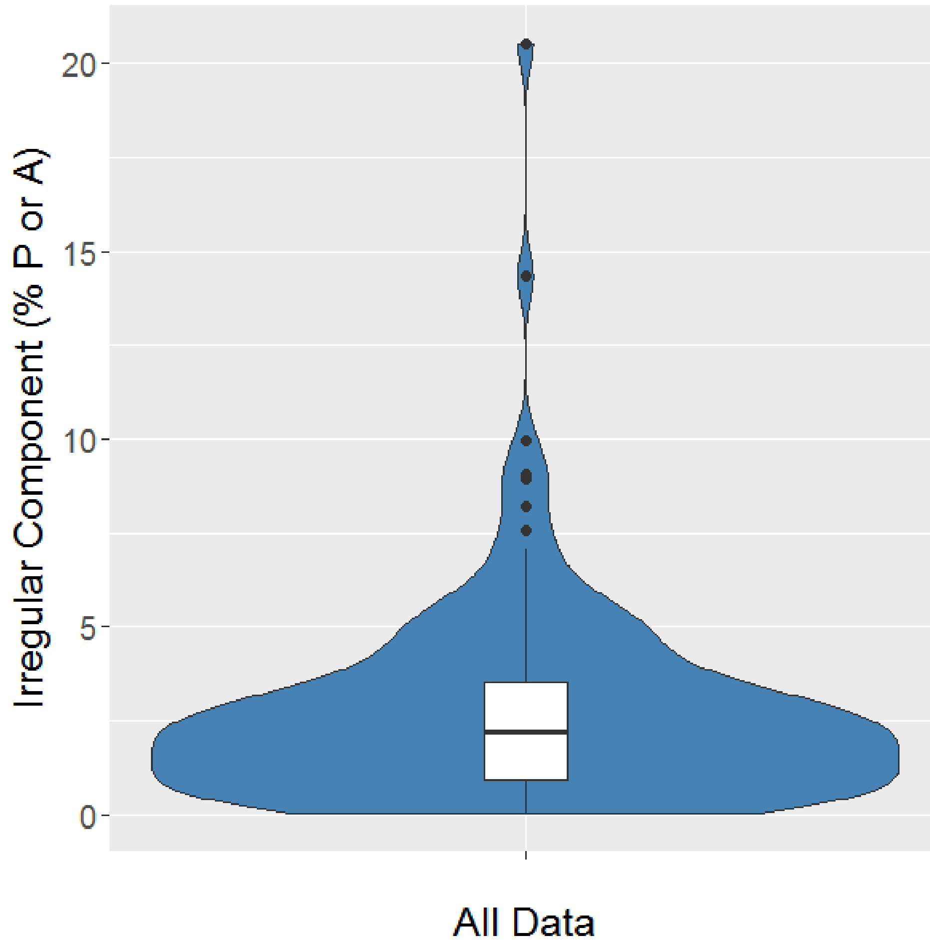

Figure 3: Distribution of Irregular Component: Math

The median of the 141 irregular components was 2.19 percent proficient or advanced. The bootstrapped 95% confidence interval for the median was 1.70 percent proficient or advanced (lower) to 2.52 percent proficient or advanced (upper).

Conclusions & Considerations

Quantifying the irregular component in longitudinal proficiency data can provide an estimate of the amount of "noise" that is present in longitudinal state assessment data. Using the state assessment data collected between SY0910 and SY1617, the Department of Research, Evaluation, and Assessment determined the following estimates for "noise" in grades 3 through 5 math and RLA achievement:

- RLA: Approximately 2 percent Proficient or Advanced (On Grade Level or Mastered)
- Math: Approximately 2.5 percent Proficient or Advanced (On Grade Level or Mastered)

These estimates should only be considered "rule of thumb" quantities rather than precise measurements of error. These estimates should be considered bi-directional, such that a 2 percent gain OR loss in school-level achievement may not be reflective systematic change. When year-over-year changes are within 1 to 2 percentage points, other data will need to be consulted to determine if schools are on sustainable positive trajectories.

For example, school X invested in a skill-based adaptive computer program that was used during centers rotations in order to support small-group Reading instruction (and changed nothing else in regards to Reading instruction from the previous year). The goal of this investment was to increase school-wide ELA proficiency scores on state assessments. If the percent of students who tested "On-Grade Level" or "Mastered" increased by 2 percentage points, the increase may be due to the implementation of the program or due to random "noise" in the TCAP data. Therefore, the school should seek to compare year-over-year changes in student work and formative assessments in order to assess if the students were more likely to demonstrate mastery of Reading skills after treatment. Similarly, if the percent of students who tested "On-Grade Level" or "Mastered" increased by 4 percentage points, it is possible that the "true" achievement gains related to the program could be estimated at anywhere between 2 and 6 percentage points by the "rules of thumb" above.

REA may continue to monitor these trends in order to see if additional data collected under the TNReady assessment lead to differences in these "noise" estimates. Additionally, further modeling could be done to determine if covariates impact estimates of the "noise" in the data. For example, these "noise" estimates may be a function of school size and/or incoming student achievement levels. If these hypotheses are true, further modeling may allow for more precise estimates of "noise" in the state assessment.

References

Brookhart, S. M. (2015). How to make decisions with different kinds of student assessment data. ASCD.

Bernhardt, V. L. (2013). Data analysis for continuous school improvement. Routledge.

Goertz, M. E. (2005). Implementing the no child left behind act: Challenges for the states. Peabody Journal of Education, 80(2), 73-89.

Jorgensen, M. A., & Hoffmann, J. (2003). History of the No Child Left Behind Act of 2001 (NCLB). Assessment Report. Pearson Education: San Antonio.

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the no child left behind act of 2001. Educational Researcher, 31(6), 3-16.

Tennessee Department of Education (2017). ePlan Resources. Retrieved from https://eplan.tn.gov.

Tennessee Department of Education (2018). 2018 Accountability Protocol. TDOE Office of Accountability.