What Counts in Calculating School and District Level Performance Index Scores: *A Summary and Analysis of Academic Performance Index Metrics across the 50 States*[1]

Xinyu Ni
Teachers College,
Columbia University
xn2115@tc.edu

Alex J. Bowers[2]
Teachers College,
Columbia University
Bowers@tc.edu

Jennifer Esswein
Education Northwest
Jennifer.Esswein@educationnorthwest.org

ABSTRACT:

The purpose of this report is to summarize the key elements of school and district level Performance Index scores (PI scores) for the 50 states and the District of Columbia (D.C.) across the United States. PI scores are partial or overall summative ratings of schools or districts currently used across US state accountability systems to assess organizational performance. In this study, we first extracted 14 elements from 49 PI calculation metrics for states in the U.S and conducted a descriptive analysis to provide an overview of which data elements are used across the different calculation metrics for each state and what role PI scores play in state accountability systems. Second, we categorized the fourteen elements into seven categories proposed by the most recent ESSA regulations (81 FR 34539 §200.14-16, 2016) and examined how each state integrated each element in their PI score calculations. Third, we conducted a multidimensional scaling (MDS) analysis to compare the similarities and differences of PI calculation metrics across the states. The results indicate that there are few commonalities in PI score calculation metrics across the states, as each state has its own methods in addressing the requirements of NCLB and now ESSA. The goal of this report is to inform decisions across states on PI score calculations through summarizing overall ratings and metrics nationally used to hold schools and districts accountable as states move toward implementing the recent Every Child Succeeds Act (ESSA) regulations.

Keywords: Accountability, School District, Evaluation, Assessment, Performance Index;

INTRODUCTION:

Over the past half century, the U.S government has promoted increasing student overall academic performance through focusing on closing achievement gaps (Barton & Coley, 2009; Freeman & Freeman, 2002, 2002; Reb EL & Wolff, 2009; Singham, 2005). How to hold districts and schools accountable for making progress in this area dates back to 1965 with the first authorization of the Elementary and Secondary Education Act (ESEA). In 1965, President Johnson signed ESEA into federal law to promote equal access to education and close achievement gaps across schools. For each year, ESEA awarded more than $1 billion to school districts serving students from low-income families. Funding is firstly distributed to state education agencies (SEA), then allocated to local education agencies (LEA) which then disperse funds to public schools with more than 40 percent of students from low-income families qualifying for the United States Census's definition of low-income. Through ESEA, schools, districts, and state departments of education are required to be accountable for improving students' performance (Bailey & Mosher,1968; Bloomfield & Cooper, 2003; Tomlinson & Jarvis, 2014).

The notable reauthorization of ESEA named Improving America's Schools Act (IASA) of 1994 was a major effort of President Clinton's administration in education reformation. The IASA (Marti, Sargrad & Batel, 2016) required each state to develop school-improvement plans and performance standards in English language arts and mathematics for elementary, middle and high schools. Additionally, districts submitted plans to their states describing which assessments were used specifically with Title I students and what strategies districts employed to coordinate services within districts to improve the performance of students from historically disadvantaged backgrounds. Thus, under the framework of IASA, states such as Massachusetts and Texas began to implement school accountability systems which are tied to rewards and designations (Herdman, 2002; Horm-Wingerd, Winter & Plofchan, 2000).

The 2002 reauthorization of ESEA, known as No Child Left Behind (NCLB), took a progressive step in holding schools accountable for students' outcomes. Under the NCLB law,

[1] This report was published online as a white paper report, August 11, 2016. Suggested citation format:
Ni, X., Bowers, A.J., Esswein (2016) What Counts in Calculating School and District Level Performance Index Scores: A Summary and Analysis of Academic Performance Index Metrics across the 50 States. A White Paper Report. New York, NY: Teachers College, Columbia University.

[2] Corresponding author. Alex J. Bowers; Bowers@tc.edu; Teachers College, Columbia University, Box 67, 525 West 120th Street, New York, NY 10027.

states must perform statewide standardized tests in mathematics and reading in grades 3 through 8 and report the results for both the whole population and subgroups including economically disadvantaged; special education; percentage of minority students, and English learners (Bloomfield & Cooper, 2003; Redfield & Sheinker, 2004). Additionally, NCLB required schools to report whether schools meet "adequate yearly progress" (AYP) goals set by the states. If a school did not make AYP for multiple years, the school would be faced with serious sanctions including losing state funds or shut down (Pruitt & Bowers, 2014). Major critiques for NCLB focused on the issues of heavy reliance on cross-sectional standardized test scores while ignoring other indicators in school or district accountability systems such as students' preparation for post-secondary schools and careers. Additionally, researchers and practitioners criticized that the one-size-fits-all evaluation system under NCLB was overly rigid and limited states' capacity in enacting a more meaningful and comprehensive accountability system (Carr, Olsen, & White, 1992; Carr, Wallin, & Andrew Carr, 2000). Thus, to address those critiques, the Obama Administration recently approved more flexibility under the new ESEA waivers for 43 states and D.C. by the year 2015 to provide flexibility around many requirements under NCLB in exchange for more holistic state-developed accountability plans designed to improve educational outcomes for all students and close achievement gaps. Instead of binary "pass-or-fail" standards required by the previous AYP accountability system, the ESEA flexibility waivers encouraged each state to assign designation letters or multiple-tiered proficiency levels to schools and school districts (Dunlap, 2011).

The recent reauthorization of ESEA was signed into law by President Obama on December 10, 2015 and amended as the Every Student Succeeds Act (ESSA), which provides states flexibility to set ambitious but achievable expectations for all students and to take a more meaningful view to measure school overall performance based on five required measures including: students' achievement performance in reading and mathematics; academic progress in elementary and secondary schools; graduation rates in high schools; rates of progress for English learners achieving language proficiency and a state-determined measure of school quality or student success. The ESSA affirms that states are required to create multi-measure statewide accountability systems while giving states flexibility to choose indicators in each required measure. Furthermore, to promote transparency in a format that is easily understandable by parents, the proposed ESSA regulations require each state to assign a comprehensive and summative rating for each school to provide a clear picture of its overall standing based on all of the measures (81 FR 34539, 2016) .

Although the newly proposed ESSA regulations do not require states to report a similar summative rating for school districts, as local education agencies (LEA), school districts play important roles in improving students overall educational performance and closing achievement gaps. Under the amended ESSA regulations, LEAs take responsibility in coordinating with state education agencies (SEA) to implement targeted improvement plans for focus schools and to consult with stakeholders (school leaders, parents) to make intervention plans. Indeed, an increasing amount of research and practice literature on the effects of school districts on student achievement has indicated that school district personnel practices can have a strong effect on individual and overall student, school and district achievement (Bowers, 2008, 2010, 2015; Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Elmore & Burney, 1999; Honig, 2003, 2008; Leithwood, 2010; Levin, Datnow, & Carrier, 2012; Purkey & Smith, 1985; Trujillo, 2013). Understanding district summative ratings in the current accountability systems nationwide will provide informative single year-over-year outcomes to stakeholders and help address the issues from past outcome measures used in previous district effectiveness research (Bowers, 2010, 2015; Trujillo, 2013).

A recent study conducted by Center for American Progress (CAP) (Martin, Sargrad & Batel, 2016) analyzed measures that states currently include in their accountability systems. In the report, the authors first organize measures into seven major categories including achievement indicators; student growth indicators; English language acquisition indicators; early warning indicators; persistence indicators; college-readiness indicators; and other indicators. The authors then analyzed which indicators are included in each state. According to the results, states include eleven indicators on average with a minimum of four and a maximum of twenty-six indicators. Additionally, the authors conducted a weighting analysis for the states that combined all indicators into one single grade or overall rating to analyze the weight of each indicator in the composite score. The results suggested that academic achievement indicators account for an average of 48 percent of a school's accountability rating on average across the states followed by student growth indicators accounting for 45 percent. The category of other indicators, such as art access and physical fitness, accounts for the least with 10 percent of school accountability scores. The CAP report provides a comprehensive overview of what indicators states currently include in their accountability systems and the weight analysis for each indicator as an informative means to combine all the indicators into a single grade. However, the report did not address the issues of how states combine each indicator and how the metrics are similar or dissimilar with each other, as a means to comply with NCLB and ESSA. Additionally, for the states that do not combine all indicators into a single

grade, many combine partial indicators into a continuous PI score, which could be used for providing year-to-year outcomes to educational researchers, practitioners and stakeholders.

The focus of the present study is to examine the extent to which each of the states in the US calculates a school or district-level PI score and how they calculate the scores. Additionally, we further explored the extent to which the summarized elements of PI scores across the 50 states in the U.S fit the measures required by the new ESSA regulations and how PI scores are used in states' accountability systems. Through examining how each element in PI score metrics fit the ESSA regulations, the report is especially informative for states which are developing summative ratings metrics required by ESSA (81 FR 34539 §200.18, 2016) . The central research questions are: 1) how many different elements or features are currently used for PI score calculations across the 50 states and DC? 2) How does each element fit the required measures of the newly proposed ESSA regulations for state accountability systems? 3) How are school or district performance index metrics across the states similar or dissimilar from each other? 4) And given this review, which PI score metrics from which states could be reported as overall ratings for schools as required by ESSA regulations?

METHODS:

Data Collection

The data sources for this summary included primary documents on the school or district level performance indices for 50 states and DC in the United States. We did a comprehensive search for the metrics and construction of performance index scores of each state following four steps. First, we searched on the department of education website of each state for the district performance index information with the following keywords: school performance; school indicator; school accountability; school score; district performance; district indicator; district score; district accountability. Second, we examined the school and district level annual accountability report to examine whether there existed a continuous variable that summarized multiple indicators for rating school or district performance for each state. Third, we examined the ESEA waiver applications for 45 states exploring the partial or summative ratings in their accountability systems provided by the United States Department of Education (USDOE) and updated the summarized table in the appendix. Fourth, we cross-referenced the 50 States Accountability Report Cards Comparison Summary (Christie, 2013) provided by the Education Commission of the States and Center for American Progress (CAP) report of "A 50-State Analysis of School Accountability Systems" (Martin, Sargrad & Batel, 2016) to assure that all the states were included with single or multiple continuous indicators to measure school or district level performance. While we carefully executed these steps, the authors recognize that accuracy of the PI calculations presented in this paper is dependent upon the accuracy of the publically available information provided by each state.

According to the ESSA proposed regulations (81 FR 34539, 2016), state accountability systems are required to include the following five measures: academic achievement; academic progress indicator; four-year adjusted cohort graduation rate; indicator in measuring progress in achieving English language proficiency for English learners; and single or multiple indicators in school quality or student success (81 FR 34539 §200.14, 2016). Here, school quality or student success is a new measure amended by ESSA in which states are required to include one or more of the following: student access to and completion of advanced coursework; postsecondary readiness; school climate and safety; student engagement and educator engagement. Additionally, ESSA regulations, same as the previous NCLB waiver, also require states' standardized test participation meet 95% standard (81 FR 34539 §200.15, 2016). ESSA proposed regulations do not provide specific methods for accounting for test participation rates, but states are required to provide a clear explanation of what roles test participation rates play and how they include test participation rates in their accountability systems. Finally, according to the ESSA proposed regulations, states are required to report subgroup educational outcomes (economically disadvantaged students, students from major racial and ethnic groups, children with disabilities and English learners) (81 FR 34539 §200.16,2016). Thus, a total of seven measures are required to be considered in the newly amended ESSA regulations. In the following analysis, we extracted elements in each state's PI score calculation metrics and summarized it within the above seven categories so as to further explore how states' PI score calculations fit the recently proposed ESSA regulations.

Analysis

Among the 50 states and the District of Columbia in the United States, 48 states and D.C. have a single or multiple continuous Performance Index (PI) score to measure the overall performance of schools or districts. Thirty-four states have a single summative PI score of all the elements in their accountability systems which these states use to determine school or district designation letter grades, proficiency levels or AYP status. Based on states' PI score calculation metrics, we first extracted fourteen main elements from the data collected across the states and summarized the extracted elements into seven measures according to the ESSA regulations. Then, we performed descriptive statistics to quantify the characteristics of the calculation metrics for the 48 states and D.C., ranked the elements that counted in the metrics across the states. We

then used multidimensional scaling (MDS) to further explore similarities among different calculation methods from the types of elements states used.

MDS is a technique for the analysis of similarity or dissimilarity data on a set of subjects (Borg & Groenen, 2005). MDS analysis is appropriate for the present study since it allowed us to further examine the hidden structure of the data set and provides a visual presentation of the pattern of similarities (Kruskal & Wish 1978; Cox & Cox 2000) . To perform the analysis, we coded each indicator as 1 if states have that element and 0 otherwise to transform the dataset to a 49 x 14 matrix. Then following previous literature (Torgerson,1956; Machado & Mata 2015), we conducted MDS analysis using on a Euclidian distance model and plotted a 2-dimensional solution to present similarities in PI calculation metrics among 48 states and D.C in the U.S. We used SPSS 22.0 for all the analysis and RStudio to plot all the hexagon maps in the results (hrbrmstr, 2015). In the hex map, equal-sized hexagons represent the states so as to eliminate the visual information of geographic size when it is not related to the information portrayed (Carr et al., 1992, 2000). Additionally, in the present study, a hex map allows us to see clearly the northeastern states in the U.S.

RESULTS:

The purpose of this report is to summarize the key elements of school and district level Performance Index scores (PI scores) for the 50 states and the District of Columbia (D.C) across the United States. In the following section, we first report descriptive results to provide an overview of PI score calculation metrics across the states in the U.S and illustrate the relationship between PI scores and state accountability systems. Second, we categorize the fourteen elements into seven categories proposed by the most recent ESSA regulations (81 FR 34539 §200.14-16, 2016) and briefly interpret each element and how each state integrated each element in their PI score calculations. Third, we report the MDS analysis to examine the similarities and hidden structure in the 48 states' and D.C PI calculation metrics.

Descriptive Statistics

Overview

Among the 50 states in the United States, 48 states and the District of Columbia have single or multiple continuous performance index scores to measure school or district level performance. Among them, 4 states (Illinois, New Jersey, Texas and Wyoming) and the District of Columbia have multiple continuous index scores rather than a single variable. Additionally, two states including Vermont and Montana do not have such PI score calculation metrics. These two states use various separate indicators to determine school and school district AYP status. For example, AYP status of schools and districts in Vermont is determined by indicators including performance on mathematical achievement tests, reading achievement tests, student participation rates and graduation rates separately (Vermont Adequate Yearly Progress , 2016). In order to meet AYP status, schools and districts in Vermont must meet all the criteria for each indicator. Similar to Vermont, Montana (Furois, 2013) determines AYP status of schools and districts based on students' performance on mathematical and reading tests, attendance rates, test participation rates and cohort graduation rates separately. Schools and districts have to meet all the criterions for each indicator to make AYP status. Thus, for these reasons, we excluded Montana and Vermont from the following analysis and analyzed PI calculation metrics for the rest of the 48 states and D.C.

Given the composition of PI elements and descriptions of how each state uses a performance index score in their accountability system, for the states which have overall PI scores, from our analysis of the data we conclude that states use PI scores for two major functions. First, 34 states use PI scores to determine school or district overall ratings, proficiency levels or AYP status. Thus, for these states, a PI score is a summative and comprehensive rating score to compare schools or districts within the specific state so as to identify priority or focus schools as required by their NCLB waiver. Second, states such as Ohio calculate a PI score mainly to measure one of the required indicators by NCLB waiver- students' academic performance-to partially fulfill the waiver agreement. In addition to PI scores, Ohio also reports other required educational outcomes including closing gaps, progress indicators, postsecondary school and career readiness in their accountability system. The Ohio SEA takes account of all the above indicators together with PI score to determine final Ohio school or district designation letters and identify priority schools to which the SEA will provide special support (ODE,2015).

The scale of the PI scores differs for each state. The smallest range of school or district performance score is used by the state of Tennessee which scales from 0-4 and the largest range of PI score is used in Mississippi high schools, which scales from 0-1000. Additionally, as mentioned above, 34 states have a summative school or district PI score and used it to determine school or district ordinal ratings, designation letter or AYP status. For example, South Dakota adopted a school performance index (SPI) metric ranging from 0-100. Each school will have an SPI score and schools with SPI scores at or above the top 5 percent of schools are classified as Exemplary schools; schools with SPI scores at or above the top 10 percent of schools are categorized as Status

Table 1: Descriptive Statistical Results Aggregated by Elements for PI Score Calculations

Elements	Number of States with element (%)
Test Score	49 (100)
Growth Indicator	46 (93.87)
Graduation Rate	45 (91.84)
Closing Achievement Gap	42 (85.71)
Test Participation	32 (65.30)
Post School Readiness	30 (61.24)
Career Readiness	27 (57.14)
Attendance Rate	20 (40.81)
Weight by Proficiency	13 (28.57)
Dropout Rate	10 (22.45)
Educator Effectiveness and School Environment	8 (16.32)
Weight by Subject	6(14.29)
English Language Indicator	6 (12.24)
Arts or Humanities	5 (10.20)

Note: Table 1 summarizes the count number and percentage of each element included in the 48 states and D.C calculation metrics. Test score is the most widely used indicator followed by growth indicator, graduation rate and close achievement gap.

schools; schools with SPI scores less than top 10 percent and greater than bottom 5 percent of schools are classified as Progressing schools; schools with SPI scores lower than bottom 5 percent of schools are classified as Priority schools. The table presented in the appendix summarizes the characteristics of each metric for 48 states and DC.

Examining Results Aggregated by Elements

We analyzed results aggregated by PI score elements to summarize the count number and percentage of each element included in the 48 states and D.C calculation metrics in Table 1**.**

The "test score" element is counted most often in the calculation metrics across the states, with all the 48 states and D.C using it in their calculations, followed by the elements of growth indicator, graduation rates and methods designed to reduce the achievement gap between historically underperforming subgroups and their counterparts. The element of arts and humanities is the least counted in the calculation metrics with only five states taking it into account for the PI score calculation. We discuss each element in turn below.

Examining Results Aggregated by States:

We aggregated the results by states to further explore the count number and proportion of elements each state used to calculate the Performance Index score. Table 2 summarizes the results aggregated by states. On average, states include 8 elements in calculating PI in their accountability systems, with a minimum of two and a maximum of ten Alabama, Louisiana and Nebraska and use the most elements (10 elements) in their calculation metrics of school or district level performance index scores. Kansas counts 2 elements in their calculation metrics, the smallest number of elements used.

Additionally, to visualize the results, we created a hexagon map to represent the percentages of elements counted in each state, with darker shades indicating higher percentages in Figure 1. In the figure, the shades indicated the proportions of elements counted in each state's PI calculation metrics with the darker shades corresponding to higher percentages of elements used in the state performance index score calculation metrics.

Since Vermont and Montana do not have a PI score of the type under discussion here, their shades are the lightest. Out of all 50 states and DC, Alabama, Louisiana and Nebraska have the darkest shades indicating the most elements used in their PI calculation metrics.

Table 2: Descriptive Statistical Results Aggregated by States

State	Number of Data Elements	Percentage all elements included%
Alabama	10	71.43
Louisiana	10	71.43
Nebraska	10	71.43
Arizona	9	64.29
Connecticut	9	64.29
Illinois	9	64.29
Mississippi	9	64.29
Missouri	9	64.29
Texas	9	64.29
Wisconsin	9	64.29
Alaska	8	57.14
Colorado	8	57.14
Delaware	8	57.14
Florida	8	57.14
Georgia	8	57.14
Hawaii	8	57.14
Idaho	8	57.14
Kentucky	8	57.14
Maryland	8	57.14
Massachusetts	8	57.14
Nevada	8	57.14
New Mexico	8	57.14
Oklahoma	8	57.14
Pennsylvania	8	57.14
South Dakota	8	57.14
Indiana	7	50.00
Iowa	7	50.00
Michigan	7	50.00
North Carolina	7	50.00
North Dakota	7	50.00
Virginia	6	42.86
Arkansas	6	42.86
Minnesota	6	42.86
New Hampshire	6	42.86
Washington	6	42.86
Wyoming	6	42.86
Oregon	5	35.71
New Jersey	5	35.71
New York	5	35.71
Rhode Island	5	35.71
South Carolina	5	35.71
Utah	5	35.71
West Virginia	5	35.71
District Columbia	4	28.57
Maine	4	28.57
Tennessee	4	28.57
California	3	21.43
Ohio	3	21.43
Kansas	2	14.29

Note. Table 2 summarizes the results aggregated by states. On average, states include 8 elements in calculating PI in their accountability systems, with a minimum of 2 and a maximum of 10.

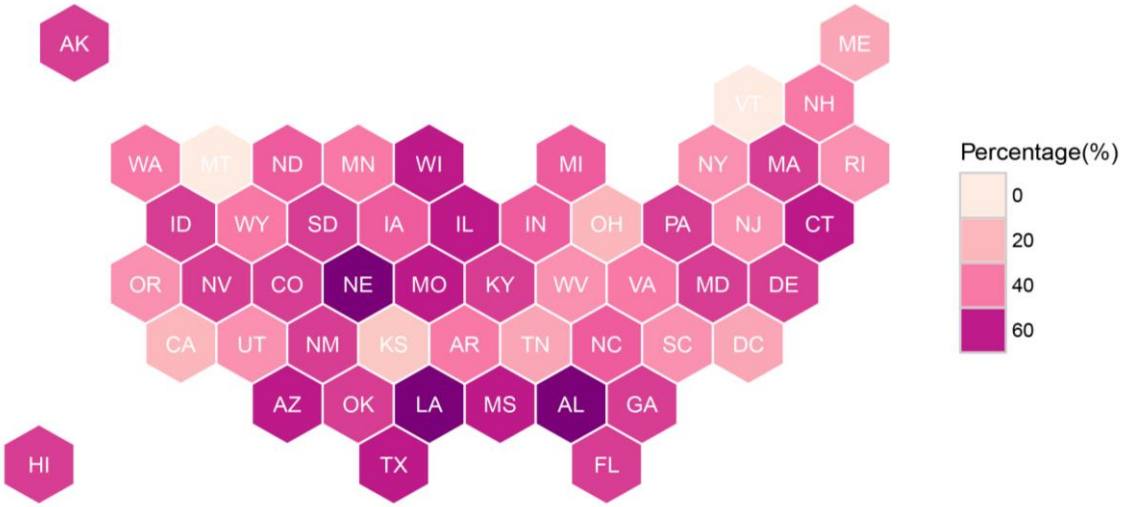


Figure 1: Hexmap of Total Percentages of Elements Included in Each State's Calculation Metric for PI Scores across the 50 states and DC.

Note. Equal-sized hexagons are used to represent the states to avoid visual bias with data that do not relate to the size of the state. Darker shades correspond to higher percentages of elements used in the state performance index score calculation metrics. Maximum=71.43% (Alabama, Louisiana, Nebraska), minimum=14.29% (Kansas).

Elements Description

Academic Achievement Elements

Test Score

The test score element in Table 1 includes K-12 students' performance on large-scale standardized tests administered by each state. All the states take this element into account in calculating school or district performance index scores. The subjects of the test scores vary from different states. Some of the states such as Nevada only accounted for the test performance of mathematics and English reading as required by NCLB for AYP. However, other states, such as Ohio and California, used all of the tested subjects including mathematics, writing, reading, social studies and science in consideration of the calculation of the district performance index score. All staes measured subjects required at the school-level – mathematics and English language arts for all students (81 FR 34539§200.33, 2016). According to the recent CAP report (Martin, Sargrad, & Batel, 2016) of states measuring other subjects, twenty-nine states measure science; four states measure writing and twelve states measure social studies. Most states calculate this element by pooling the percentage of meeting standard for all tests into the PI calculations, while other states assign different weights to percentages of different proficiency tiers before integrating this element into calculating the final PI.

Weight by Proficiency Level

Thirteen states use student performance levels for each subject in PI calculations. By using proficiency level weights, some states award high performance while other states underweight low performance. For example, Ohio calculates the final district performance index score by using a multiplier of 1.2 to the percentage of students who achieved an advanced level on the state standardized tests and multiplied 0.3 by the percentage of students below the proficiency level to calculate the final district performance index score (Kucinski, 2007). As a second example, Missouri assigned 16 points to the advanced performance on the test score of each subject and assigned 0 points for not meeting the proficiency level (Missouri ESEA Flexibility Request, 2015).

Weight by Subject

Six states use different weights on different course subjects when calculating the performance index score. As an example, the California SEA multiplies 0.48 by the reading proficiency percentage, 0.32 by the mathematics proficiency percentage and 0.2 by the writing and science proficiency percentage to calculate the weighted average performance score-Academic Performance Index (API) (CA Dept of Education). Similarly, Oregon calculated the weighted average student test score with reading and mathematics'

weight of 0.39 and writing's weight of 0.22 (Oregon ESEA Flexibility Request, 2015) .

Growth indicator

We defined the growth indicator as the students' improvement on test performance over constitutive years. The growth indicator is one of the required elements by ESSA regulations. In total, 45 states and D.C calculate this element for their current PI score calculations. Different states calculated this element using different models. After conducting a thorough review of those states which incorporate this element into their PI calculations, we summarize three major methods that states use to measure performance growth including 1) annual progress model; 2) student growth percentile (SGP) or adequate growth percentile (AGP) model, and 3) value-added models.

Annual progress model is an improvement model which provides a calculation of school or district progress in PI score points, test scores and other educational outcomes (Goldschmidt et al.,2005). For instance, Minnesota measured student improvement by current student performance on the Minnesota Comprehensive Assessments to student performance in the most recent test (Minnesota ESEA Flexibility Request, 2015). Each student receives a growth score, and each school then receives an aggregated growth score based on the average growth of all students in the school. Alaska, as another example, first assigned progress and proficiency scores to individual students according to one-year progress then calculated the average for each school or subgroup (Alaska ESEA Flexibility Request, 2015). Another example is Illinois, which used a value-table model created by their own to measure annual progress. The value-table model establishes performance categories and awards points to individual students based on their growth between performance categories on statewide achievement over two years. Each student is assigned a growth score based on how much student performance increased from last year with larger increases being assigned higher scores. The individual student scores are averaged for all the scores for a school or a district to obtain a growth score. Thus faster-paced progress earns higher scores and slower-paced progress receives lower scores (Illinois ESEA Flexibility Request, 2015). Given the above calculation methods, a value-table model is actually a weighted annual progress model. Of states taking this element into consideration for PI score calculations, a total of twenty-two states measure growth indicators by annual progress.

A second way to measure school or district growth is a student growth percentile (SGP) model or adequate growth percentile (AGP). The SGP model measures the amount of growth a student makes relative to peers (Betebenner, 2009). Growth for schools or districts is calculated as the median growth percentile of all students. For example, the basis of the growth component is the West Virginia Growth Model, which calculates a student growth percentile—a descriptive estimation of how much growth has occurred for a given student when compared with students across the state with similar prior academic scale scores. Similar to the SGP model, the AGP model is calculated as the median percentile growth for up to three consecutive years. This calculation is also known as the Colorado growth model (Bonk et al., 2012). A total of nineteen states and D.C used SGP or AGP model to measure schools or districts growth.

The third major method of measuring the growth indicator is to use a value-added model (VAM). Five states including Wisconsin, Pennsylvania, New Mexico, Tennessee and North Carolina use this model to measure students' growth. A value-added model first calculates the amount of growth expected for a school and the amount of growth the school actual gains. The difference between the expected growth and actual growth is the "value" schools or districts added (Andrejko, 2004). For example, Pennsylvania used a VAM to measure school and district cohort growth from year to year for reading and mathematics tests, known as the Pennsylvania value-added assessment system (PVAAS). Each school and district will have a PVASS index score scaled from 50 to 100, which is weighted 40% in the final PI score calculations (Pennsylvania ESEA Flexibility, 2015).

To visualize the distribution of these three growth measures, including the annual progress model, SGP or AGP model and value-added model used for each state across the U.S, we created a set of three hexagon maps to summarize student growth indicator measures by states in Figure 2. As shown in Figure 2, the colored hexes (red, blue, yellow) indicate the extent that each state uses each model to measure student growth.

English Language Indicator

The English language indicator is a new requirement by the recent ESSA proposed regulations (81 FR 34539 §200.14(b)(4), 2016). Currently, a total of six states – Arizona, Colorado, Georgia, Illinois, Massachusetts, and Texas – incorporated this measure in PI calculations. Texas, for example, adopted English language learner (EL) progress to measure English language learner proficiency progress in Texas standardize tests. Each EL student was placed to a one-to-four year plan according to the number of years the student has been enrolled in U.S. schools and the performance on the Texas English Language Proficiency Assessment System (TELPAS) composite proficiency levels the first time the student takes the test. Then the Texas Education Agency sets year-to-year cut score expectations on each standardized test score domain. EL students receive credits for meeting their year-to-year plans. The individual

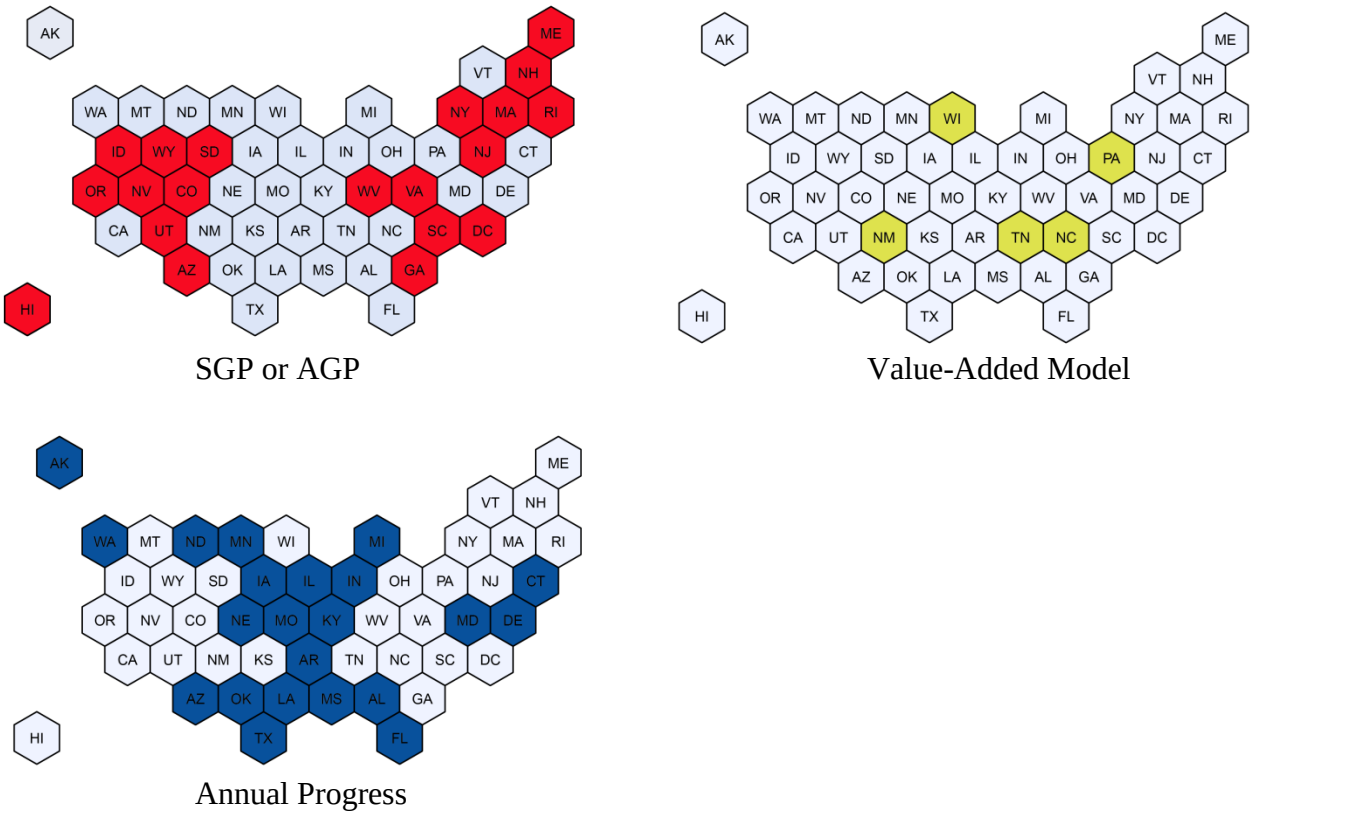


Figure 2 Student Growth indicators by state

Note: Equal-sized hexagons are used to represent the states so as to avoid bias because of state sizes. In the figure, the colored (red, blue, yellow) hexes indicate that the states use the model labeled below to measure student growth.

student scores are averaged for all of the scores for a school or a district to obtain a growth score (Texas ESEA Flexibility Request, 2015). Additionally, Georgia and Arizona measured EL progress based on performance on the Assessing Comprehension and Communication in English State-to-State for English learners exams (ACCESS for ELs). For Georgia (Georgia ESEA Flexibility Request, 2015), it measures the percentage of ELs increasing to a higher performance band on the test determined by the state. In comparison, in Illinois, the percentage of ELs making state-determined progress on ACCESS for ELs into the calculation (Illinois ESEA Flexibility Request, 2015).

Graduation Rates

The Graduation rate is the third most used element calculated in the metrics for different states. There are 45 states that include graduation rates into PI calculations, but all the 49 states and D.C. in the analysis report graduation rates in their current accountability systems. According to ESSA regulations (81 FR 34539 §200.34, 2016), all states are required to include the four-year adjusted cohort graduation rate in determining school overall summative ratings or proficiency levels. Additionally, states could also add the extended year adjusted cohort graduation rates if states plan to use the extended year cohort adjusted graduation rates in their accountability systems. But the ESSA regulations require the goals based on the extended year cohort adjusted graduation rate to be more rigors than the four year adjusted cohort graduation rate. The calculation guidelines for the cohort adjusted graduation rates are described in ESEA flexibility documents (USDOE, 2008). Some of states like Arkansas use four-year graduation rates (Arkansans ESEA Flexibility Request, 2015), while others such as Texas takes the average of four-year and five-year cohort graduation rates into consideration in the calculations (Texas ESEA Flexibility Request, 2015). As example, Alaska schools receive points based on either four or five- year adjusted cohort graduations rates,

whichever result in higher point according to the state pre-determined tiers. But the state set higher standards for the five-year adjusted cohort graduation rates, compared with the four-year adjusted cohort graduation rates. For instance, schools will attain full scores on the graduation rate indicator if their four-year adjusted cohort graduation rates are larger than 80% but have to reach 85% for five-year adjusted cohort graduation rates. Graduation rates is not only required by the recent proposed ESSA regulations (81 FR 34539 §200.14,2016) but also a standard measure of students' completion of high school and students' persistence in high school coursework (Alaska ESEA Flexibility Request, 2015).

School Quality or Students Success Elements

School quality or student success is a new category amended by ESSA (81 FR 34539 §200.14,2016). States are required to include one or more of the following: student access to and completion of advanced coursework; postsecondary readiness; school climate and safety; student engagement and educator engagement. According to the descriptions of above-mentioned indicators, the following seven elements appeared in PI calculation metrics across the states.

Postsecondary School Readiness

Unlike graduation rates, postsecondary school readiness measures how well students are prepared for postsecondary education. Variables included in this element include average ACT or SAT score, the percentage of students taking AP courses, average college admission rates and average GED test participation and among others. There are 30 states that use this element in their PI score calculation. For example, states such as Pennsylvania measured postsecondary school readiness by taking the percentage of students who take PSAT/PLAN participation rates into consideration in their calculations (Pennsylvania ESEA Flexibility Request, 2015). Oklahoma assigned bonus points for participation and performance in advanced coursework, college enrollment and industry certification courses (Oklahoma ESEA Flexibility Request, 2015).

Career Readiness

Similar to post-secondary school readiness, career readiness measures students' preparation for careers. States use different methods to include career readiness into calculations. Louisiana, for example, included average scores on the career readiness test developed by states such as Louisiana and percentages of students getting a specific career certificate into the calculation (Louisiana ESEA Flexibility Request, 2015). Alaska, as another example, assigned points to student proficiency levels and participation rates on WorkKey Certificate tests (Louisiana ESEA Flexibility Request, 2015). Tewnty0seven states that account for career readiness in their PI score calculations.

Attendance Rate

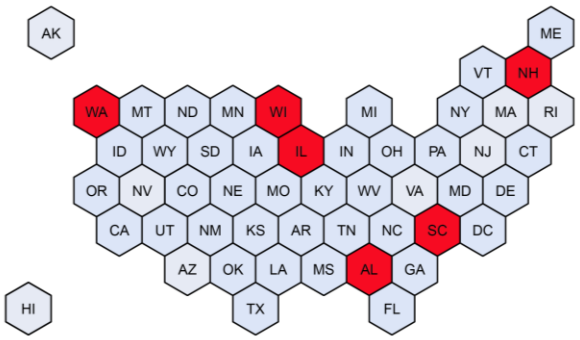
Twenty states used attendance rates in their PI score calculation. Some states, such as Washington (Washington ESEA Flexibility Request, 2015), assigned a specific amount of credits to this element, such as a score of five when the attendance rate is larger than 95%. Other states such as Alaska will use the actual attendance rate and weight it by 0.25 in the overall school or district performance index score.

Drop Out Rate

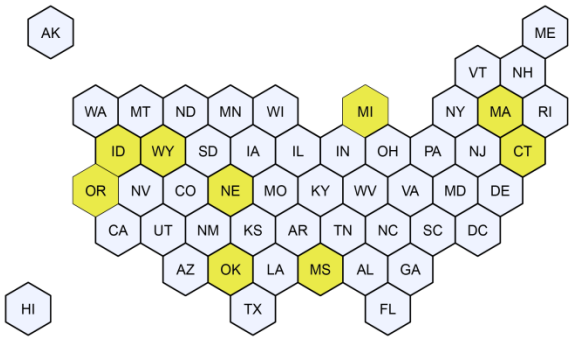
There are ten states which count dropout rates in the calculation metrics for school or district level PI scores. For many of these states, schools or districts are deducted points for high dropout rates in PI calculations. For example, in Colorado, schools are assigned different points based on state-determined dropout rates tiers. If a school's dropout rate is less than 1%, it will earn full points for the dropout rate indicator. Otherwise schools will be deducted points for dropout rates higher than 1% (Colorado ESEA Flexibility Request, 2015). Other states like Wisconsin deduct specific credits (6 credits) directly from the total score if the district dropout rate is higher than a specific criterion (6%) (Wisconsin ESEA Flexibility Request, 2015).

Educator Effectiveness and School Environment

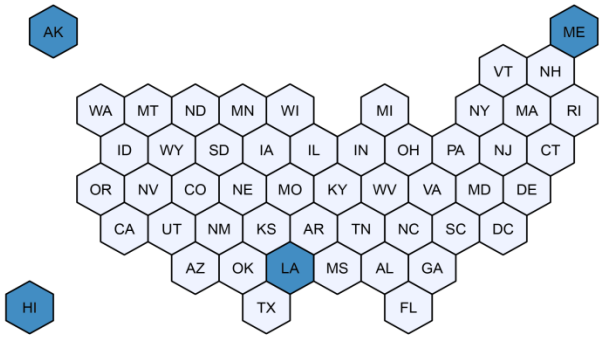
Eight states take this element into consideration for the calculation of the district or school level performance. Educator effectiveness measures the effectiveness of the school leadership and teachers, indicators such as the percentage of highly qualified teachers, or percentage of teachers with advanced degrees are used. School environment includes variables measuring school learning climate, academic expectations and other aspects. For example, one component in the Kentucky PI score calculation is Next-Generation Professional index, which consists of multiple measures of leadership, instruction, learning climate and assessment practice. The Next-Generation Professional index weights 10 percent in the final Kentucky PI score (Kentucky ESEA Flexibility Request, 2015). As another example, New York assigned 15 out of 100 points to the School Environment index, which is measured by the NYC school Survey. The school environment index includes four aspects including academic expectation, communication, engagement and safety and respect (NYDOE, 2013).



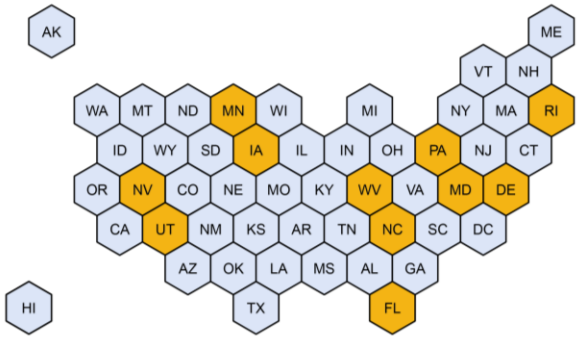
Deduct Points



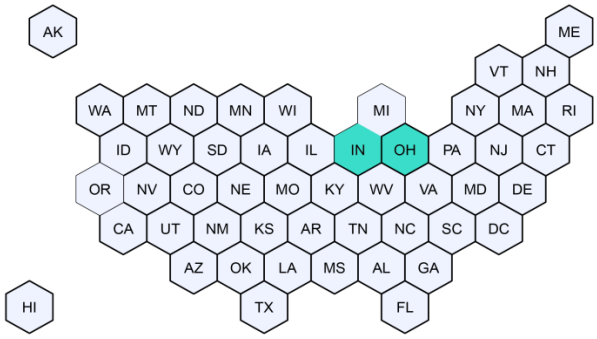
Classifying to a Lower Rating Level



Classifying Non-Test-Takers
to Non-Proficient Groups



Classifying to Focus Schools
or Gate Keeper



Weight by Test Participation Rates

Figure 3 Test participation rates by state

Note: Equal-sized hexagons are used to represent states so as to avoid bias because of state sizes. In the figure, the colored hexes indicate that the states use the method labeled below to include test participation rates into calculations.

Arts or Humanities

Arts or Humanities measured students' performance on subjects in history, philosophy, arts and subjects that were not examined as part of state standard test. Five states including Alabama, Connecticut, Georgia and Kentucky and Mississippi account for this element into PI calculation

For instance, one component of Kentucky PI score is named Next Generation Instruction and Support, which assign credits to the access of arts and humanities courses (Kentucky ESEA Flexibility Request, 2015). Connecticut, as an another example, measured art access by percentage of students in grade 9 through 12 participating in at least one dance, theater, music or visual arts course in the year.

Closing Achievement Gaps

There are 42 states that include closing achievement gap into their PI calculations. For these states' PI score calculations and reward positive progress in closing achievement gaps. Pennsylvania, as an example, the achievement gap is determined by comparing the percent of students who are proficient or advanced in baseline year with 100% proficiency in all the tested subjects. Once the achievement gap is determined, schools are measured on the success in closing that gap based on the preset state benchmark. In comparison, Illinois assigned different points based on whether school meet state-determined gap reduction targets for historically underperforming groups including racial and ethnic minorities, economically disadvantaged, English learners and students with disabilities in each school.

Test Participation Rate

Thirty-two states use test participation rates in some form for the PI score calculations. According to how states incorporate test participation rates into PI calculations, we summarized five ways of using this element including deducting points, classifying schools with less than 95% test participation rates to lower ratings, assigning non-test takers to non-proficiency level or a zero score, classifying schools directly into focus schools if they do not meet test participation standards and weight by test participation rates. Six states including Washington, Wisconsin, Illinois, Alabama, South Carolina and New Hampshire deduct points directly if states fail to meet test participation rates of 95% required by ESSA(81 FR 34539 §200.15,2016). Alabama, for instance will deduct points directly from the total PI score if test participation rates is less than 95% (Alabama ESEA Flexibility Request, 2015).

The second way of including test participation rates in accountability systems is to classify overall school designation ratings to lower designations if schools failed to achieve 95% test participation rates. Wyoming for example, when a school does not meet the participation threshold, the school is considered as not meeting expectations and a lower designation letter grade will be assigned to the school. A total of nine states will classify overall school designation ratings to a lower level if schools' test participation rates are lower than 95%.

Similar to this method, eleven states are more rigid in accounting test participation rates by classifying schools directly into focus or priority schools if they failed to meet the test participation standard of 95%. Another example of West Virginia, in calculating the West Virginia Accountability Index, test participation rates serve as "on/off switches" in the sense that 95% test participation rates will be minimum requirements for all schools. If a school fails to meet the 95% participation criteria for any valid subgroup, it will automatically be identified as a Support School. In Rhode Island, if a school fails to test at least 95% of its students in either reading or math, it is classified as a "Warning School," at best, regardless of the Composite Index Score (Rhode Island ESEA Flexibility Request, 2015).

The fourth method is to classify non-test takers directly to the "not meet" standard category or the non-proficiency group if school test participation rates are lower than 95%, a total of four states including Hawaii, Alaska, Maine and Iowa use this method. For instance, Hawaii proposes to maintain the 95% for participation rate all students and disaggregated subgroups as annual measure objectives for reading, mathematics and science tests (Hawaii ESEA Flexibility Request, 2015). A non-proficient outcome will be assigned to any non-participant in schools not meeting the 95%test participation rates.

The fifth method to include test participation rates in state accountability system is to multiply the actual test participation rates with PI scores. Thus, the lower the test participation rates, the lower the PI scores. Two states including Ohio and Indiana use this method.

To visualize distribution of these five measures including deduct points, classifying to a lower rating level, classifying non-test takers to non-proficient groups, classifying to focus schools or gatekeeper and weight by test participation rates for each state across the U.S., we plot five hexagon maps to summarize test participation rate measures by state in Figure 3. In the figure, the colored hexes indicate the extent that each state uses each model to measure test participation rates.

Multidimensional Scaling Analysis (MDS)

To examine similarities among different PI calculation methods from the types of elements states counted in their calculation metrics, we followed previous literature (Torgerson,1956; Machado & Mata 2015) and conducted a multidimensional scaling analysis (MDS) with the

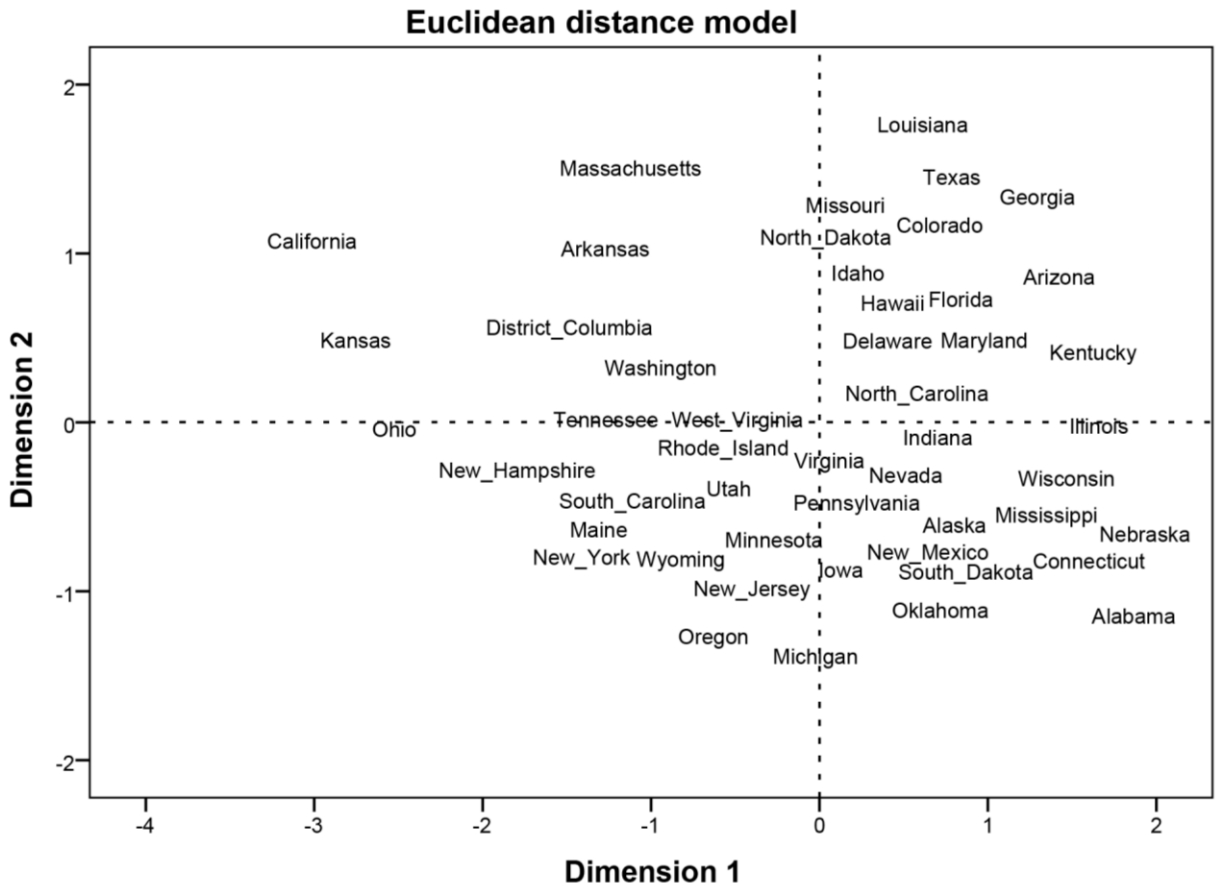


Figure 4 Multidimensional Scaling (MDS) stimulus configuration for similarities of calculation metrics across 48 states and DC with single performance index scores.

The horizontal dimension represents the number of elements in each state calculation metric while the vertical dimension represents whether states integrate closing achievement gaps into the final PI score. The closer the two states the more similar their elements in the performance index score calculation.

Euclidean distance model. The two-dimensional measure fit the data well with R-square equal to 0.799, explaining 79.9% of the variance in the data set. We then plotted each state in 2-dimensions (see Figure 4).

The multidimensional scaling in Figure 4 provides a visual representation of the complex set of relationships between the different metrics across the states. In Figure 4, the closer the two states in the two dimensions, the more similar the elements counted in their calculation metrics of performance index scores. For example, the similarities of calculation metrics between Kansas and Alabama are smaller than that between Alabama and Michigan. From Figure 4, we can conclude that states have a large diversity in calculating PI scores given the elements used in their performance index scores. Additionally, by further exploring the states' calculation metrics for the two dimensions, we came to the following findings. First, for dimension 1, states on the right side have more elements in their PI score calculations and states on the left side have fewer elements in calculating PI scores. For dimension 2, most states in the bottom side of the MDS plot incorporate attendance rates in their calculation metrics. Conversely, most states on the upper side of the plot in Figure 4 do not take this element into account in their PI score calculations.

DISCUSSION:

The results of the descriptive statistics and multidimensional scaling analysis bring us to five central conclusions. First, as shown across the figures and tables, there are few commonalities in PI score calculations across the states, as each state appears to have forged their own path forward in addressing the requirements of NCLB and now ESSA. Indeed, our analysis shows that no two states use exactly the

same system to calculate their PI score metrics. Second, among the 50 states and DC in the U.S., 49 of them have their own PI score to measure the school or district level performance. Third, 14 elements could be extracted from the 49 calculation metrics, and the number of elements counted in the performance index score calculation metrics varied across the states. For the states with such metrics, Louisiana, Nebraska and Alabama incorporated ten elements in their calculation metrics, which is the most. Kansas counts two elements in their PI calculations, which is the least. Fourth, from the descriptive statistic results aggregated by elements, we can conclude that the test score, growth indicator, graduation rates and closing achievement gap are the top four elements included across the states, while arts and humanities is the least counted element in these calculations. Finally, the results of the multidimensional scaling analysis suggest that there is large diversity in calculating PI scores across the states, indicating that contents or constructs of PI scores vary substantially from each state.

From the above conclusions, we cannot justify which metric may be "the best" given the multiple different indicators included in the calculations. However, PI scores including more indicators might provide more detailed potential guidelines for in-depth qualitative studies and more information for practitioners and researchers to identify "effective" practices. For instance, outperforming schools in Alabama may outperform their peer schools in aspects of improving students' test score, graduation rates, promoting students' improvement and closing the achievement gaps, higher attendance rates, more effective educators. Additionally, the traditional percentage of proficiency in specific tests under AYP accountability systems has been criticized for its limited and unrepresentative depictions of large-scale test score trends and lack of recognition for low status but high progressing schools (Ho, 2008). PI scores summarized in the present studies may provide a means for addressing these types of issues. As a distribution-wide score, PI scores allow researchers and practitioners to report other statistics such as standard error, percentile and effect sizes to better reflect the overall distribution and trend of schools or district performance within a state. Thus, while a PI score is an informative indicator for researchers and practitioners to compare schools or districts within the same state, we encourage states to not only report the overall proficiency levels for schools and districts but also report the actual score schools or districts gain for each year.

As states moving forward in implementing ESSA, a summative rating for schools overall performance is required by the ESSA regulations (81 FR 34539 §200.18, 2016). The present report is especially informative for states that are now developing indicators for school accountability to fulfill the ESSA requirements since we examine the alignment of 14 elements in PI score calculations with ESSA requirements nationwide. For states that include all of the required indicators by ESSA regulations in their PI calculations they may potentially consider incorporating the use of PI scores as one of many possible performance criteria to determine a school or district overall designation, letter grade or proficiency level. For states that include partial required indicators by ESSA in PI calculations, these states may potentially use PI scores as one composite of summative ratings while developing specific plans that may include the states' own long-term objective and expectations for holding schools and districts accountable. For example, for states setting a long-term goal as improving overall college-entrance rates, these states could give additional weight to a post-secondary readiness indicator. As a different example, states with a large percentage of English learners could choose to weight more on an English language indicator.

However, researchers and practitioners should standardize PI scores before comparing school or district year-to-year overall performance within the same state. Additionally, PI scores calculated by different states in their current accountability systems cannot easily be compared across states. In addition to no two states using the same calculation metrics for their PI scores, PI scores for each state are reported in ordinal scales. Within states, this allows a state to potentially rank order schools or districts, but precludes the ability to compare relative positional differences between states. For ordinal scale scores, the intervals between two adjacent points are not equal (Cohen, Swerdlik, & Phillips, 1996; Nunnally Jr., 1970). In other words, a one unit change in PI scores across years without standardization within the same state and across different states are not the same. Additionally, different states conduct different standardized tests and incorporate various elements in calculating PI scores. Therefore PI scores for schools and districts from different states are not on the same scale, and thus are not comparable.

Recently, to address these issues, a series of studies funded by the US Department of Education Institution for Education Sciences and conducted by the Center for Education Policy Analysis (CEPA) at Stanford University examined this issue of comparing standardized test scores for districts across states. Reardon et al. (2016) conducted a linear equating procedure (Kolen & Brennan, 2014) to link each of the 50 states' standardized achievement test scores to the state's corresponding performance on the National Assessment of Education Progress (NAEP) scale. This analysis is one of the first to provide a means to compare NAEP scale scores across districts and states over multiple years. Given the findings of the present report, we argue here that future research and state-level policymaking on school and district PI score calculations should take into

account the variability of these calculations reported here across the states combined with the recent research on the possibility of also equating these types of metrics across states and time at a national level.

Conclusion

This report summarized 14 key elements of school and district level Performance Index scores (PI scores) for the 50 states and the District of Columbia (D.C) across the United States. The descriptive analysis and multidimensional scaling results indicate that there are few commonalities in PI score calculation metrics across the states, as each state has its own methods for addressing the requirements of NCLB. As states move forward towards implementing the recent Every Child Succeeds Act (ESSA) regulations, we recommend states enact PI calculation metrics aligning with each state's long-term goals and expectations for holding schools and districts accountable and that states report PI scores in addition to letter grades or proficiency levels in school and district reports cards.

Suggested Citation Format:

Ni, X., Bowers, A.J., Esswein (2016) What Counts in Calculating School and District Level Performance Index Scores: A Summary and Analysis of Academic Performance Index Metrics across the 50 States. A White Paper Report. New York, NY: Teachers College, Columbia University.

REFERENCES:

81 FR 34539. (2016). Retrieved June 23, 2016, from https://www.federalregister.gov/articles/2016/05/31/2016-12451/elementary-and-secondary-education-act-of-1965-as-amended-by-the-every-student-succeeds

Academic Performance Index (API) (CA Dept of Education). (n.d.). Retrieved June 23, 2016, from http://www.cde.ca.gov/ta/ac/ap/

Alaska ESEA Flexibility. (2015, July 23). [Letters (Correspondence)]. Retrieved June 23, 2016, from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/ak.html

Andrejko, L. (2004). Value-Added Assessment: A View from a Practitioner. *Journal of Educational and Behavioral Statistics*, *29*(1), 7–9.

Arkansans ESEA Flexibility Request, 2015. (n.d.). Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/arrenewalreq2015.pdf

Barton, P. E., & Coley, R. J. (2009). *Parsing the Achievement Gap II. Policy Information Report*. Educational Testing Service. Retrieved from http://eric.ed.gov/?id=ED505163

Betebenner, D. (2009). Norm- and Criterion-Referenced Student Growth. *Educational Measurement: Issues and Practice*, *28*(4), 42–51. http://doi.org/10.1111/j.1745-3992.2009.00161.x

Bloomfield, D. C., & Cooper, B. S. (2003). NCLB: A New Role for the Federal Government: An Overview of the Most Sweeping Federal Education Law since 1965. *T H E Journal (Technological Horizons In Education)*, *30*(10), S6.

Bonk, W., Copa, J., Gibson, N., Gillin, T., Nau, J., Peoples, A. L., … others. (2012). GROWTH MODELS. Retrieved from https://nces.ed.gov/Programs/SLDS/pdf/guide_growth-model.pdf

Borg, I., & Groenen, P. J. F. (2005). *Modern Multidimensional Scaling: Theory and Applications*. Springer Science & Business Media.

Bowers, A. J. (2008). Promoting Excellence: Good to great, NYC's district 2, and the case of a high performing school district. Leadership and Policy in Schools, 7(2), 154-177.

Bowers, A. J. (2010). Toward Addressing the Issues of Site Selection in District Effectiveness Research: A Two-Level Hierarchical Linear Growth Model. Educational Administration Quarterly, 46(3), 395-425.

Bowers, A. J. (2015). Site Selection in School District Research: A Measure of Effectiveness Using Hierarchical Longitudinal Growth Models of Performance. School Leadership & Management, 35(1), 39-61.

Bryk, A. S., Sebring, P. B., Allensworth, E. M., Luppescu, S., & Easton, J. Q. (2010). Organizing schools for improvement: Lessons from Chicago. Chicago: The University of Chicago Press.

Carr, D. B., Olsen, A. R., & White, D. (1992). Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data. *Cartography and Geographic Information Systems*, *19*(4), 228–236.

Carr, D. B., Wallin, J. F., & Andrew Carr, D. (2000). Two new templates for epidemiology applications: linked micromap plots and conditioned choropleth maps. *Statistics in Medicine*, *19*(17–18), 2521–2538.

Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement (3rd ed.)* (Vol. xxviii). Mountain View, CA, US: Mayfield Publishing Co.

Cox, T. F., & Cox, M. A. A. (2000). *Multidimensional Scaling, Second Edition*. CRC Press.

Dunlap, A. (2011). *ESEA Flexibility: Department of Education Criteria and State Responses*. Mid-continent Research for Education and Learning (McREL). Retrieved from

EducatorGuide_EMS_20131118.pdf. (n.d.). Retrieved from http://schools.nyc.gov/NR/rdonlyres/7B6EEB8B-D0E8-432B-9BF6-3E374958EA70/0/EducatorGuide_EMS_20131118.pdf

Elementary and Secondary Education Act (ESEA) - Oregon Department of Education.

Elmore, R. F., & Burney, D. (1999). Investing in teacher learning: Staff development and instructional improvement. In L. Darling-Hammond & G. Sykes (Eds.), Teaching as the learning profession : Handbook of policy and practice (pp. 263-291). San Francisco: Jossey-Bass.

Federal Register | Elementary and Secondary Education Act of 1965, As Amended by the Every Student Succeeds Act-Accountability and State Plans.

Freeman, Y. S., & Freeman, D. E. (2002). *Closing the Achievement Gap: How To Reach Limited-Formal-Schooling and Long-Term English Learners.* Heinemann, 88 Post Road West, P.O. Box 5007, Westport, CT

Georgia ESEA Flexibility Request, 2015.). https://www2.ed.gov/policy/eseaflex/ga.pdf

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., & Williams, A. Policymakers' Guide to Growth Models for School Accountability: How Do Accountability Models Differ? Retrieved from http://www.ccsso.org/Documents/2005/Policymakers_Guide_To_Growth_2005.pdf

Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under no child left behind. Educational Researcher, 37(6), 351-360.

Herdman, P. A. (2002). Understanding the Basic Bargain: A Study of Charter School Accountability in Massachusetts and Texas.

Horm-Wingerd, D. M., Winter, P. C., & Plofchan, P. (2000). Primary Level Assessment for IASA Title I: A Call for Discussion. Series on Standards and Assessments.

Honig, M. I. (2008). District Central Offices as Learning Organizations: How Sociocultural and Organizational Learning Theories Elaborate District Central Office Administrators' Participation in Teaching and Learning Improvement Efforts. American Journal of Education, 114(4), 627-664.

Honig, M. I., & Venkateswaran, N. (2012). School–Central Office Relationships in Evidence Use: Understanding Evidence Use as a Systems Problem. American Journal of Education, 118(2), 199-222.

hrbrmstr. (2015, May 14). GeoJSON Hexagonal "Statebins" in R. Retrieved July 1, 2016, from https://rud.is/b/2015/05/14/geojson-hexagonal-statebins-in-r/

Kentucky ESEA Flexibility (2015, April 17). [Letters (Correspondence)]. Retrieved June 24, 2016, from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/ky.html

Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling*. SAGE.

Kucinski, S. (2007). *OGT: Ohio Graduation Test in Reading and Writing*. Barron's Educational Series. Retrieved from http://books.google.com/books?hl=en&lr=&id=p80XPJ_bUhcC&oi=fnd&pg=PA27&dq=%22in+the+Accelerated%22+range+will+count+as+if+it+is+in+the+%22Advanced%22+range%22+%22through+eight+(8),+plus+the+English+Language+Arts+and+math+OGT+assessments%22+&ots=uBtE5YpZDU&sig=MXuJ6gmHs96pf6lyVWWFdgYNsGs

Leithwood, K. (2010). Characteristics of school districts that are exceptionally effective in closing the achievement gap. Leadership and Policy in Schools, 9, 245-291.

Levin, B., Datnow, A., & Carrier, N. (2012). Changing school district practices. Boston, MA: Students at the Center: Teaching and Learning in the Era of the Common Core: A Jobs for the Future Project.

Machado, J. A. T., & Mata, M. E. (2015). Analysis of World Economic Variables Using Multidimensional Scaling. *PLoS ONE*, *10*(3).

Martin, C., Sargrad, S., & Batel, S. (2016). Making the Grade: A 50-State Analysis of School Accountability Systems. Retrieve from https://cdn.americanprogress.org/wp-content/uploads/2016/05/17094420/AccountabilityLandscape-report2.pdf

Minnesota ESEA Flexibility Request. (2015, April 17). [Letters (Correspondence)]. Retrieved June 23, 2016, from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/mn.html

Missouri ESEA Flexibility Request. (n.d.). Retrieved from https://dese.mo.gov/sites/default/files/qs-MO-2015-ESEA-Waiver-Renewal-FINAL.pdf

Nunnally Jr., J. C. (1970). *Introduction to psychological measurement* (Vol. xv). New York, NY, US: McGraw-Hill.

ODE,2015 Retrieved from https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Achievement-Measure/Technical-Documentation-PI-Score.pdf.aspx

Okalahoma ESEA Flexibility Request, 2015. (n.d.). https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/okrenewalreq7282015.pdf

Pennsylvania ESEA Flexibility Map Page. (2015, September 8). [Letters (Correspondence)]. Retrieved June 23, 2016, from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/pa.html

Pruitt, P.L., Bowers, A.J. (2014) At What Point Do Schools Fail to Meet Adequate Yearly Progress and What Factors are Most Closely Associated with Their Failure? A Survival Model Analysis. A paper presented at the annual meeting of the Association of Education Finance and Policy, San Antonio, TX: March 2014. http://dx.doi.org/10.7916/D8VQ326P

Purkey, S. C., & Smith, M. S. (1985). School reform: The district policy implications of the effective schools

literature. The Elementary School Journal, 85(3), 352-389.

Rebell, M. A., & Wolff, J. R. (2009). *NCLB at the Crossroads: Reexamining the Federal Effort to Close the Achievement Gap*. Teachers College Press.

Redfield, D., & Sheinker, J. (2004). *Framework for Transitioning from IASA to NCLB. Series Overview*.

Shepard, R. N. (n.d.). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, *27*(2), 125–140.

Singham, M. (2005). *The achievement gap in U.S. education: canaries in the mine*. Lanham, Md: Rowman & Littlefield Education.

Texas ESEA Flexibility (2015, September 29). [Letters (Correspondence)]. Retrieved June 24, 2016, from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/tx.html

Tomlinson, C. A., & Jarvis, J. M. (2014). Case Studies of Success Supporting Academic Success for Students With High Potential From Ethnic Minority and Economically Disadvantaged Backgrounds. *Journal for the Education of the Gifted*, *37*(3), 191–219.

Torgerson, W. S. (n.d.). Multidimensional scaling: I. Theory and method. *Psychometrika*, *17*(4), 401–419.

USDOE.2008. (n.d.). Retrieved from https://www2.ed.gov/policy/elsec/guid/hsgrguidance.pdf

Trujillo, T. (2013). The Reincarnation of the Effective Schools Research: Rethinking the Literature on District Effectiveness. Journal of Educational Administration, 51(4), 426-452.

Vermont Adequate Yearly Progress (AYP). (n.d.). Retrieved June 23, 2016, from http://education.vermont.gov/data/accountability/faqs#lea_accountability

West Virginia ESEA Flexibility Request. (2015, April 17). [Letters (Correspondence)]. Retrieved June 23, 2016, from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/wvrenewalreq2015.pdf

Appendix

This part of the appendix listed the manuals or guidelines describing how the 50 states and District of Columbia calculate their performance index score.

Alabama:
http://web.alsde.edu/docs/documents/908/School%20A%20Example%20Booklet.pdf#search=school%20performance%20score
http://www2.ed.gov/policy/eseaflex/approved-requests/alapprovalreq.pdf

Alaska:
http://ecs.force.com/mbdata/mbstprofexcL?Rep=arst&st=Alaska
http://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/akrenewalreq2015.pdf

Arizona:
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/azrenewalreq2015.pdf
http://www.azed.gov/eseawaiver/files/2013/09/esea-informational-flyer-09-20-13.pdf

Arkansas:
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/arrenewalreq2015.pdf

California
http://www.cde.ca.gov/ta/ac/ap/

Colorado
http://www.cde.state.co.us/sites/default/files/documents/accountability/downloads/1ccr301-1-june2012.pdf
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/coflexrenewal11192015.pdf

Connecticut
http://www.sde.ct.gov/sde/lib/sde/pdf/nclb/waiver/performance_index_computational_guide.pdf
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/ctrenewalreq2015.pdf

Delaware
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/derenewalreq2015.pdf

District of Columbia
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/dcrenewalreq2015.pdf

Florida
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/flrenewalreq2015.pdf P60
http://schoolgrades.fldoe.org/pdf/1415/SchoolGradesCalcGuide15.pdf

Georgia
http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Accountability/Documents/Accountability%20Resources/2015%20Indicators.pdf
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/garenewalreq2015.pdf

Hawaii
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/hirenewalreq2015.pdf

Idaho
https://sde.idaho.gov/topics/accountability/files/appeals/Star-Rating-Accountability-System-Business-Rules.pdf
https://sde.idaho.gov/topics/accountability/files/appeals/Star-Rating-Accountability-System-Business-Rules.pdf

Indiana
http://www.doe.in.gov/sites/default/files/news/asrpgb1-rp.pdf
http://www.doe.in.gov/sites/default/files/accountability/basic-summary-f_1.pdf

Illinois

https://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/il.html
Iowa
https://www2.ed.gov/policy/eseaflex/ia.pdf
Kansas
http://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/ksrenewalreq2015.pdf
Kentucky
http://education.ky.gov/comm/UL/Documents/Revised%20Approved%20KY%20ESEA%20flexibility%20waiver%20Sept%2028%202012%20final%20version%20mam.pdf
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/ky.html
Louisiana
http://www.louisianabelieves.com/accountability/school-performance-score
https://www2.ed.gov/policy/eseaflex/approved-requests/lareq11192015.pdf
Maine
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/merenewalreq22016.pdf
Massachusetts
http://www.doe.mass.edu/apa/ayp/2013/LEAbrochure.pdf
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/marenewalreq2015.pdf
Maryland
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/mdrenewalreq2016.pdf
http://www.marylandpublicschools.org/NR/rdonlyres/BD02D2CB-A55C-45E4-8707-5CBB53CBF520/34084/SPI_Informational_PowerPoint_101112_.pdf
Michigan
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/mirenewalreq2015.pdf
https://www.michigan.gov/documents/mde/ScorecardGuide_426897_7.pdf
http://www.michigan.gov/documents/mde/Accountability_Scorecards_At-A-Glance_425302_7.pdf?20130802143046
Minnesota
http://education.state.mn.us/mdeprod/groups/educ/documents/hiddencontent/bwrl/mdm0/~edisp/mde034431.pdf
Mississippihttp://ors.mde.k12.ms.us/report/lettergrade.aspx
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/msrenewalreq2015.pdf
Missouri
http://dese.mo.gov/sites/default/files/MSIP%20Accountability-CSIP%20-%20Striving%20for%20Continued%20Improvement%20(Spalty,%20Reese,%20Ricker).pdf
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/morenewalreq2015.pdf
Montana
http://opi.mt.gov/PDF/AYP/2013/2013-AYP-Manual.pdf
Nebraska
https://www.education.ne.gov/assessment/pdfs/NE_School_District_Accountability.pdf
Nevada
http://nspf.doe.nv.gov/Home/Points
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/nvrenewalreq2015.pdf
New Hampshire
http://education.nh.gov/instruction/accountability/ayp/documents/2012_summary_report.pdf
http://www.education.nh.gov/instruction/school_improve/documents/appeal_2007-08_e_m_status_index_rpt.pdf

New Mexico
https://www2.ed.gov/policy/eseaflex/approved-requests/nmrequest12082015.pdf
New Jersey
http://www.state.nj.us/education/pr/1213/80/806036921.pdf
http://www.nj.gov/education/pr/1213/Interpretive%20Guide%202014.pdf
New York
http://www.regents.nysed.gov/common/regents/files/documents/meetings/2012Meetings/September2012/912p12a2.pdf
http://schools.nyc.gov/NR/rdonlyres/7B6EEB8B-D0E8-432B-9BF6-3E374958EA70/0/EducatorGuide_EMS_20131118.pdf
North Carolina
https://www2.ed.gov/policy/eseaflex/approved-requests/nc3req32015.pdf
North Dakota
http://www2.ed.gov/policy/eseaflex/nd.pdf
Ohio
https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Achievement-Measure/Technical-Documentation-PI-Score.pdf.aspx
Oklahoma
http://sde.ok.gov/sde/sites/ok.gov.sde/files/documents/files/AtoFReportCardGuide.pdf
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/okrenewalreq7282015.pdf
Oregon
http://www.ode.state.or.us/wma/data/schoolanddistrict/reportcard/docs/rc_rating_policy_technical_manual_1314.pdfhttps://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/orrenewalreq2015.pdf
Pennsylvania
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/parenewalreq2015.pd
Rhode Island
http://www.eride.ri.gov/eride40/reportcards/13/documents/RI%20Accountability%20System%20-%20Techn%20Bulletin-May%202013.pdf
South Dakota
http://doe.sd.gov/secretary/documents/AccPresen.pdf
Texas
http://doe.sd.gov/secretary/documents/ESEAflex2.pdfTexas
https://rptsvr1.tea.texas.gov/perfreport//account/2013/20130328coe/overview_20130423.pdf
http://www2.ed.gov/policy/eseaflex/approved-requests/txrenewalreq2015.pdf
Tennessee
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/tnrenewalreq2015.pdf P44
Utah
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/utrenewalreq2015.pdf
Vermont
http://education.vermont.gov/data/accountability/faqs
Virginia
http://www.doe.virginia.gov/statistics_reports/school_report_card/accountability_guide.pdf
https://www2.ed.gov/policy/eseaflex/approved-requests/va4req32015.pdf
Washington
http://www.sbe.wa.gov/documents/AccountabilitySystemInitialrecommendationsfinal2.pdf
West Virginia

http://wvde.state.wv.us/esea/support/Documents/Technical%20Fact%20Sheet_Understanding%20the%20WV%20Accountability%20Index.pdf
https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/wvrenewalreq2015.pdf
Wisconsin
http://dpi.wi.gov/accountability
https://apps2.dpi.wi.gov/reportcards/
http://dpi.wi.gov/accountability/report-cards
http://www2.ed.gov/policy/eseaflex/approved-requests/warequestamended022713.pdf
Wyoming
https://www2.ed.gov/policy/eseaflex/approved-requests/wyapprovalrequest4152013.pdf

Appendix B Table: Elements in PI Score Calculations by State.

Note: Vermont, Montana and Porto Rico are not in the table since we do not find a continuous PI score in their current accountability system

The abbreviation in the chart is interpreted as follows:

- SOR: School or District Level
- SOM: Single or Multiple Indicators
- Scale: The Range of the Score
- E1-E14 are the following elements in order left to right: test score, dropout rate, graduation rate, growth indicator, English language indicator, closing achievement gap, weight by proficiency, weight by subject, attendance rate, test participation rate, post school readiness, career readiness, arts or humanities, effective educator and school environment.

State	SOR	SOM	Scale	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
Alabama	Both	Single	1-90 (PII 1) 2-200(PII 2)	X		X	X		X			X	X	X	X	X	X
Alaska	School	Single	0-100	X		X	X		X			X	X	X	X		
Arizona	Both	Single	0-200	X	X	X	X	X	X				X	X	X		
Arkansas	School	Single	0-300	X	X	X	X		X	X							
California	Both	Single	200-1000	X						X	X						
Colorado	Both	Single	0-100	X	X	X	X	X	X					X	X		

Ni, Bowers & Esswein (2016)

Connecticut	Both	Single	850-1350 (elementary-high school) 0-100 %	X		X	X		X			X	X	X	X	X	
Delaware	Both	Single	0-100%	X		X	X		X		X		X	X	X		
District of Columbia	Both	2 Indicators	0-110	X			X		X	X							
Florida	Both	Single	0-100%	X		X	X		X		X		X	X	X		
Georgia	Both	Single	0-100	X		X	X	X	X					X	X	X	
Hawaii	Both	Single	0-400	X		X	X		X		X		X	X	X		
Idaho	School	Single	0-100	X		X	X		X	X			X	X	X		
Indiana	School	Single	0-100%	X		X	X		X				X	X	X		
Illinois	Both	4 indicators	0-100	X		X	X	X	X				X	X	X		X
Iowa	Both	Single	0-100	X		X	X					X	X	X			
Kentucky	Both	Single	0-100	X		X	X		X					X	X	X	X
Kansas	School	Single	0-100	X						X							
Louisiana	Both	Single	0-150	X	X	X	X		X	X	X		X	X	X		
Maine	school	Single	0-100	X		X	X						X				
Massachusetts	Both	Single	0-100	X	X	X	X	X	X	X			X				
Maryland	School	Single	0-100%	X	X	X	X		X				X	X	X		

Ni, Bowers & Esswein (2016)

Michigan	Both	Single	0-100%	X		X	X		X			X	X				X
Minnesota	Both	Single	0-100	X		X	X		X			X	X				
Mississippi	Both	single	700,900 or 1000	X		X	X		X			X	X	X	X	X	
Missouri	Both	Single	0-140 (k12) 0-80 (k8)	X		X	X		X	X	X	X		X	X		
Nebraska	School	Single	0-100%	X	X	X	X		X			X	X	X	X		X
Nevada	School	Single	0-100	X		X	X		X			X	X	X	X		
New Hampshire	Both	single	0-200	X		X	X		X	X		X					
New Mexico	Both	Single	0-100	X		X	X		X			X		X	X		X
New Jersey	School	3 indicators	0-100%	X		X	X							X			
New York	Both	Single	0-100%	X			X		X								X
North Carolina	School	Single	0-100%	X		X	X		X				X	X	X		
North Dakota	School	Single	0-100	X		X	X		X	X				X	X		
Ohio	Both	Single	0-120	X						X			X				
Oklahoma	School	Single	0-110	X	X	X	X					X	X	X	X		
Oregon	School	Single	0-100%	X		X	X					X	X				
Pennsylvania	School	Single	0-107	X		X	X		X			X	X	X	X		
Rhode Island	School	Single	20-100	X		X	X		X				X				
South Carolina	Both	Single	0-100	X		X	X		X				X				
South Dakota	Both	Single	0-100	X		X	X		X			X		X	X		X
Texas	Both	4 indicators	0-100 for each	X		X	X	X	X	X		X		X	X		
Tennessee	Both	Single	0-4	X		X	X		X								
Utah	School	Single	0-600	X		X	X		X				X				

Ni, Bowers & Esswein (2016)

Virginia	Both	Single	0-100	X	X	X	X		X				X			
Washington	Both	Single	0-10	X		X	X		X	X			X				
West Virginia	Both	Single	0-100	X		X	X		X			X	X				
Wisconsin	Both	Single	0-100	X	X	X	X		X			X	X	X	X		
Wyoming	School	3 indicators	0-100	X		X	X		X			X	X				