August 2016

CCSSO
Council of Chief State School Officers

Key Issues in Aggregating Indicators for Accountability Determinations under ESSA

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public.

Key Issues in Aggregating Indicators for Accountability Determinations under ESSA

Author:

Ryan Reyna, Senior Associate at Education Strategy Group

One Massachusetts Avenue, NW, Suite 700 • Washington, DC 20001-1431
Phone (202) 336-7000 • Fax (202) 408-8072 • www.ccsso.org

CONTENTS

With the passage of the Every Student Succeeds Act (ESSA), states have wide authority to construct a school accountability model that can best advance college- and career-ready outcomes in their unique context. The law requires each state to meaningfully differentiate the performance of its schools on an annual basis, using a set of defined "academic" indicators (e.g., academic achievement, student growth, graduation rate, progress in achieving English language proficiency) and requiring at least one indicator of school quality or student success. In making the annual differentiations, the state must give "substantial weight" to each of the "academic" indicators, and in the aggregate, "much greater weight" than it provides to the school quality and student success indicator(s). The U.S. Department of Education's proposed regulations[1] further require a single summative rating overall for the school and that each indicator receives a rating, with both the summative and indicator ratings each having at least three performance levels.[2]

Over the next year, ESSA requires that all states review and revise their state accountability systems consistent with federal requirements and new state authority. The Council of Chief State School Officers (CCSSO) has worked with states to reaffirm key principles and produce a roadmap and resources to help guide this work. Those documents define how states can build high-quality, college- and career-ready systems of school accountability and supports, and meet ESSA requirements in that context. This paper is designed to go deeper on the options states have in aggregating indicators to make school accountability determinations.

This paper is intended as a resource for state education agency (SEA) staff as they develop their new ESSA accountability models. There is no single "right" way to aggregate multiple indicators within an accountability system to provide annual school performance classifications; rather there are a variety of key decisions that a state must make to develop a system that best fits its vision for accountability. As such, multiple options are presented along with additional key questions to prepare state education leaders for the critical task ahead.

The following presents an initial landscape of decisions and options for states with regard to the aggregation of indicators for accountability determinations. There are five big decision points that states must consider:

1. System purpose
2. Type of aggregation
3. Weights
4. Method for rating calculation
5. Communication of rating(s)

1 References are made throughout this resource to the U.S. Department of Education's proposed ESSA regulations. All references are contingent upon finalization of the regulations.

2 There are a number of additional accountability requirements in ESSA and the U.S. Department of Education's proposed regulations that are not specifically discussed in this resource. For example, the annual meaningful differentiation of schools must be based on all indicators for all students and each subgroup of students. This paper is largely silent on the specific inclusion of accountability indicators, or the measures that make up those indicators. States will need to consider the construction of their accountability system within the full context of the law and subsequent regulations.

SYSTEM PURPOSE

As described in CCSSO's *State Strategic Vision Guide*, state education leaders should approach the design of an accountability system by starting with key questions of purpose, including: what outcomes are the state hoping to achieve for students? What needs to change to achieve those outcomes? And, what information does the state need to inform that change? State leaders also will need to consider how the system helps focus the actions of students, parents, teachers and administrators. It is critical that the state engage with each of these groups to ensure that the purpose of the accountability system meets the needs of those most affected by the system.[3]

While accountability systems are intended to measure the impact of schools on student learning, they must also:

- Establish transparency in school performance for parents and policy makers;
- Make determinations for additional reward, support or consequence;
- Enable the continuous improvement of teaching and learning in the school.

Each purpose is important for states to consider when designing an accountability system. However, states will also need to determine which of these purposes is "first among equals." The decision regarding purpose will significantly influence the approach a state takes in addressing the other areas identified below.

It is paramount that all accountability systems reliably measure the impact that a school has on its students, rather than the impact student characteristics have on school performance. Constructing a model that fails to do this will likely provide inaccurate information to parents, misidentify schools in need of support, and downplay the exceptional work of educators and administrators in schools that are excelling with large populations of traditionally underperforming students.

Key Questions

When defining the state's vision for the accountability system, state leaders should consider the following questions:

- What is the most important outcome for the accountability system today? In five years?
- What parts of the current state system are driving the desired outcomes and what does the state want to change?
- How can the state accountability system drive desired behaviors and instructional practices?
- What process will the state use to engage stakeholders in defining the state's vision for accountability? How will that vision be communicated broadly?

3 For more information on stakeholder engagement, see CCSSO's guide to stakeholder outreach in the development and implementation of ESSA plans: http://www.ccsso.org/Documents/2016/ESSA/CCSSO%20Stakeholder%20Engagement%20Guide%20FINAL.pdf.

TYPE OF AGGREGATION

To make annual determinations of school performance, states will need to "aggregate" the multiple indicators in their accountability system.[4] There are four main ways that states can aggregate the group of indicators that make up the state's system for annual school determinations: index, goal-based, matrix, and dashboard. While these approaches may seem distinct, some states have blended these to fit their unique context. See below for examples of each method of aggregation, as well as examples of "mixed" state aggregation methods, using the required ESSA indicators for demonstration purposes. For additional examples of current and proposed accountability models that meet the requirements of ESSA, see CCSSO's *Accountability Models Matrix*.

In connection to the aggregation models presented below, states may also need to include additional business rules to address ESSA requirements such as inclusion of subgroup performance in school determinations, participation of at least 95 percent of students on the state assessment, and identification of comprehensive and targeted support and improvement schools. While not discussed in significant detail in this resource, it is important to recognize that there are additional decision points that may need to be made prior to a state finalizing its indicator aggregation methodology.

Index – Currently in use by many states, an index is the numerical aggregation of performance across the full spectrum of measures. The measures are often assigned a "weight" according to policy preferences. The index can be used to assign scores to groups of measures (i.e., performance on all state assessments) and/or the system as a whole. Index systems typically calculate a single summative school rating and communicate the rating through a number, label, letter, or symbol (see section below on communication for a more detailed description). Using an index aggregation, a state would select the bottom 5% of scores on the index to identify its comprehensive support and improvement schools.[5] States that employ an index system include (among others): Delaware,[6] Kentucky, Florida, Louisiana, and West Virginia.

The value of an index aggregation is that it provides clear information to the public about the policy value of each indicator in the accountability system. Additional benefits, limitations, and considerations are listed below.

4 This paper is silent on the specific measures that can fulfill the required ESSA indicators. Examples are provided for illustrative purposes only to demonstrate how aggregation occurs. The examples are not intended as a specific recommendation for the measures that a state should include in its system. Some of the state examples of aggregation are based on the state's ESEA waiver, and thus, do not include all of the required ESSA indicators.

5 There are additional requirements in ESSA and the proposed regulations for identification of comprehensive support and improvement schools, such as any high school with a graduation rate below 67 percent.

6 Note that under the accountability system approved in its ESEA waiver renewal application, Delaware did not provide a single summative rating on its school report cards; rather it provided individual ratings in four different categories. The full aggregation of indicators was only used for identification of Reward, Focus, and Priority schools, not for public communication. This system was designed to meet the ESEA waiver requirements, and as such, does not include certain required components of ESSA (i.e., progress in meeting English language proficiency).

- Benefits:
  - Simple for parents and the public to understand overall rating
  - Can maximize differentiation between schools and create clarity between the rating levels
- Limitations:
  - Can minimize transparency of performance on individual measures
  - May be difficult to weight the aggregation of individual measures appropriately
  - Performance expectations may be "behind-the-scenes"[7]
- Considerations:
  - Establishing policy "weights" requires political agreement regarding accountability system values (i.e., importance of status vs. growth)
  - Policy weights may not match numerical weights in calculation (see weights discussion below)
  - Performance expectations can be set for individual metrics and/or overall, which can lead to unintended consequences if not properly aligned
  - Performance expectations can be normative (i.e., top 5% of schools) or criterion-based (i.e., >90% performance or distance from goal)

Figure 1. Index Example – Delaware [8]

Area/Measures	Weight	Points
Academic Achievement	***25%***	***125***
Proficiency ELA	7.5%	37.5
Proficiency Math	7.5%	37.5
Proficiency Science	5%	25
Proficiency Social Studies	5%	25
Growth	***45%***	***225***
Growth in ELA	22.5%	112.5
Growth in Math	22.5%	112.5
On Track to Graduation	***20%***	***100***
On Track in 9th Grade	5%	25
4-year Cohort Graduation Rate	10%	50
5-year Cohort Graduation Rate	3%	15
6-year Cohort Graduation Rate	2%	10
College and Career Preparation	***10%***	***50***
College and Career Preparation	10%	50
Total	**100%**	**500**

7 Depending on the structure, subgroup performance may also be "behind-the-scenes."

8 Delaware's accountability system under the ESEA waiver: http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/404/DSSF_RefGuide_15-16_7-25-16.pdf

Goal-based – Similar to Adequate Yearly Progress (AYP), and still in use in a small number of states, the goal-based approach to aggregation establishes certain performance benchmark(s) and determines school performance based on whether the school met the benchmark(s) or how close it was to meeting the benchmark(s). In this approach, the goals would directly connect to the long-term and interim progress goals required under ESSA, and could be set in two different ways. First, the state can establish a single consistent performance expectation for all schools. The performance expectation can be determined in response to a policy priority (i.e., greater than 90% graduation rate for all high schools) or through the use of statistical methods to determine the feasible goal range. In general, statistically based goals are desirable to those set in an "ad hoc" manner because they are more actionable and defensible; however, statistically based goals can also lead to undesirable goals from a policy perspective (i.e., a school improving its graduation rate from 50 to 55% over three years). State leaders will have to balance the dual priorities of establishing goals that are both rigorous and attainable. Second, the state can establish goals for accountability determinations based on performance improvement (i.e., year-to-year change) or subgroup performance (i.e., gap closure).

School determinations result from whether a school met its goal(s) (1) on a set number of indicators (i.e., a school that meets six out of eight goals is identified as "Met Expectations"); or (2) a minimum "performance gate" on specific indicators (i.e., 60% Proficiency in ELA and Math). Schools that do not meet either of these thresholds are automatically identified for support and intervention. Goal-based systems typically calculate a single summative school rating and communicate the rating through a label (i.e., Met, Approaching, Not Met) or number (i.e., six out of eight goals met). Using a goal-based system, a state would likely identify its comprehensive support and improvement schools based on the greatest number of goals missed and/or distance from goals. State examples of a goal-based model include (among others): Tennessee (mix of goal and index), California (proposed[9]; mix of matrix and goal), and Vermont (proposed).

The value of a goal-based model is that it can be used to directly align cross-sector policy goals to create coherence across the education system. For instance, the state could align its goals for students' college and career readiness upon exiting high school with its higher education attainment goals. Additional benefits, limitations and considerations include are listed below.

- Benefits:
    - Clear expectations of performance for all schools
    - Has historical precedence that is relevant for stakeholders
    - Can take advantage of new predictive analytic models to set goals
- Limitations:
    - May narrow focus on improvement to students near benchmark
    - Requires multiple decisions about performance expectations because goals are set for multiple indicators
    - Depending on number of goals to meet, may be difficult for educators to focus their efforts
    - Requires resetting performance expectations as a result of test or cut score changes

9 Note that at time of writing, CA currently does not plan to provide a single summative rating for its schools. If the state moves forward as such, that would technically place the model under the "dashboard" heading; however, the dashboard would provide information about school performance in a mix of matrix (at the individual indicator level) and goal approaches.

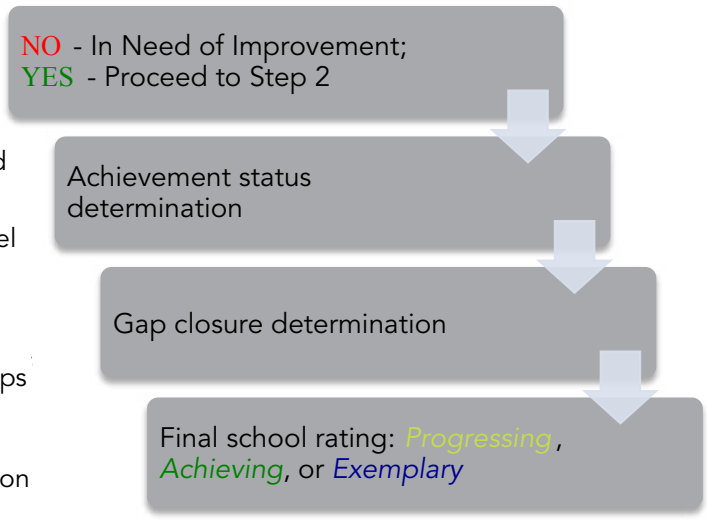
- Considerations:
  - Determinations can be binary (i.e., met/not met goal of 70%) or based on distance from the goal (i.e., points based on distance from 70%).
  - Predictive analytic models to set goals are under development and could take this established idea into the next generation of policy making. However, negative historical connotations of AYP may discourage next-generation models.
  - Proposed regulations require at least three performance levels, which increases the number of decisions about the "appropriate" level of performance, and may lead to conversations about whether and how the goals differ for different schools or student populations.

Figure 2. Goal-based Example – Theoretical

Measure	Performance	Goal	Status
Proficiency ELA	63%	60%	Meets
Proficiency Math	49%	60%	Does Not Meet
Proficiency Science	91%	70%	Exceeds
Progress in ELP	14%	50%	Does Not Meet
Growth ELA	Above Average	Average	Meets
Growth Math	Above Average	Average	Meets
4-year Grad	81%	83%	Approaching
6-year Grad	86%	85%	Meets
College and Career Readiness	41%	35%	Meets
Postsecondary Enrollment	22%	25%	Does Not Meet

Figure 3. Mix of Goal and Index Example – Tennessee[10]

- **Step 1**: Did school meet minimum Achievement and Growth expectations?
- **Step 2**: Assign points based on: Did school meet its targets? How did it perform relative to peers? What level of growth demonstrated?
- **Step 3**: Assign points based on whether school closed identified gaps
- **Step 4**: Average Achievement and Gap closure to get final determination



10 Tennessee's accountability system under the ESEA waiver: https://www.tn.gov/assets/entities/education/attachments/ESEA_flexibility_request_approval_summary_2015.pdf. This system was designed to meet the ESEA waiver requirements, and as such, does not include certain required components of ESSA (i.e., progress in meeting English language proficiency).

Matrix – Currently in use internationally, by states in a number of teacher accountability systems, and proposed in at least one state's new ESSA accountability plan, the matrix model reviews school performance in two (or more) domains, often status and growth (or improvement). Determinations are based on where the school falls within each domain. Practically, this is done through separate aggregation of indicators in each domain, and then plotting school performance on each axis (either in terms of relative performance statewide or with respect to state performance goals). Decisions must be made about which specific measures fall into the domains and whether one domain receives preference (or "weighting") when making determinations. Matrix systems typically calculate a single summative school rating, but communicate the school's performance in the two domains directly (i.e., High Status, Average Growth). Using a matrix model, a state could identify its comprehensive support and improvement schools based on the 5% of schools in the bottom left corner of the matrix. Because this is a relative measure, it is likely that the "bottom left" would need to be reset every three years for new identification of schools.

The value of a matrix model is that it provides parents, educators and the public more nuanced information about two different avenues of school performance. Additional benefits, limitations and considerations are listed below.

- Benefits:
    - Communicates the policy values of the system through the school rating (i.e., that status and growth are equally important)
    - Does not combine dissimilar measures, thus skirting many of the accountability weighting issues (discussed below)
- Limitations:
    - Few school accountability examples currently exist so the model can be more difficult to explain to stakeholders
    - May be harder to establish federal school classifications of comprehensive support and improvement and targeted support and improvement
    - May require additional aggregation across indicators to incorporate subgroup performance or produce a single summative rating
- Considerations:
    - Must decide how to incorporate specific indicators (or improvement) into each domain and whether to add additional domains, such as peer performance comparisons, to the model to further differentiate support and intervention
    - Establishing performance expectations requires political agreement regarding accountability system values (i.e., importance of status vs. growth)
    - Proposed regulations requiring each indicator to have a performance designation of at least 3 levels can complicate the communication of a matrix

Figure 4. Matrix Example – Theoretical Aggregation of Indicators

Status
- ELA
- Math
- Science
- 4-year Grad Rate
- On-track in 9th
- CCR

Progress
- ELA growth
- Math growth
- Progress in EL Proficien
- Extended year Grad Ra
- CCR Improvement

Status	Below Average	Average	Above Average
Above Average	Satisfactory	Promising	Exemplary
Average	Concern	Satisfactory	Promising
Below Average	Priority	Concern	Satisfactory

Progress

Figure 5. Mix of Matrix and Goal – California (proposed)[11]

Measure	Status	Change
ELA Assessment	Intermediate	Improved
Math Assessment	Low	Maintained
English Learner Proficiency	Very Low	Improved
4-year Grad Rate	Intermediate	Declined
Chronic Absenteeism	High	Improved
Suspension Rate & Local Climate Survey	High	Improved Significantly
College & Career Readiness	Intermediate	Maintained
Basics	Met	
Implementation of Academic Standards	Not Met for One Year	
Parent Engagement	Met	

Dashboard[12] – This approach does not attempt to aggregate indicators into a single school rating, rather it is designed to present actual school performance on each measure in a public format. In this way, a full picture of school performance is provided, without attempting to combine performance across multiple measures. ESSA does require annual determinations to identify low-performing schools for support and intervention, so a state still needs to develop and publicize its process (likely one of the three options described above) for making those determinations,

11 http://www.cde.ca.gov/be/ag/ag/yr16/documents/jul16item02.doc and http://www.cde.ca.gov/be/ag/ag/yr16/documents/july16item2addendum.doc

12 This refers specifically to the aggregation of accountability indicators, not to the idea of providing a compelling, publicly transparent collection of school performance measures. All states under ESSA are required to provide school report cards that display information about school performance across multiple indicators. States that use the other aggregation options and choose to display a single summative rating will still provide a "dashboard" of data.

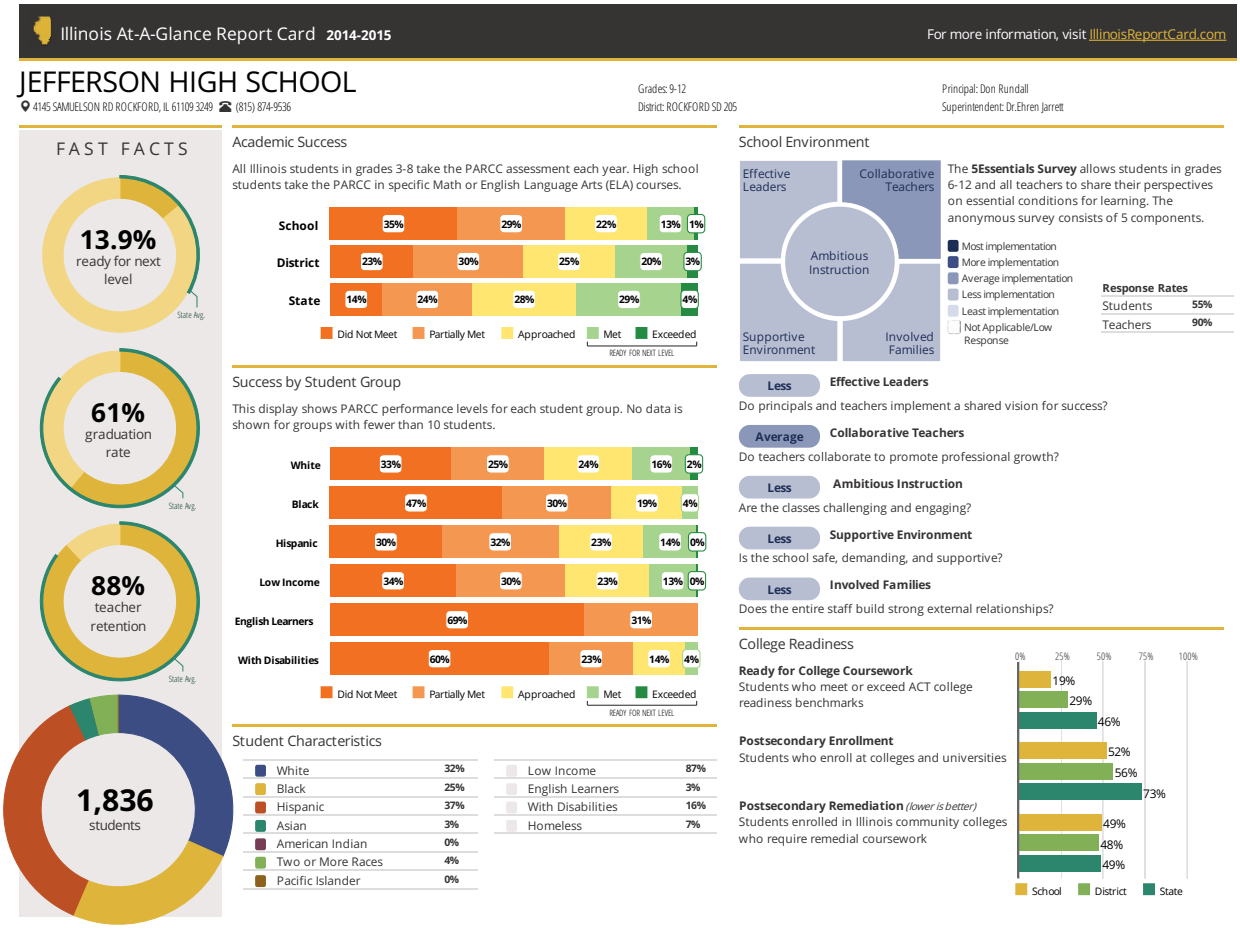
even if the information is not publicly displayed through a rating label.[13] This does not preclude a state from prioritizing the display of performance information through a dashboard; rather, it means that the state will need to identify a specific set of business rules to identify schools for the required federal classifications. Illinois is an example of a state that previously communicated its accountability information through the use of a dashboard.

The value of the dashboard model is that it aims to treat the accountability system as a tool for continuous improvement while avoiding the need to combine multiple measures into a single rating. Providing insight to performance on each measure can highlight gaps in even the highest performing schools. Additional benefits, limitations and considerations are listed below.

- Benefits:
    - Maximizes transparency of performance on individual measures
    - Allows stakeholders to determine their own values about the data
    - Does not combine dissimilar measures, and thus could skirt many of the accountability weighting issues (discussed below)
- Limitations:
    - May be difficult for the public to interpret overall performance across schools
    - Does not provide clear performance expectations for educators and administrators, thus may be challenging to understand why a specific school was identified for one of the federal improvement categories
    - Effective communication with dashboards takes considerable design work
- Considerations:
    - Another approach needs to be used for the identification of comprehensive and targeted support and intervention schools

13 The current regulations call for at least three performance categories. Under this rule, a state could identify schools for (1) comprehensive support and improvement, (2) targeted support and improvement, and (3) all other schools. ESSA does call for "meaningful" annual determinations and it would be up to interpretation whether that approach would meet the letter (or spirit) of the law. Additionally, the state would be required to publicize on its school report cards if a school has been identified in either the comprehensive or targeted support and improvement category and the method(s) used to identify those schools.

Figure 6. Dashboard Example – Illinois



Key Questions:

When deciding the approach the state will take in aggregating accountability indicators, state leaders should consider the following questions:

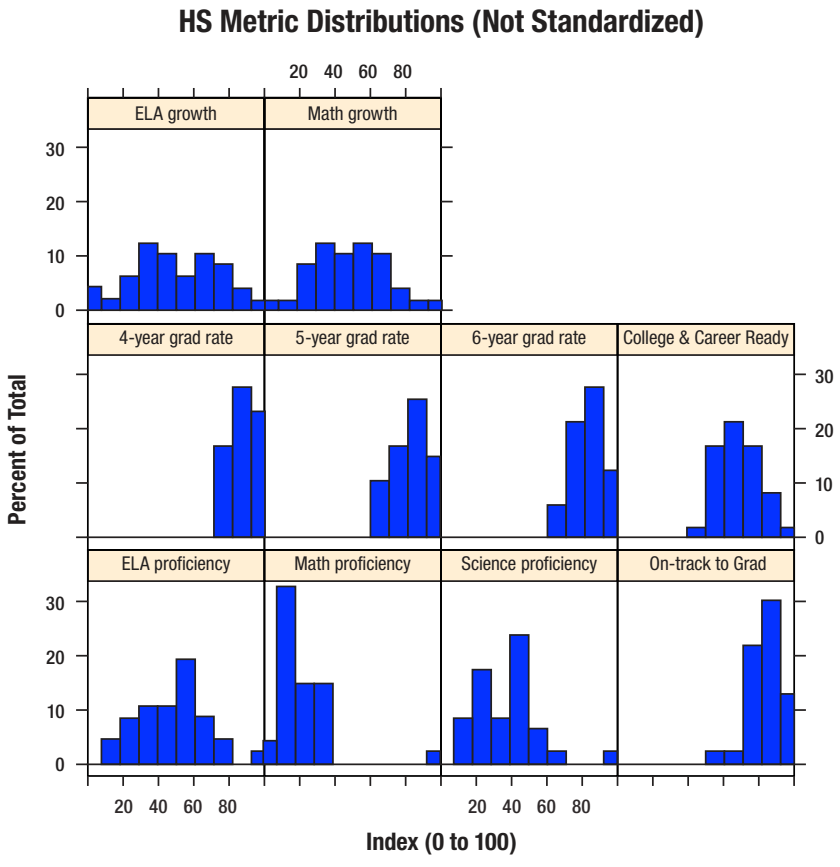
- How can the state balance ease of communicating the determination (or rating) with the transparency of performance on individual indicators?
- What process does the state plan to use to establish the performance expectation cut points? Does the state plan to establish performance expectations for each indicator or the whole system?
- What is an appropriate level of performance for each designation?
- What additional measures beyond those identified under ESSA will the state report to provide a complete picture of school performance?

WEIGHTS

Accountability systems based on multiple measurements of performance often require the "weighting" of individual measures to calculate a summative rating. Under ESSA, states will be required to weight accountability indicators to determine which schools to identify for comprehensive and targeted support and improvement. ESSA calls for the first four "academic" accountability indicators to account for "much greater weight" than the fifth indicator (e.g., school quality or student success). Determining appropriate weights for each indicator in the system is critical because it communicates to educators and the public what type of performance the state values most. It is likely that decisions on indicator weighting will require stakeholder engagement and potentially policy maker approval beyond the SEA.

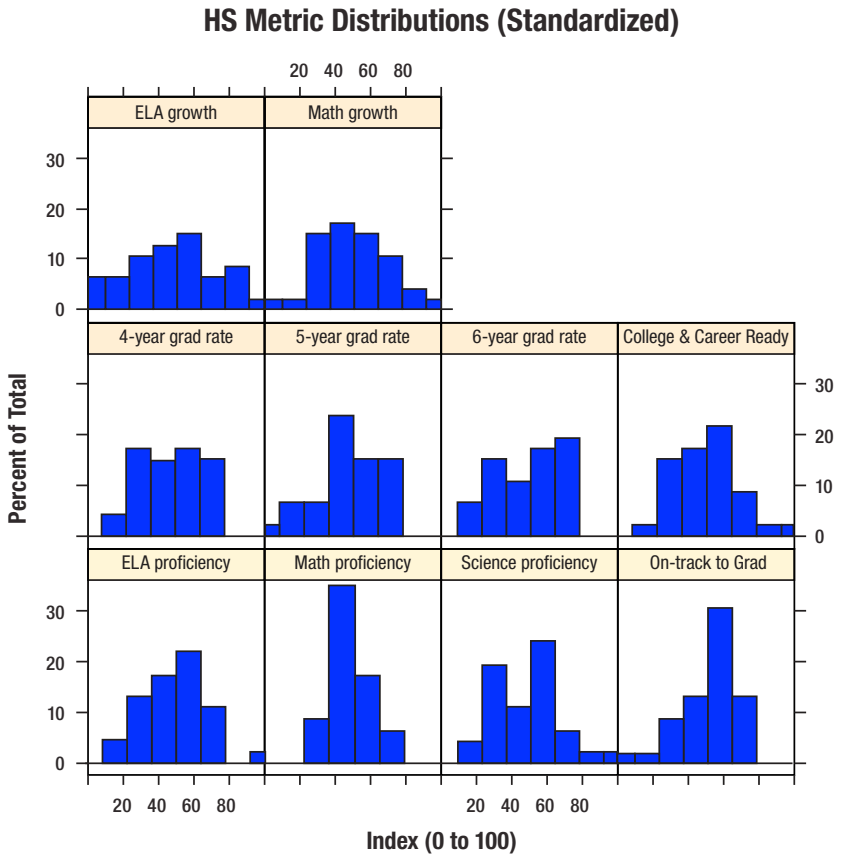
It is important to recognize that the term "weight" in this scenario is almost always referring to the policy weight assigned to a measure, which reflects system values (i.e., the state values growth more than proficiency) rather than the actual numerical weight of a measure. The ultimate value that a measure has in the overall calculation of performance is more affected by the spread of performance for that measure than the weight assigned in the system. An example of this phenomenon is provided below in Figure 7. The figure displays a set of potential high school accountability measures. Some of the measures, like English language arts and math growth, have a large spread or distribution of performance, while others, like 4-year graduation rate and On-track to Graduation, have relatively little spread. A state could assign significant "weight" in its accountability system to the 4-year graduation rate and On-track to Graduation measures (say 50%); however, because school performance is relatively clumped, very little differentiation between schools would result. Rather, the measures with the greatest spread would most significantly contribute to the overall differentiation of school ratings.

Figure 7. Example of non-standardized high school accountability measures.



There is a solution to avoid this scenario, but it too can lead to complications with the calculation and communication of ratings at the individual measure level. States can "standardize" (e.g., a statistical process to create spread in metrics, as demonstrated in Figure 8) all of the measures in their accountability system, so that the "policy" weights chosen match the "numerical" weights in the overall calculation. Standardizing makes performance relative, and thus may "force" schools to the "bottom" of performance on a specific measure, even if the level of performance is acceptable. For instance, a school with a 70% 4-year graduation rate may receive little to no credit (or points) in the overall rating for that performance level if the measure is standardized and that is the lowest graduation rate in the state. However, this issue can also be corrected with additional statistical techniques. Thus, it is important to seek technical assistance—from agency data analysts, state Technical Advisory Committees, or external research organizations—on statistical weights regardless of the state's method for aggregation or calculation of performance for individual measure.[14] Further, it is important to seriously consider the value of additional complications in the system. While standardization does allow for more accurate weighting among measures, it can lead to mistrust among stakeholders if they feel that the calculations are too complicated or done in a "black box." Communication is essential in this area.

Figure 8. Example of standardized high school accountability measures.



14 From a technical standpoint, states should also compare the overall school ratings with the individual indicator ratings to ensure that there is alignment.

Key questions:

When deciding how to "weight" multiple indicators of performance in the accountability system, state leaders should consider the following questions:

- How will the state determine the weighting of the multiple indicators in its accountability system? Who will be involved in that process?
- How will the state ensure that the "numerical weights" of individual indicators match the "policy weights" set forth by the state and its stakeholders? How will the state ensure that it meets the ESSA statutory language for the "academic" indicators to count for "much greater weight" than the school quality or student success indicator?
- What are the trade-offs of using standardization to create a spread of performance in indicators or correct a perception caused by such a technique? How does that align with the state's goals around transparent reporting?
- What capacity—internal or external—can the state rely on to conduct appropriate statistical tests on the accountability system and its individual components?

METHOD FOR CALCULATION

There are five main methods currently in use for calculating school performance on specific accountability indicators. Some states use a mix of the following approaches, so they should not be thought of as mutually exclusive.[15]

1. *Numerical* – In use in most states, this method of calculation takes the performance of a school on a specific measure and converts it to a scale for the mathematical aggregation of multiple measures. For instance, 4-year graduation rate is worth 10 points in the accountability system, so a school with 75% on that measure would receive 7.5 points. Alternatively, the state may set performance ranges (i.e., ≥70% and <80%) to attribute a certain amount of points. While this calculation method is mathematically transparent, it can also suffer from issues related to differences in the scales of specific measures. Metric weights in particular may become distorted without careful consideration. Nevada is an example of a state that employs this method.[16]

2. *Distance to goal* – In use in a few states, this calculates performance based on how close or far a school is from the performance expectation for a specific measure. For example, a state sets 4-year graduation rate goal of 90%, so a school with 75% graduation rate would receive 8.3 points (e.g., (75/90)*10=8.3). This method requires setting a goal for each measure and using numerical aggregation once performance against that goal is weighted. Connecticut is an example of a state that employs this method.[17]

15 For instance, Tennessee calculates performance through a number of methods, including: numerical, distance to goal, performance against self and performance against peer. For more information, see: https://www.tn.gov/assets/entities/education/attachments/ESEA_flexibility_request_approval_summary_2015.pdf

16 http://nspf.doe.nv.gov/Home/Points

17 http://www.sde.ct.gov/sde/lib/sde/pdf/evalresearch/next_generation_accountability_system_march_2016.pdf

3. *Performance against self* – This method of calculation takes into account a school's year-to-year improvement. For instance, a school with 75% graduation rate in the current year and 70% graduation rate in the prior year may receive full credit (i.e., 10 points) on the measure for demonstrating improvement above state expectations or points based on a sliding scale of improvement (i.e., improvement >4% = 10 points, >3% and <4%=8 points). The challenge with this method is that the comparison occurs between different cohorts of students, so there is greater potential for year-to-year fluctuations. The use of rolling, multi-year averages may be one way to address this issue. California's proposed model (shown above) provides an example of this method.

4. *Performance relative to peers* – This method of calculation seeks to evaluate performance in relation to similarly populated schools. For instance, take twenty schools with similar populations of low-income students. Nineteen of the schools have a 50% graduation rate and one school has a 60% graduation rate. Even though all of the schools may be underperforming state expectations, the school with the 60% graduation rate would receive additional credit for outperforming its peers (i.e., 10 points vs. 5 points for schools at the average of peer performance). The difficulty in implementing this approach is establishing meaningful peer groups. New York City previously employed this method.[18]

5. *Conjunctive/Bonus* – In use in a few states, often in addition to another approach, this method of calculation identifies specific measures that can move a school's rating up or down a designation scale based on performance. This method is often used to uphold or place significant emphasis on a specific accountability value, such as greater than 95 percent of students testing. In the conjunctive model, a school may be required to meet a specific performance threshold on a measure (or set of measures); otherwise their rating may be predetermined or reduced. For instance, a school may be required to meet the 95% assessment participation threshold, and if they miss that target, they automatically receive a low school designation. A similar approach can occur to move schools up the designation scale. Often termed a "bonus," a number of states provide additional credit to schools for meeting certain expectations on accountability measures. For instance, a school may earn up to an additional 5 points based on the percentage of students who enroll directly in postsecondary education or training after high school graduation. Georgia is an example of a state that uses this method through its "Exceeding the Bar" indicators.[19]

18 http://schools.nyc.gov/NR/rdonlyres/BD3585E6-B686-43F2-97F2-8F0EA3BF71FD/0/EducatorGuide_HS_11_25_2013.pdf

19 http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Accountability/Documents/Indicators%20and%20Targets/2014%20CCRPI%20Indicators%2004.01.14%20v2.pdf

Key Questions

When deciding how the state will calculate accountability indicators, state leaders should consider the following questions:

- What data, information systems, or technology supports will the state need in place to deliver on its approach to accountability aggregation? How should the SEA's chief information officer or data management team be involved in the collection, calculation, and aggregation of accountability indicators?
- Which, if any, indicators are included to demonstrate an accountability expectation for all schools (i.e., any measure of performance that automatically puts a school in an improvement category or moves a school up a designation category)?
- To what extent will the state use multiple years of data (or rolling averages) to calculate ratings or account for improvement in ratings?
- How will the state incorporate the 95% assessment participation requirement in ESSA? How will the state incorporate the requirement in ESSA to include subgroup performance in annual school determinations?

COMMUNICATION OF RATING(S)

States use a myriad of ways to display their accountability determinations.[20] Again, these tactics are not necessarily mutually exclusive; many states include specific ratings for both overall determinations and performance in specific categories of measures. This approach can increase public transparency.

1. *Numbers* – This can take the form of traditional (ex. 0-100, 1-5) or non-traditional (ex. 0-150, 1-4) scales. Use of traditional scales can be helpful for public transparency as stakeholders often have a frame of reference for approaching the numbers. However, that frame can create issues if performance does not fit the model as the public is used to experiencing it (for instance, a score of 70 out of 100 may be the highest performing school in a state, but would seem like a middle of the road school for someone who understands a 70 to reflect a "C" in school grades). The inverse is true of non-traditional scales. Additionally, the use of large scales may lead stakeholders to interpret significant differences in school performance where they do not actually exist. For instance, a 10-point difference on a 1000-point scale is unlikely to signal significant performance differences.
2. *Labels* – This can take the form of state-determined language (i.e., below expectations, meet expectations) or federally required categories (i.e., comprehensive support and improvement). Using words can limit school-to-school rankings. On the other hand, the language used can make it difficult for stakeholders to interpret how well a school is actually performing or how close a school may be to the next performance category.

20 The proposed regulations call for at least three levels of performance for the summative rating and for each indicator. For the summative rating, at a minimum, ESSA requires that states will identify and publicly communicate schools designated for (1) comprehensive support and improvement and (2) targeted support and improvement.

3. *Letter grades* – In use in a number of states, this places the designation of schools on an A-F scale, similar to grades. This style of determination faces many of the same challenges as raised in both the number and word approaches, as it is a specific combination of those two models.

4. *Symbols* – This approach takes a familiar symbol from other rating systems (for instance, Forbes rates hotels based on a quality rubric and communicates the rating on a 5-star scale) and uses it for school designations. This can provide the public with a frame of reference for reporting in a somewhat neutral way. However, it may be difficult for the public to understand how the calculation led to the specific rating, and the extent to which there is differentiation between the levels.

5. *Colors* – Often used in conjunction with words or symbols, colors are added to provide an additional level of context about performance to support the public's interpretation of the results. For instance, a high performing school may receive 5 green stars while a low performing school receives 1 red star. Questions arise about the adequate number of colors to include, and whether it is useful to go beyond the traditional red-yellow-green stoplight spectrum. For instance, some states use blue to signify the highest level of performance.

6. *Dashboard* – This approach pivots away from providing any specific ratings; rather, the goal is to provide school performance data in its actual form. While it is possible to both display individual performance data and an accompanying rating (and many states do), this approach differs in that there is a purposeful choice not to publicly label schools with state-defined rating(s). On one hand, this approach can encourage the use of the accountability system as a continuous improvement tool. On the other hand, it can make it difficult for the public to understand how well a school is performing. In this approach, a state could provide comparison data (to other schools, district or state) as a way of differentiating performance for the public without providing a specific rating for each data element.

Key questions:

When deciding how the state will communicate accountability determinations, state leaders should consider the following questions:

- How will the state engage with parents, educators, policymakers and the public to gather feedback on the communication of school rating(s)? Which approach to communication do stakeholders find to be most valuable?

- How many school rating categories will the state use? What are the benefits and limitations to providing additional nuance in the school determinations?

- What process will the state use to set the performance thresholds? Who will be involved in that process?

- How much differentiation exists between the state's chosen designation categories?

CONCLUSION

Designing a system to aggregate multiple indicators is one step in an ongoing process toward building a robust accountability system that provides clear information to the public, empowers educators with data and information necessary for improvement, and targets supports and interventions to the schools with the greatest needs. Throughout the process, state leaders should consider how the accountability system could evolve over time, and establish systems for periodic review and improvement. It is paramount that states prioritize communication with, and feedback from, diverse stakeholders through development, implementation and future revisions. Accountability can be a critical tool to support students on their path to college and career readiness. However, accountability loses value when students, parents, educators, and policymakers struggle to understand what is measured, why it is measured, and how it is communicated.

This paper was written by Ryan Reyna, Senior Associate at Education Strategy Group and former Director of Accountability and Data Management for the Delaware Department of Education, on behalf of the Council of Chief State School Officers. They author can be contacted at rreyna@edstrategy.org. The author wishes to thank Katie Carroll and Scott Norton at CCSSO for their feedback, and special thanks to Andrew Rice at Education Analytics for his support in the initial development of this paper.

Council of Chief State School Officers

One Massachusetts Avenue, NW, Suite 700
Washington, DC 20001-1431
voice: 202.336.7000 | fax: 202.408.8072