

Setting Goals in School Accountability Systems: A Focus on 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.

Setting Goals in School Accountability Systems: A Focus on ESSA
Brian Gong, National Center for the Improvement of Educational Assessment

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

Melody Schopp (South Dakota), President
Chris Minnich, Executive Director

Council of Chief State School Officers
One Massachusetts Avenue, NW, Suite 700
Washington, DC 20001-1431
Phone (202) 336-7000
Fax (202) 408-8072
www.ccsso.org

Copyright © 2017 by the Council of Chief State School Officers, Washington, DC
All rights reserved

Table of Contents

Introduction	4
Goals and ESSA.....	4
Requirements of ESSA.....	4
Relationships of Goals with Other ESSA Accountability Requirements.....	5
Goals Tightly Coupled with Accountability Ratings	7
Interim Goals as Criteria for Summative Determinations.....	7
Coherent Accountability Goals for Less Correlated Measures	8
Characteristics of Goals	10
Specifying Goals	10
Goals and Theory of Action	13
Systems with Multiple Goals.....	14
Aspirational and Empirical Goals	15
Goals and Accountability Consequences	17
Goals and Types of Performance.....	18
Goals on Assessments with Meaningful Achievement Levels	19
Patterns of Intermediate Goals.....	20
Goal Performance that can be Measured Reliably.....	22
Summary.....	24
Appendix – ESSA Provisions Related to Goals	25

Setting Goals in School Accountability Systems: A Focus on ESSA (*Every Student Succeeds Act*)

Introduction

Goals are an essential part of educational systems and serve several purposes. Goals embody values for rallying support from the public, policymakers, and others. Goals provide a focus for evaluating performance and an entryway into more comprehensive plans for making improvements.

For these reasons, setting appropriate goals is essential for having a powerful and credible educational system. However, there are many challenges to setting appropriate goals, including balancing ambitious and achievable within the state's accountability context, defining goals in a way that guide selection of measures that will support valid interpretations, and setting goals when there is a wide range of baseline performance. This paper discusses these challenges and provides practical guidance for state policymakers and designers of school accountability systems to set appropriate goals. Goals will be discussed in general, but the focus of this paper is on helping states develop long-term goals that meet their needs based on the requirements of the *Every Student Succeeds Act* (ESSA) regarding goals in the context of states' school accountability systems.

Goals and ESSA

Requirements of ESSA

The federal ESSA law (*Every Student Succeeds Act*, 2015)¹ specifies a number of requirements regarding goals for states' school accountability systems. The most important requirements of the statute are:

- The state must establish long-term goals and measurements of interim progress for indicators of academic achievement, graduation rate, and progress of English Learners on assessments of English proficiency.
- The goals must be expressed in terms of proficiency for indicators of academic achievement. The goal for graduation rate must be in terms of the four-year graduation rate; the state may also set a goal using an extended-year graduation rate, which must be more rigorous than the four-year rate.

¹ The *Every Student Succeeds Act* (ESSA, 2015), is the reauthorization of the federal 1965 *Elementary and Secondary Schools Act*, of which Title I is a major part. ESSA succeeds the *No Child Left Behind Act* (NCLB), which was the reauthorization of ESEA enacted in 2002. The ESSA statute can be found at: <https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf>. Appendix A contains the sections of the statute that refer directly to goals and their use.

- Goals must be set for the school and federally required subgroups (e.g., race/ethnicity as determined by the state, students with disabilities, English Learners, and economically disadvantaged students).
- The long-term goals must be ambitious.
- The goals must reduce the achievement gap between subgroups.
- Aspects of the school accountability system established by the state in compliance with ESSA must “be based on” the long-term goals.
- States may choose to identify schools for Targeted Support and Intervention (TSI) for Consistently Underperforming Subgroups (CUS) on the basis of subgroup(s) not meeting the long-term goals and/or associated interim goals; this is one possible but not required approach to identifying schools for TSI/CUS.

Relationships of Goals with Other ESSA Accountability Requirements

ESSA specifies three main aspects of accountability:

1. Reporting against long-term and interim progress goals
2. Reporting an overall summative determination and indicators for schools and subgroups within schools
3. Identifying schools for Comprehensive and Targeted Support and Intervention

Unlike previous federally-specified accountability systems (i.e., NCLB, ESEA-Flex), ESSA does not require a tight connection between the three aspects of goals, ratings, and identification for consequences. The state may choose to design its system so performance against the goals and overall determinations are closely related, but this is not required. The basis for identifying schools for Comprehensive and Targeted Support and Intervention (Low-Performing Subgroup(s) or Consistently Underperforming Subgroup(s)) by the state is specified in ESSA and does not require any measure of performance in relation to the goals.

If the state chooses to tightly link the long-term and interim progress goals with its system of summative determinations, there are a few general approaches for doing so, each with its own important considerations, which will be discussed below.

However, the state may choose to make its long-term and interim progress goals loosely coupled with its system for assigning summative determinations. In such a case the goals serve primarily as a reporting framework for some key aspects the state may wish to draw attention to, but which are not linked strongly to summative ratings or accountability consequences.

Figure 1: Main accountability components included in federal Every Student Succeeds Act (ESSA)

Three Main ESSA Accountability Components		
Goals	Summative Determination	Identification for Support & Intervention
<ul style="list-style-type: none"> Long-term, ambitious goals must be set for three indicators: Proficiency, English Learner (EL) Progress, and High School Graduation. Not required for Other Academic Indicator or School Quality/Student Success indicator. Must be measured in Percent Graduation using the four-year Cohort for Graduation Rate. Interim progress goals must increase over time up to the long-term goal. Must set long-term and interim progress goals for schools and each subgroup. Goals for subgroups of students who are behind on a measure must take into account the improvement necessary to make significant progress in closing statewide proficiency and graduation rate gaps. 	<ul style="list-style-type: none"> Based on all indicators, not just those required to have long-term goals. States may use extended-year adjusted cohort graduation rates in addition to four-year cohort rate. 	<ul style="list-style-type: none"> Required criteria for identifying Comprehensive and Targeted Support (low-performing subgroup) are related to “lowest performing 5% of Title I schools,” without explicit reference to either Goals or Summative Determinations. Identification of Targeted Support (Consistently Underperforming Subgroup) may be based on subgroup not meeting interim goals or not being on track to achieve long-term goal, but this approach to identification is optional; state may devise any method it wishes.

Goals Tightly Coupled with Accountability Ratings

The most common accountability design under NCLB and ESEA Waivers was to set intermediate goals (e.g., AMOs) and then evaluate schools' performance in relation to those intermediate goals. If a school did not meet the goal, then it was given a lower performance rating and perhaps identified for mandated support by the state. Although ESSA permits—but does not require—such tight coupling between goals, ratings, and identification for support, this section discusses two important aspects of this well-known design:

1. The role of interim progress goals as criteria for performance and summative determinations
2. Combining performance on goals for indicators that are not highly related (positively correlated)

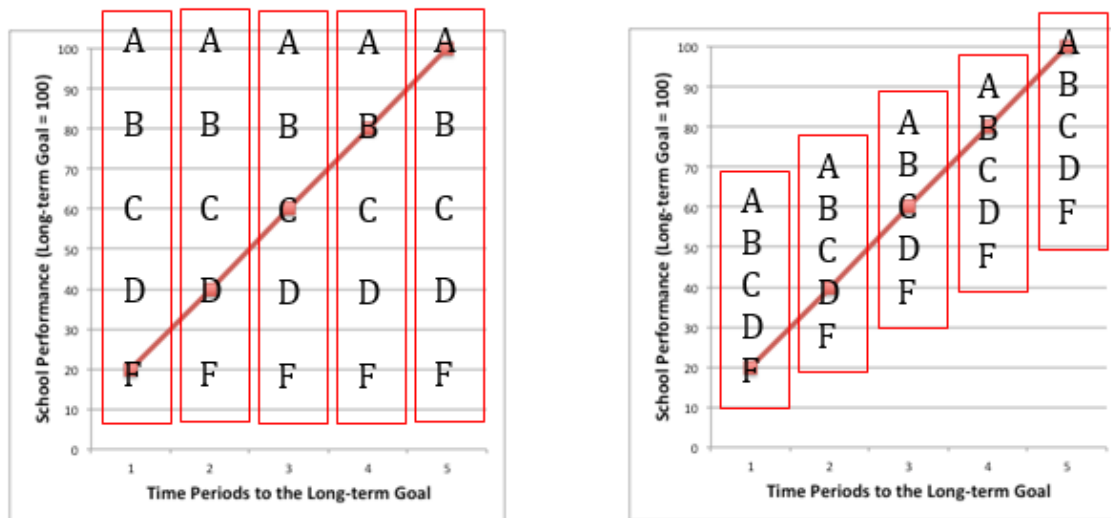
Interim Goals as Criteria for Summative Determinations

Consider a state that wants to use its long-term and interim progress goals to determine the shape of its accountability system ratings, such as its ESSA-required summative determination of an overall rating and/or indicator ratings. One way to do this is to use the interim progress goals to establish the criteria for “good enough” performance. That is, a “good school” is one that met the interim progress goals. Two fundamentally different ways this may be conceived are discussed below.

The first way is that school performance is reflected in the interim progress goals, regardless of the point in time. That is, the interim progress goals describe quality performance in an absolute way. Under this view, an excellent school would be one that met the long-term goal; a very good school would be one that met the highest interim progress goal; a good school would be one that met the second highest interim progress goal; and so on. (See figure below, left, where letter grades are used to indicate school summative designations.)

Another way to incorporate interim progress goals is to have each interim progress goal define “good/acceptable” performance for a specific time period. In this approach, the “good enough” criterion increases over time. (See figure on next page, right.)

Figure 2: Two different ways interim goals may be used to inform summative determinations over time



A state might choose the approach on the left if:

- It values having the descriptions and criteria of “good” performance be constant over time
- It values having entrance and exit criteria for identification of consequences be constant over time
- It either has schools across the full rating spectrum, or having many schools with high ratings during the initial interim progress time periods is not a priority

A state might choose the approach on the right if:

- It wants to emphasize that “good enough” definitions are increasing over time; This is especially useful if the “good enough” definition of interest is a minimal “just acceptable” performance standard
- It doesn’t have many schools that could meet a much higher criterion than the interim progress goal at the time
- It has defined progress as qualitatively better performance over time, similar to a learning progression (i.e., an “A” in fourth grade represents not only better performance but also more advanced content knowledge and skills than an “A” in third grade)

It may be more difficult to explain the relationships between summative determination ratings and interim goals if the interim goals and/or ratings do not increase in an even pattern over time.

Coherent Accountability Goals for Less Correlated Measures

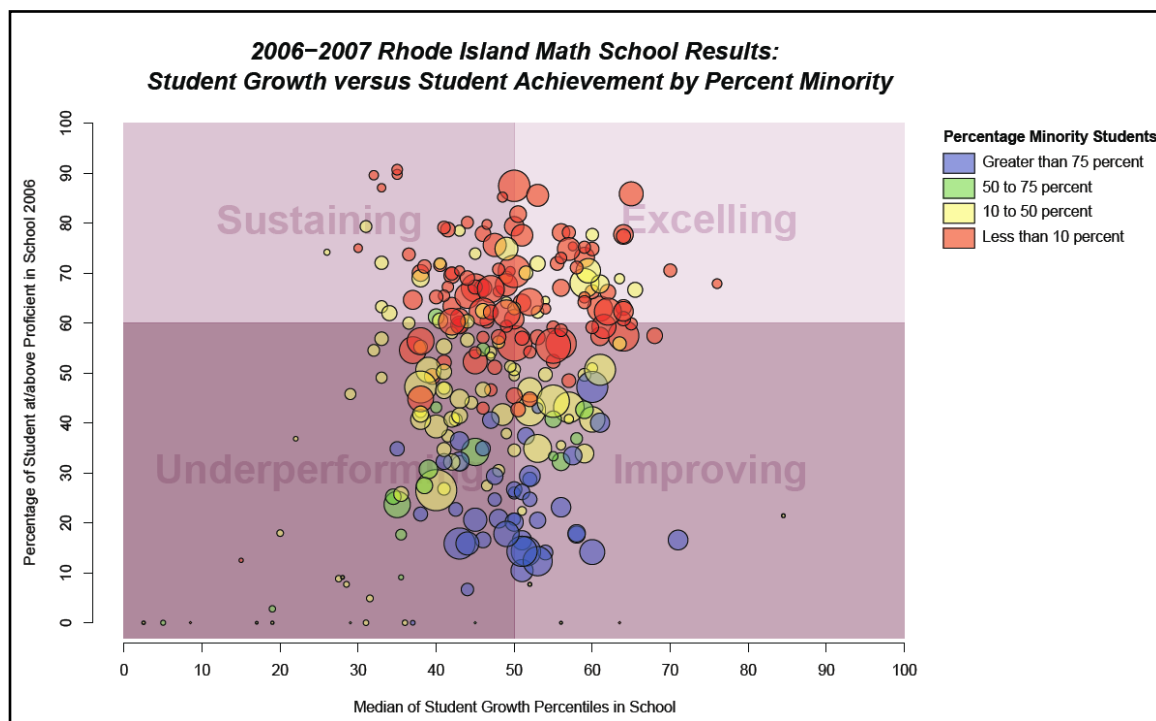
If interim progress goals are to be used as performance criteria in making summative determinations, then it is often assumed that because each set of interim progress goals increases over time that higher performance on one measure will

likely occur with higher performance on another. This view would underlie an accountability system that conjunctively requires “good” performance on every indicator in order to receive an overall rating of “good.”

However, school performance on multiple indicators may not be consistent because of variations in focus, resources, and other factors. In addition, there may be some measures that are known not to be highly related with each other. In such situations, the state must take care to set coherent accountability goals, particularly to combine performance measures against those goals to determine a rating.

An example is the relationship between Status (Proficiency at a point in time) and Growth (Increasing proficiency of the same students over time). As the figure below shows, it is often the case that Proficiency and Growth are not highly correlated; that is, schools that perform higher on Proficiency often do not perform higher on Growth, and vice versa. Empirical data such as in those Figure 3 show that it is rare for these schools to be high both in Proficiency and in Growth.

Figure 3: Display of interactions of Percent Proficient and Growth performances



Source: Rhode Island Department of Education and D. Betebenner. See <http://www.ride.ri.gov/InformationAccountability/RIEducationData/GrowthModelVisualization.aspx>

So a high overall summative rating would be assigned to a school that scored high in both Proficiency and Growth, and a very low overall summative rating would be assigned to a school that scored very low in both areas. But what about the “off-

diagonal” combinations of Proficiency and Growth, represented by the upper left and lower right quadrants?

Situations where there are non-obvious and non-linear relationships between performance on measures indicate that goals should be set by considering the measures in combination with each other, in order to have a coherent and rational accountability system based on the goals.

This section has briefly discussed some considerations for states where goals are the basis for summative determinations, though this is not a requirement of ESSA.

The remainder of the paper will focus on setting sound goals within an accountability system, but will not assume that the goals are or are not directly translated into summative determinations.

Characteristics of Goals

There are several important considerations for states when setting goals:

- Specificity
- Timing (i.e. is this goal measuring a point in time, or progress over time?)
- How the goals fit with state’s theory of action
- Aspirational goals vs. practical goals
- Implications for accountability systems
- Distinguishability between different performance measures
- Credibility against achievement levels
- Continued relevance over time
- Systemic coherence (i.e. how well a set of goals work together to achieve outcomes).
- Reliability of evaluations

Each of these points is discussed in more detail below.

Specifying Goals


Goals for accountability systems may be defined at many levels of generality and are not necessarily mutually exclusive. Three dimensions that may increase in specificity are a) defining the content area of the goal, and b) defining what would constitute evidence of performance, and c) defining “good enough” or baseline performance. ESSA requires states to set “long-term goals” and “measures of interim progress towards meeting those goals.” These two specific types of goals are discussed within the context of how goals fit within the state’s theory of action below.

Four helpful levels of goal specificity are:

- State Strategic Vision² – high-level description of desired outcomes. The identified outcome provides the general focus for accountability, and drives the relevant evidence and evaluation.
 - Examples: “All students are prepared for success”, “All students graduate college-/career-ready”; “Excellence with equity”; “Top 10 in 10 Years.”
- Accountability Areas of Focus – definition of the areas to be included in accountability systems to support the state strategic vision. Areas of Focus typically include: students/schools/districts; knowledge & skills; general outcomes.
 - Example: “All Students are Prepared for Success” is further defined to address 1) which students: all students and student subgroups enrolled in school, including students with cognitive disabilities and English learners, in any public school; 2) what knowledge and skills: reading or English language arts (depending on state content standards) and mathematics in grades 3-8 and once in high school; 3) general outcome: all students demonstrate proficiency on state tests, and all subgroups demonstrate proficiency at the same rate as the average of all students in the school.
- Long-term Goals – definition of the performance to be achieved, usually specifying the specific measures that will be accepted as evidence, the metrics, and what is “good enough.”
 - Example: “All Students are Prepared for Success” means all students and subgroups, in all content domains and grades as defined previously, will achieve at least a score of “Proficient” on the state’s assessment in each grade/content area, such that 100% of students as a whole and each student group scores at least “Proficient” by 2035.
- Interim Progress Goals – definition of the performance to be achieved as “good enough” at time points before the long-term goals. These also have been referred to in the literature and in other accountability settings as “intermediate performance goals,” “benchmark objectives,” and “annual measurable objectives.”
 - Example: “All Students are Prepared for Success” means all students score at least “Proficient” on state tests by 2035, and annually an increased percentage of students score at least “Proficient” between 2017 and 2035, e.g., if the school started with 38% of its students “Proficient” in 2017, over the next 18 years, each year at least an additional 4% of students would score “Proficient” so that by 2035, 100% would score “Proficient” ($100\% - 38\% = 72\%$; $72\% / 18 \text{ years} = 4\% \text{ per year}$).

² CCSSO’s [State Strategic Vision Guide](#) is a resource for Chiefs as they solidify, reform, and enact their vision for their state in the context of increased flexibility now provided in the federal law.

Figure 4: Different types of goals and associated degrees of specificity

Type of Goal	Description becomes more specific 	
State Strategic Vision	General outcome	
Accountability Areas of Focus	Define which students, knowledge/skills, more detail about outcome	
Long-term Goals		Define measures for evidence, including metric; “good enough” in terms of metric and time
Interim Progress Goals		Define “good enough” at time points leading up to the longer-term performance goals

These examples of goals and descriptions are intended to be useful guidance. There is no general rule regarding the number, names, or specifications for goals in educational accountability. These examples illustrate that is helpful to have more than one type of goal in terms of specificity, and to avoid confusion by differentiating the goals by name and type.

“State Strategic Vision” and “accountability areas of focus” goals are essential and provide the substantive focus for the accountability system. This is especially important since accountability systems must prioritize elements to include, omitting some things in favor of others. Goals at these more general levels already have shaped what will be incentivized and what will become consequential for schools and educators. State policy makers should pay attention to these goals, and consider the possible impact of what they decide to include in their accountability focus. Those considerations will necessarily be specific to each state and its context at a particular time.

It is important for states to have a long-term vision and accompanying long-term goals. For functioning accountability systems, it also may be important to create intermediate and longer-term goals that are distinct, but move towards that long-term vision.

In this paper, “long-term goal” refers to the ultimate state of an accountability system, according to each state’s specific timeline. Historically, the maximum amount of time observed for an accountability system goal has been 20 years³; in many states policy makers demand a shorter timeline.

³ See the long-term performance goals set in the early accountability systems in Louisiana and Kentucky in the 1990’s.

In the following sections, “goals” refers to both “long-term” and “interim goals”, unless otherwise specified.

Goals and Theory of Action

A “theory of action” refers to an analysis of what actions would lead to what results. Sometimes also called a “logic model,” a theory of action may be expressed as a set of “if... then” propositions. In relation to goals, a theory of action would express how the goal is to be achieved. A more complete theory of action would include not only the intended set of actions, but also plausible unintended consequences.

In an accountability system, it is essential that the state have a theory of action that plausibly indicates how schools will achieve the goals. Having a theory of action is not the same as having empirical data showing that some schools are meeting the goals. Empirical data show that achieving goals is possible; a theory of action lays out the steps or actions that could be taken to be successful. A theory of action should acknowledge the multiple ways to achieve a particular goal, and articulate what will likely lead to achieving the goal—it is a “theory” that should be monitored, evaluated, and improved over time.

To illustrate how a theory of action works, consider that of a hypothetical state. Suppose a state sets a goal that schools will increase the number of students scoring Proficient or higher by 10% within two years. The state’s theory of action says, “We believe that lower performing students often are taught by less expert teachers who lack understanding of the state’s content standards and how to help lower performing students learn them. To address this, the state will identify three content standards in each grade that students statewide score lower on. The state will gather and disseminate districts’ curriculum guides and model lessons addressing those standards. The state will incentivize districts to use professional development funds to teach teachers of lower performing students how to use those materials. We think a goal of an increase of 10% is reasonable because our analysis of districts that were performing at that level showed their teachers had this type of knowledge and skills.”

Note that a complete theory of action should consider possible unintended consequences to the accountability system, especially negative consequences.

Goals should be both ambitious and achievable and have a plausible theory of action that describes how the goals can be achieved.

Systems with Multiple Goals

State accountability systems will often have multiple longer-term goals and multiple sets of intermediate goals. Different goals—longer-term and/or intermediate—apply to different units, such as grade levels (elementary, middle, high schools), content areas (e.g., reading/ELA, mathematics, science), or student subgroups (e.g., White, African American, Asian). The timeline for these goals must be the same but the goals themselves may vary.

When there are multiple goals, the state should work to ensure the system of goals is coherent. States can require goals to be the same across units, or be consistent at least in principle. States may also create a coherent theory of action that explains differences.

Goals typically embody a set of dimensions. A set of goals may be the same on all, some, or none of the dimensions. Some important dimensions for specifying (sets of) goals include:

1. Starting point (baseline)
2. Ending point (longer-term goal)
3. Time
4. Shape of trajectory
5. Metric
6. ‘Good enough’ criterion
7. Is the baseline recalculated over time

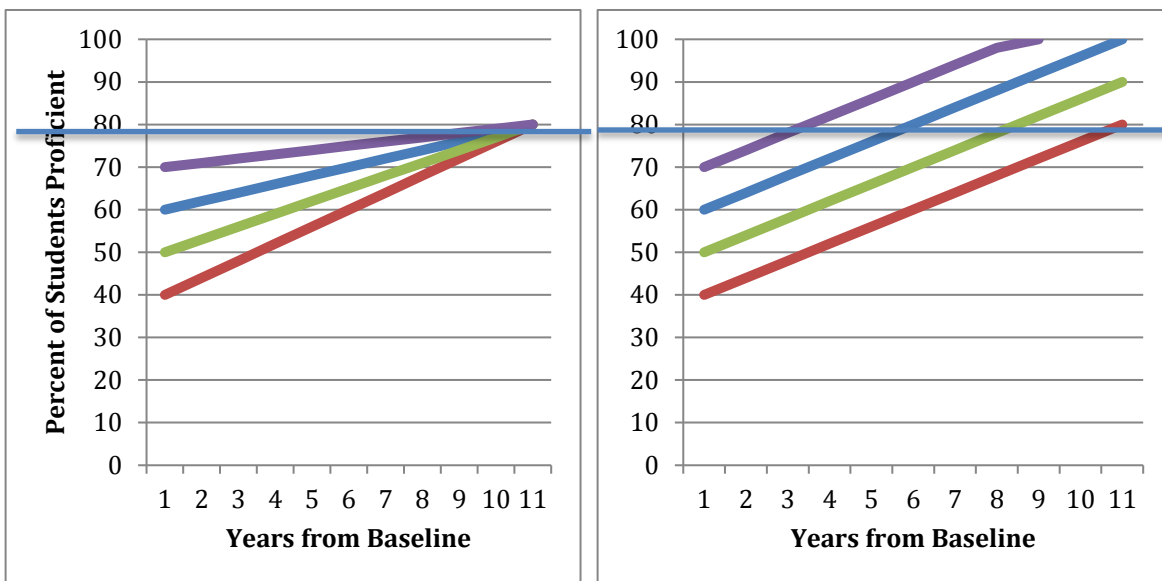
These seven dimensions must be specified to produce an ESSA long-term goal and interim goals (to inform “measures of interim progress”). When the accountability reflects different initial performances for schools and/or subgroups—i.e., different starting points—then there will be a set of goal trajectories, one for each starting point. Specifying each of the seven dimensions, and their interactions, reflects the values, intended outcomes, and incentives of the state sponsoring the accountability system. Each dimension is critically important; changing even one may lead to very different goals.

For example, the two figures below show patterns of intermediate goals that have the same values for six of the seven dimensions identified above: they have the same starting points (baselines) for four different groups; they have the same longer-term goal (80% proficient); the shape of the trajectories is the same (straight line); the metrics are the same (percent of students scoring proficient on the state test); the “good enough” criteria are the same (meet the interim goal); and baselines and trajectories are constant over time for both examples (not recalculated).

The main difference between the two figures is how time to reach the long-term goal is specified. The figure on the left requires the same amount of time for all

Figure 5: Two sample sets of goals that share six parameters out of seven

student groups to reach the long-term goal. The figure on the right requires different amounts of time for each group to meet the long-term goal (i.e., a student group that starts higher is required to meet the goal sooner).



Obviously, changing the parameter of one dimension results in changing other aspects. For example, slope of goal trajectories—the rate of change—is determined by baseline, end goal, and time. The figure on the left has trajectories with different slopes; the figure on the right has trajectories with the same slope.

The state must carefully decide the specifications that reflect its values, constraints, and data and then choose a set of goals that meets those values and specifications.

Aspirational and Empirical Goals

Goals should challenge at least some schools to significantly improve performance. However, when goals articulate a performance level that hasn't previously been observed at any school in the state, those goals are considered aspirational. Aspirational goals can range from aiming just slightly above observed performance levels, to performance levels that are much higher than any observed achievement. In general, policy makers in the past have tended to set aspirational goals in accountability systems that have been highly aspirational, e.g., "100% of all students," or "reduce percent not proficient by 50% in six years." Aspirational goals usually are fueled by a perception of what is good, just, or politically acceptable.

In contrast, empirical goals are based on what has been observed in the past and represent what is practical and possible. In order to establish empirical goals, states must gather and analyze data.

For example, if a state wanted to know about achievement in terms of students being “college-ready,” it could determine what percentage of students met a recognized college-ready benchmark on a widely accepted college entrance exam. The state might also gather empirical data for several years to inform setting expectations for improvement.

Figure 6 shows the data for a state published by ACT for five years.⁴ (This state was chosen for this example because it has administered the ACT statewide for many years, and tied results to accountability.) If the state wanted to use these data to inform expectations for improvement, they might look at the percentages of students who met ACT’s college-ready benchmarks over the five years. Here we see that the statewide achievement remained relatively steady in English and mathematics, decreased 6% in reading, and increased 3% in science. Before setting an aspirational goal that holds schools accountable for improving their students’ college-readiness more than a few percentage points in reading, English, or math, state leaders would take into account that those improvements are not reflected by empirical data for the state over the past five years.

Figure 6: Percentages of students who met ACT benchmarks, 2010-2014

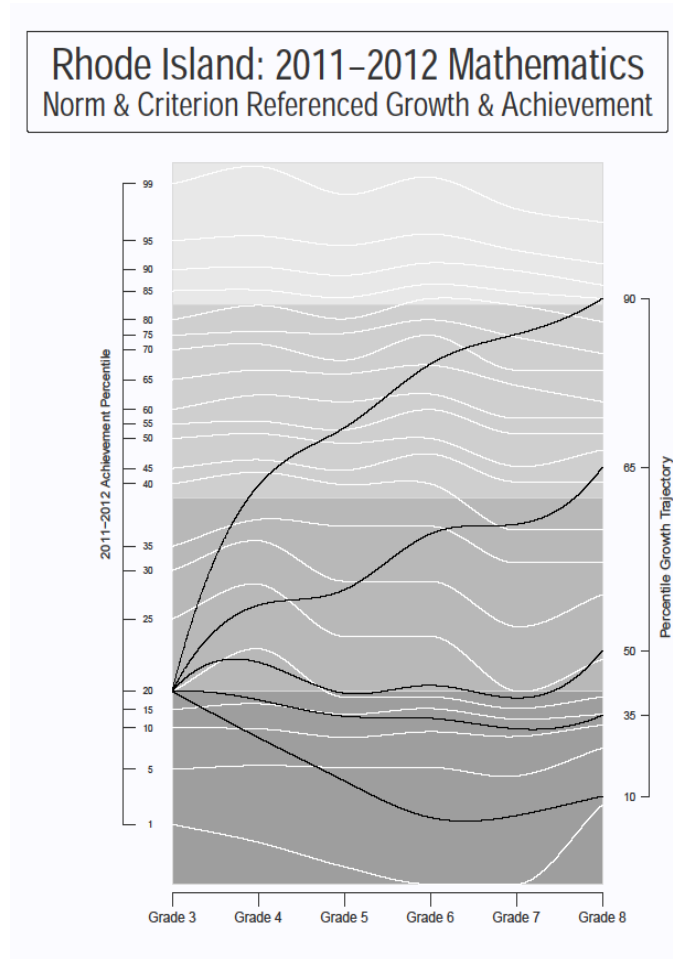
Year	Number of Students Tested		Percent Who Met Benchmarks									
	State	National	English		Mathematics		Reading		Science		Met All Four	
			State	National	State	National	State	National	State	National	State	National
2010	145,520	1,568,835	64	66	40	43	48	52	27	29	23	24
2011	144,469	1,623,112	65	66	42	45	48	52	28	30	23	25
2012	146,822	1,666,017	65	67	44	46	47	52	30	31	25	25
2013	160,066	1,799,243	63	64	42	44	41	44	35	36	25	26
2014	158,037	1,845,787	62	64	41	43	41	44	35	37	26	26

Another example of how empirical data can inform decision making around goals is shown in Figure 7 below, which summarizes growth data from a state.⁵ This figure shows not only the average student growth, but the amount of growth achieved by students at different percentile ranks (students at the 10th, 35th, 50th, 65th, and 90th percentiles). Using these data, a policy maker could say, “A student who starts just above Below Basic in grade three needs to grow at the 65th percentile in order to score Proficient by grade eight.” This provides an empirical basis for determining how to set a goal that all students become Proficient by grade eight. At the same time, these data also allow the policy maker to acknowledge, “There is a 35% probability that students who start at just above Below Basic will achieve this amount of growth” if their performance continues to improve at the same rate as it has in the past.

⁴ ACT State Profile Report, Illinois, Graduating Class 2014. Retrieved 1/31/17 from <https://www.act.org/content/dam/act/unsecured/documents/Natl-Scores-2014-Illinois.pdf>. Similar reports are available from ACT and College Board for every state and for the nation.

⁵ Betebenner, D. (2009). Norm- and Criterion-referenced Student Growth. *Educational Measurement: Issues and Practice*, 28(4), pp. 42-51.

Figure 7: Student Growth Percentile trajectories, grades 3-8



In most cases, goals should be both ambitious and plausible. If the goals are too aspirational, many educators and students will feel discouraged, and so may not perform at their highest level. If goals simply ask students and educators to maintain the status quo, the accountability system will not inspire improvement. Setting goals that strike this balance usually requires an engagement strategy with policy makers that incorporates a combination of values-based (e.g., what is good or just), and data-based (e.g., what has happened in the past) arguments.

Goals and Accountability Consequences

One of the defining characteristics of an accountability system is the established consequences based on performance in relation to goals. These consequences are assigned to a school depending on whether students did or did not meet the performance goals.

A consequence assigned for not meeting a goal is usually negative (undesirable); conversely, a consequence assigned for meeting a goal is usually positive

(desirable). Accountability system designers should consider balancing negative and positive consequences. States should decide whether their accountability goals articulate the minimum achievement expectations for schools and failure to meet the goal is met with a negative consequence, or whether the goals are “stretch goals,” where achieving the goal is commendable in and of itself.

When an accountability system has significant consequences, it is said to be “high-stakes.” When an accountability system stipulates that a student must pass a test in order to graduate, courts have identified that system as being “high-stakes” for the student because a) the consequence is valuable (i.e., a high school diploma can be seen as a property right), b) the decision and/or its implications are difficult to reverse (e.g., for a student who was mistakenly denied a high school diploma might not be able to enroll in college in the fall), and c) the consequence is largely tied to the accountability system. ESSA stipulates only that states must identify certain schools as qualifying for “Comprehensive” or “Targeted Support and Intervention” (TSI) and that certain TSI schools must establish a school improvement plan; the law allows each state to determine the actual consequences. Even for Comprehensive Support schools, the state has freedom to decide which support/intervention will be implemented.

Goals should be matched appropriately with consequences and stakes. The consequences of meeting/not meeting a goal should flow naturally from the goal itself. Implementing higher-stakes consequences necessitates implementing corresponding goals that are possible to meet. States should consider the coherence between their systems of goals, i.e. the state’s strategic goals should be translated in some way into the state’s accountability system goals and consequences.

Goals and Types of Performance

Accountability systems must define what schools will be held accountable for. A centrally important decision is what type of performance is related to school quality. One very useful typology distinguishes between four distinct types of performance, referred to as “status”, “improvement”, growth”, and “acceleration.”⁶

These four types of performance derive from two questions, which can each be subdivided into two additional questions:

- 1. How good is this school?**
 - a. What is the achievement level of the students? **(Status)**
 - b. Is this an effective school? Given the achievement level of students when they enter, how much do they develop or grow while they are in the school? **(Growth)**
- 2. Is the school getting better?**

⁶ Carlson, D. (2002). Focusing State Educational Accountability Systems: Four Methods of Judging School Quality and Progress. Available at www.nciea.org.

- a. Is the achievement level of this school improving or declining?
(Improvement)
- b. Is the quality or effectiveness of this school improving? How much more, or less, are the students learning than they did the year before? **(Acceleration)**

It has been shown that status, improvement, growth, and acceleration are all conceptually and empirically different.⁷ Notably, a school may have high status but low growth, or high status but low improvement, etc. In addition, each requires different types of data collection and analysis.

For these reasons, it is essential in setting goals, states are very clear about which type of performance is valued for the accountability system, and why. This idea has been discussed in other guidance for designing accountability systems.⁸ Because status is the simplest and most familiar to most people, this paper will use status in most of the examples. Keep in mind, however, that the principles discussed in the paper also apply to setting goals using improvement, growth, and acceleration.

A subsequent section discusses setting multiple goals coherently, including goals for status and improvement and growth.

Goals on Assessments with Meaningful Achievement Levels

In contemporary assessment and accountability systems, many goals will be based on the performance metric set by student achievement on a state assessment (e.g., Basic, Proficient, and Advanced). For the goals to be meaningful, those achievement levels must be meaningful. One way to show meaning is to establish a relationship between achievement levels and some other credible and valued outcome measure.

For example, when states established definitions of “Proficiency”, each definition was relatively similar in wording. However, the empirical data showed that states’ results did not correspond with each other, or with the National Assessment of Educational Progress. An influential graphic shows states’ 2003 Grade 8 Reading results for NAEP and the states’ own 2003 state assessments in terms of percentages of students who scored Proficient or higher.⁹ It was observed that in almost every case states reported more students Proficient on their state assessments than NAEP reported for the state. Also, some results were simply

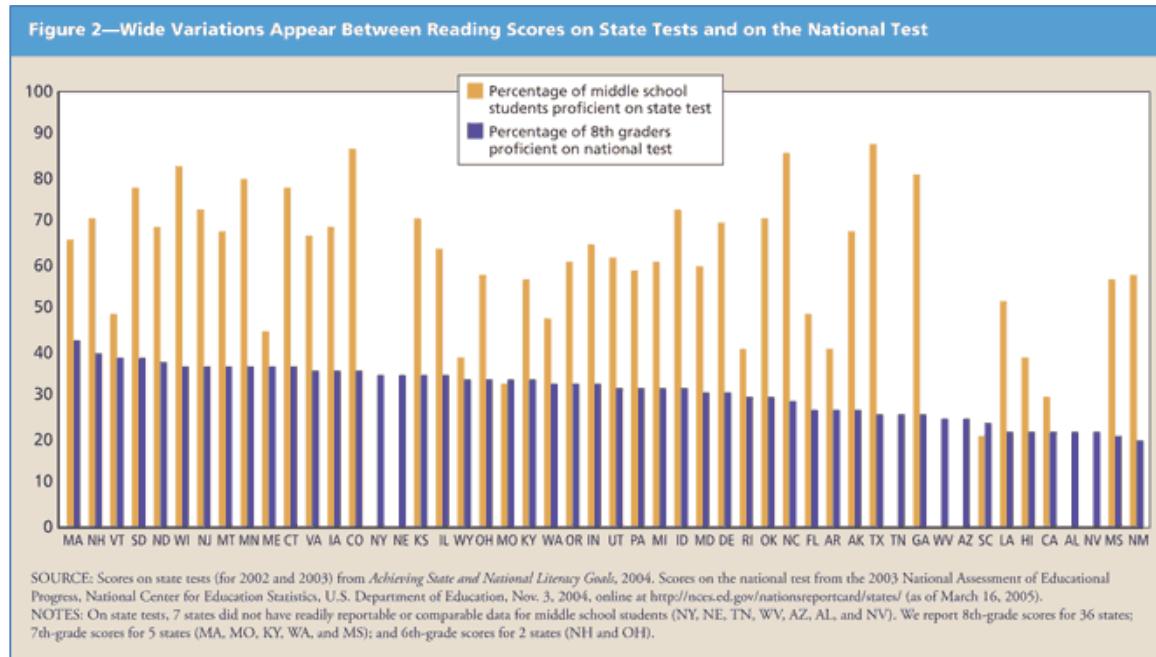
⁷ Ibid; Hill, R. (2001). Different Ways of Defining School Quality. Presentation at the 2001 Reidy Interactive Lecture Series, Oct. 4-5, 2001, Nashua, NH. Retrieved from http://www.nciea.org/publication_PDFs/RILS2001_Hill01.pdf.

⁸ For example, Gong, B. and ASR SCASS. (2002). Designing accountability systems: Towards a framework and process. Washington, DC: Council of Chief State School Officers. Available at: http://programs.ccsso.org/content/pdfs/designing_school_acct_syst.pdf.

⁹ McCombs & Carroll. (2005). Ultimate Test: Who is accountable for education if everybody fails? Santa Monica, CA; RAND.

implausible—lower-performing states reported very high proficiency rates on their state assessments.

Figure 8: Percent proficient on state assessments versus on NAEP, Reading,



Whether NAEP results are a good external criterion has been debated, but the point remains that an external reference is useful in validating the meaningfulness of assessment achievement levels. These differences between states' definitions of "Proficiency" are a leading reason why states and the federal government's ESSA law have shifted to defining desired achievement levels in terms of "college/career readiness". "College-ready," in particular has some empirical basis: every state has students who graduate from high school and then go on to college. These students' performance provides a basis for defining and interpreting performance on state assessments, including achievement levels.

If goals are set in terms of achievement levels, the achievement levels should have evidence that they are meaningful and credible. The same principle would apply to goals that use metrics other than achievement levels: the underlying measure must be meaningful and credible.

Patterns of Intermediate Goals

Intermediate goals set expectations between the point at which an accountability system starts measuring student progress, and the longer-term goals of the state. Intermediate goals should lead up to the longer-term goals in a plausible way over

time. That is, the pattern of successive intermediate goals should be supported by a theory of action and by empirical evidence, just as states support longer-term goals.

If an accountability system includes multiple measures, some of which are precursors or pre-requisites of others, then reasonable intermediate goals would be to improve performance on the “leading” indicators before improving performance on the “lagging” indicators.

If the intermediate goals are on the same measure as the longer-term goal, then the intermediate goals would be expressed more as a straight-forward performance measure, e.g., a higher percent of students are Proficient each year until the longer-term goal is met.

Common ways to establish intermediate goals are:

1. Use a formula or rules to generate the intermediate goals and the longer-term goals
2. Generate a starting point and timeline for the longer-term goal and then create of set of intermediate goals that connect the two.

Some examples of a formula are:

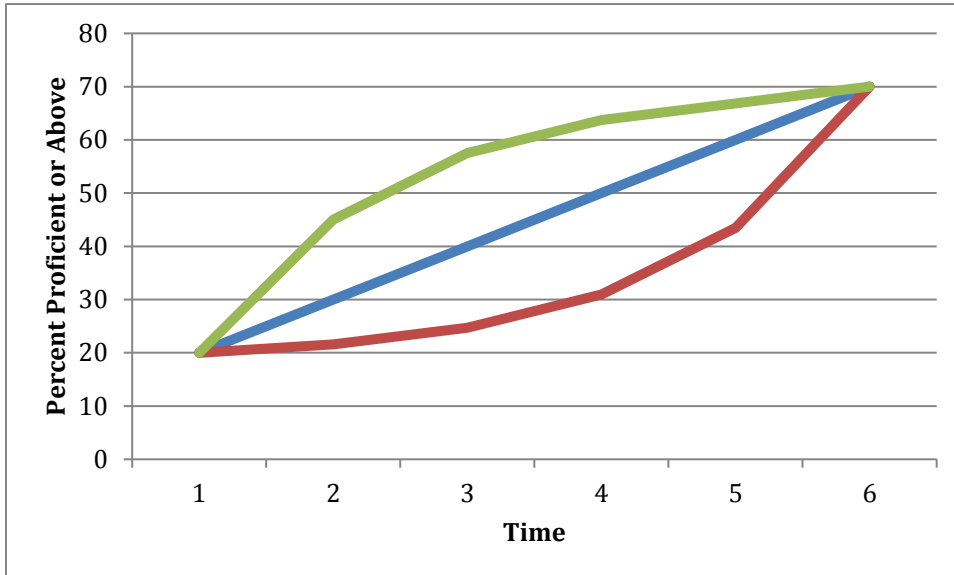
- Every school will increase its percentage of students scoring at or above the Proficient achievement level by 5% each year.
- Every school will decrease its percentage of students scoring below Proficient by 10% each year.
- Schools will be rewarded for an improvement in percentage of students scoring at or above the Proficient achievement level that is statistically greater than chance at the .10 level.

Formulas or rules such as these easily generate a series of intermediate goals. To create a longer-term goal, the state must decide the level of performance that is “high enough” and/or what a maximum timespan is.

In contrast, when the state first sets a longer-term goal (including a timeline for achieving that goal), the process involves creating a baseline and then filling in the intermediate goals.

In either process, the state must identify a pattern of intermediate goals that it considers reasonable and desirable to attach to accountability consequences. For example, two distinct patterns of intermediate goals are a) linear and b) curvilinear. A linear pattern indicates that constant improvement is expected and acceptable. A curvilinear pattern indicates that improvement is expected to accelerate or decelerate over time.

The figure below shows three patterns for intermediate goals that share the same baseline and longer-term goal, which embody deceleration, linear, and acceleration.



Of course, the patterns of intermediate goals do not have to be smooth lines or curves. For example, a “stair-step” pattern might be adopted if having stability of intermediate goals for a period of time followed by a larger change is preferable to a more gradual change every year.

Another common variation is to reset the baseline that determines the intermediate goals. This is usually done to recognize the actual performance of the school, which might not coincide exactly with its prior intermediate goal. For example, a school that started with a baseline of 20 percent of its students scoring proficient or above on the state assessment, and a longer-term goal of 70 percent proficient or above over five years might have intermediate goals of 30%, 40%, 50%, 60%. However, if the school in the first year achieved, 26% (instead of 30%), its intermediate goals might be adjusted to be 37%, 48%, 59%—to keep the intermediate goals an equal amount each time period, still ending with the longer-term goal of 70% proficient or above. The rationale might be that continuous improvement is expected to be linear from the school’s observed baseline to the longer-term goal, and the school should be held accountable for achieving the longer-term goal regardless of whether at some point it gets ahead or behind its intermediate goals.

There are several ways to set intermediate goals. The pattern and method for setting intermediate goals should be justified by a theory of action of how improvement will occur and what schools should be held accountable for.

Goal Performance that can be Measured Reliably

States pay close attention to ensure assessments are valid and adequately reliable for their intended purposes. In a similar way, accountability decisions and

classifications should be reliable, valid, and useful. Goals should be formulated with an intention to support accountability decisions that are stable and precise enough for their intended purposes.

Research has shown that the reliability/precision of accountability classifications and decisions are influenced by different variables than assessment reliability/precision. In other words, it is possible to have highly reliable assessment measurements and not have very reliable accountability decisions, and vice-versa.

Some of the variables that affect the reliability/precision of accountability classifications and decisions include:

- **Metric:** In general, more fine-grained scales or metrics are more reliable. A count or percentage of pass/fail decisions about individual students is usually less reliable than an average scale score of the same students.
- **Combining rules:** Decisions or classifications that result from a compensatory combination of rules are generally more reliable than those based on conjunctive or disjunctive rules. For example, a decision that allows higher performance on one measure to compensate for lower performance on another measure is more reliable than a decision that requires higher or lower performance on both measures and where performance on one measure can overrule the combined performance.
- **Distance to goal:** It is easier to reliably classify performance the farther it is from the goal or classification criterion. In other words, small goals (e.g., a 1% improvement in student performance) are more difficult to measure reliably, since there is less room for differentiation from the goal. For example, a school that scores 20 points below an achievement goal may clearly be categorized as off track; whereas a school that scores one point below the same goal might fall into more of a gray area (taking into consideration all variables that may make the categorization decision less certain).
- **Amount of measurement error:** The more measurement error, the more error in the accountability measure, assuming they are being measure in the same unit. Status measures, which are based on a singular measurement, are more reliable than similar growth or improvement measures, which are based on at least two measurements, and thus subject to more measurement error.
- **Sample size:** The sample size—most likely the number of students—is a major factor in determining the reliability/precision of accountability classifications and decisions. The smaller the sample, the larger the sampling error (sampling error only applies if the inference and data are based on having a sample rather than the complete population of students).

Designers and users of accountability systems, and those setting goals, should seek out technical advice to ensure the goals and decisions are adequately reliable/precise for the intended purposes.

Summary

This paper has discussed goals as they relate to ESSA. Key points include:

Goals and Specificity: Different types of goals should be identified, with different names and increasing specificity.

Longer-term and Interim Performance Goals: Longer-term and intermediate goals should specify the students, content domain, evidence, metric, “good enough” performance baseline metric, and time.

Goals and Theory of Action: Goals should be ambitious and achievable, and embedded within a plausible theory of action that describes how the goals will be achieved.

Aspirational and Empirical Goals: Goals should strike an appropriate balance between being ambitious and possible in order to maximize improvement and engagement. This can be done by considering both policy-driven desired performance (aspirational) and historical performance data (empirical).

Goals and Accountability Stakes: Goals and accountability consequences should be appropriately matched.

Goals and Types of Performance: Goals should distinguish between status, improvement, growth, and acceleration, and states should have a clear rationale for which components are included in accountability goals.

Goals on Assessments with Meaningful Achievement Levels: For goals that are set in terms of achievement levels, those achievement levels must be meaningful and credible.

Patterns of Interim Goals: Interim goals may show a pattern of accelerating, linear, or decelerating improvement between the baseline and longer-term goals. Setting intermediate goals requires a theory of action and consideration of empirical evidence.

Systems of Goals: Most goals are characterized by a limited number of aspects that can be varied to produce sets of goals. Systems goals must be coherent in every aspect that is valued. The state must carefully decide which specifications reflect those values, and then choose a set of goals accordingly.

Goal Performance that can be Measured Reliably: Designers and users of accountability systems, and those setting goals should seek out technical advice to ensure the goals are reliable and precise for the intended purposes.

Appendix – ESSA Provisions Related to Goals

SEC. 1111. STATE PLANS

(c) Statewide Accountability System

(4) Description of System – The statewide accountability system described in paragraph (1) shall be based on the challenging State academic standards for reading or language arts and mathematics described in subsection (b)(1) to improve student academic achievement and school success. In designing such system to meet the requirements of this part, the State shall carry out the following:

(A) Establishment of long-term goals – Establish ambitious State-designed long-term goals, which shall include measurements of interim progress toward meeting such goals—

(i) for all students and separately for each sub-group of students in the State—

(I) for, at a minimum, improved—

(aa) academic achievement, as measured by proficiency on the annual assessments required under subsection (b)(2)(B)(v)(I); and

(bb) high school graduation rates including—

(AA) the four-year adjusted cohort graduation rate; and

(BB) at the State’s discretion, the extended-year adjusted cohort graduation rate, except that the State shall set a more rigorous long-term goal for such graduation rate, as compared to the long-term goal set for the four-year adjusted cohort graduation rate;

(II) for which the term set by the State for such goals is the same multi-year length of time for all students and for each subgroup of students in the State; and

(III) that, for subgroups of students who are behind on the measures described in items (aa) and (bb) of subclause (I), take into account the improvement necessary on such measures to make significant progress in closing statewide proficiency and graduation rate gaps; and

(ii) for English learners, for increases in the percentage of such students making progress in achieving English language proficiency, as defined by the State and measured by the assessments described in subsection (b)(2)(G), within a State-determined timeline.

(B) Indicators – Except for the indicator described in clause (iv) [ELP], annually measure, for all students and separately for each subgroup of students, the following indicators:

- (i) For all public schools in the State, based on the long-term goals established under subparagraph (A), academic achievement—
- (iii) For public high schools in the State, and based on State-designed long term goals established under subparagraph (A) — [graduation rate]

--S.1177, pp. 34-35

(h) Reports –

(1) Annual state report card –

(C) Minimum requirements

- (vi) Information on the progress of all students and each subgroup of students, as defined in subsection (c)(2), toward meeting the State-designed long term goals under subsection (c)(4)(A), including the progress of all students and each such subgroup of student against the State measurements of interim progress established under such subsection.

-S.1177, pp. 45, 47