

July 2018

Establishing Performance Standards

for School Accountability Systems

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.

Establishing Performance Standards for School Accountability Systems

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

Carey Wright (Mississippi), President
Carissa Moffat Miller, Executive Director

Chris Domaleski, Center for Assessment
Juan D'Brot, Center for Assessment
Leslie Keng, Center for Assessment
Russ Keglovits, Nevada Department of Education
Ann-Michelle Neal, Utah State Board of Education

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



© 2018 by the Council of Chief State School Officers, *Establishing Performance Standards for School Accountability Systems*, except where otherwise noted, is licensed under a Creative Commons Attribution 4.0 International License <http://creativecommons.org/licenses/by/4.0> it is available at <http://www.ccsso.org/resource-library/establishing-performance-standards-school-accountability-systems>

CONTENTS

Establishing Performance Standards for School.....	2
Accountability Systems	2
Executive Summary.....	2
Introduction.....	4
Principles.....	5
Framework	5
Develop Policy Definitions	6
Develop School Performance Level Descriptors.....	7
Convene Standard Setting Panel	8
Preparation for Standard Setting	9
Standard Setting Event	9
Independently Identify Threshold Schools for Each Category.....	10
Establish Group Recommendation	11
Evaluation and Documentation.....	11
Conclusion.....	11
Appendix A – Case Study: Nevada.....	12
Appendix B – Case Study: Utah.....	14

ESTABLISHING PERFORMANCE STANDARDS FOR SCHOOL ACCOUNTABILITY SYSTEMS¹

EXECUTIVE SUMMARY

Most states have developed or revised school accountability systems in response to the Every Student Succeeds Act (ESSA). While these systems include multiple indicators, to many stakeholders the outcome of central interest is the overall rating or classification produced for each school. These ratings are often used to identify schools that merit reward or require support and to evaluate the efficacy of educational programs and policies.

States vary in their approach to producing school ratings. In some states, the accountability system culminates in a state-specific classification such as an A-F letter grade, awarding one to five stars, or other designations that communicate performance to the public. Other states do not provide an overall rating apart from the ESSA required categories of Targeted Support and Improvement (TSI), Additional Targeted Support and Improvement (ATSI), and Comprehensive Support and Improvement (CSI). Whether an overall or composite rating is provided, many states communicate performance using report cards or “dashboards” that often describe indicator level performance in terms thresholds (e.g., high/low; met expectations/ did not meet expectations; etc.).

Given the central importance of the accountability rating at the indicator or overall level, it is only reasonable to require compelling evidence that the rating has a high degree of validity for its intended interpretation and uses. A substantial part of that validity argument is in the design and implementation of a sound process for establishing performance standards that credibly reflects the state’s vision for the accountability system.

While there is a substantial research base in support of standard setting for assessments, very little attention has been given to establishing performance expectations in the context of school accountability systems. In many cases, accountability ratings are set normatively (e.g., the top 10% of schools receive an ‘A’), but using such procedures alone fails to ensure that the system reflects the policy values and prioritized outcomes that have been established by state leaders. In this context, this paper outlines recommended principles to guide the establishment of a standard setting-process for accountability systems and describes a framework for implementing standard setting.

The principles that guide an accountability standard-setting process are rooted in the *Standards for Educational and Psychological Testing* (2014) and include:

- Document rationale, procedures, and results;
- Ensure the process allows participants to apply their knowledge and experience; and,
- Include information associated with relevant consequences and criteria.

¹ This is the second of a three-paper set of resources presented at CCSSO’s *State Plan Implementation for ESSA* in April 2018. The other resources include <http://www.ccsso.org/resource-library/where-rubber-meets-road> and <http://www.ccsso.org/resource-library/accountability-identification-only-beginning>.

The framework is a systematic process to establish accountability performance standards that includes the following steps:

1. *Policy Definitions (PD)*. The state starts by deciding what performance categories should be established and a general Policy Definition for each category. For example, if the system will produce five performance levels (e.g., one to five stars, or letter grades A-F), then a brief description of each level consistent with the objectives of the system should be produced. These definitions should include any consequences associated with the level.
2. *School performance level descriptors (SPLDs)*. Next, the state should produce more specific School Performance Level Descriptors, which are detailed descriptions of what it looks like for a school to achieve each performance level in the state system. The SPLDs are based on the PDs but are written at a level of detail that can be used to inform the decision of panelists participating in the standard setting event. SPLDs should make clear whether performance across indicators, measures, and student groups is conjunctive, compensatory, or disjunctive to reflect statutory requirements and the intent of the system design. Additional guidelines for developing SPLDs are provided in the full paper.
3. *Standard Setting Panel*. The state may then convene a broad-based panel of leaders and stakeholders to evaluate information and make recommendations regarding performance expectations for the accountability system. The goal is to assemble a panel that is broadly representative of the state's interests and able to articulate a vision for education in the state.
4. *Standard Setting Preparation*. In preparation for the standard setting event, the state should generate multiple documents and resources that are needed to implement the standard-setting process. These include meeting agendas, facilitator scripts, presentations, as well as various handouts given to panelists.
5. *Standard Setting Event*. In convening the standard setting event, the state should identify a skilled and experienced facilitator who is very familiar with all aspects of the state system and context, has worked closely in developing the PDs and SPLDs, and can both operate and be perceived as independent. The event should include the following activities with the panelists:
 - a. Review and elaborate SPLDs
 - b. Independently identify threshold schools for each category
 - c. Establish group recommendations
 - d. Evaluate and document

Each of these activities is described in more detail in the full paper. Case studies from two states, Nevada and Utah, that implemented the described accountability standard-setting process, are also presented in the full paper to illustrate promising state practices.

In light of the importance of overall accountability school classifications, the process for establishing performance thresholds should be based on a well-defined, defensible procedure that reflects the state's vision for the system. Doing so bolsters the validity of the system for supporting the intended interpretations and uses. This document may be helpful to education leaders seeking to develop a process for establishing accountability performance classifications that fit their system characteristics and policy priorities.

INTRODUCTION

Most states have developed or revised school accountability systems in response to the Every Student Succeeds Act (ESSA). While these systems include multiple indicators, to many stakeholders the outcome of central interest is the overall rating or classification that is produced for each school. These ratings are often used to identify schools that merit reward or require support and to evaluate the efficacy of educational programs and policies.

States vary with respect to the approach to producing school ratings. In some states, the accountability system culminates in a state-specific classification such as an A-F letter grade, awarding one to five stars, or other designations for communicating performance to the public. Other states do not provide an overall rating apart from the ESSA required categories of Targeted Supports and Improvement (TSI) and Comprehensive Supports and Improvement (CSI). Whether or not an overall or composite rating is provided, many states communicate performance using report cards or 'dashboards' that often describe indicator level performance in terms thresholds (e.g. high/low; met expectations/ did not meet expectations etc.).

Given the central importance of the culminating accountability rating at the indicator or overall level, it is only reasonable to require compelling evidence that the rating has a high degree of validity for the intended interpretation and uses. A substantial part of that validity argument is the design and implementation of a sound process for establishing performance standards that credibly reflects the state's vision for the accountability system.

While there is a substantial research base in support of standard setting for assessments, very little attention has been given to establishing performance expectations in the context of school accountability systems. In many cases, accountability ratings are set normatively (e.g. the top 10% of schools receive an 'A'), but using such procedures alone fails to ensure that the system reflects the policy values and prioritized outcomes that have been established by state leaders.

In this context, the purpose of this paper is to outline recommended principles to guide establishment of a standard setting process for accountability systems and to describe a framework for implementing standard setting. Finally, two case studies are presented to illustrate promising state practices.

PRINCIPLES

As with setting standards for an assessment program, there are likely multiple approaches that can be defended given the context, features, and purposes of the system. Regardless, some basic principles should be considered when developing or evaluating any approach. The principles presented below are taken from the *Standards for Educational and Psychological Testing* (2014), adapted to apply to school accountability systems.

- **Document rationale, procedures and results.**

Whatever process is used, one should provide an explanation of the reasons for selecting that approach, including justification for how the method fits the context and supports the purposes and uses of the system. Further, the procedures and results should be described in detail, including development of performance expectations, selection and qualification of participants, and variability in judgments.

- **Ensure the process allows participants to apply their knowledge and experience.**

The individuals selected to provide judgments should be well qualified and selected to represent the range of perspectives and interests that should be considered when establishing school performance expectations (e.g. teachers, administrators, representatives from key interest groups). Moreover, the procedures and expectations of the judges should be clear and straightforward and all participants should be well trained.

- **Include information associated with relevant consequences and criteria.**

Judges should have access to appropriate data to inform their decisions (e.g. data for each accountability indicator for a range of schools and by subgroup). Additionally, to the extent that data associated with key criteria that are not otherwise available in the accountability system are available, these should also be included in the process to inform decision making (e.g. college-going rates might be included in a system that privileges college readiness). Finally, participants should have access to information about the meaning and consequences associated with each performance classification.

FRAMEWORK

In this section, a framework is presented for a systematic process to establish accountability performance standards. The framework is presented in the context of establishing overall school ratings. However, modifications could be made to the process to set standards at the indicator level only.

Develop Policy Definitions

The process starts by deciding what performance categories should be established and a general Policy Definition (PD) for each category. For example, if the system will produce five levels identified by an A-F letter grade, then a brief description of each category consistent with the objectives of the system should be produced². These definitions should include any consequences associated with the level. Table 1 shows an **example** of PDs for a hypothetical ‘letter grade’ system for high schools.

Table 1. Policy Definition Example

Category	Policy Definition
A	Recognizes a superior school that exceeds expectations for all students and student groups on every indicator category. These schools have high graduation rates and exhibit evidence that graduates are academically well prepared for success in college or careers. An A school will be rewarded for distinguished performance.
B	Recognizes a school that meets expectations for all indicator categories for all students and student groups. These schools have high graduation rates or rates that are significantly improving and exhibit evidence that graduates are academically well prepared for success in college or career. The school is exempt from submitting an improvement plan.
C	Identifies a school that has minimally met the state’s standard for performance. The all students group has met expectations for academic performance or academic growth and most student groups at least partially meet expectations for academic performance or growth. The graduation rate is not less than 75% for all students. The school must submit an improvement plan that identifies supports tailored to student groups and indicators that are below standard.
D	Identifies a school that has partially met the state’s standard for performance. The school has partially met expectations in academic performance or growth and the graduation rate is not less than 67%. The school must submit an improvement plan that identifies supports tailored to student groups and indicators that are below standard. A ‘D’ school in consecutive years is subject to state intervention.
F	Identifies a school that has not met the state’s standard for performance. An F school has not met the criteria for a D school or fails to achieve a graduation rate of 67%. The school must submit an improvement plan that identifies supports tailored to student groups and indicators that are below standard. The school is subject to state inventions including reconstitution if rated ‘F’ in more than 2 consecutive years.

PDs are critical to the process and, depending on decision-making protocols, states may elect to have them endorsed by high-level leaders such as the state chief, the state board, and/or leaders from key policy and advocacy groups. PDs should be linked to the state’s strategic plan and the theory of action for the accountability system. Importantly, this should occur before accountability standard setting is implemented and the PDs should guide the process at each stage.

² We use letter grades as an example in this paper, but this should not be interpreted as placing any special value or emphasis on this approach. There are many ways to provide an overall classification or rating (e.g. assigning ‘stars’ or performance category labels such as exceeding expectations or meeting expectations). We believe the methods described in this paper can generalize to these various approaches and simply use letter grades as one illustration.

Develop School Performance Level Descriptors

Next, the state should produce more specific School Performance Level Descriptors (SPLDs) for each classification. These SPLDs are based on the PDs but are written at a level of detail that can be used to inform the decision of panelists in standard setting.

SPLDs make clear whether performance across indicators, measures and student groups is intended to be:

- **Conjunctive:** minimum performance must be observed in all areas in order to meet performance expectations
- **Compensatory:** higher performance in some areas can offset lower performance in other areas
- **Disjunctive:** minimum performance can be observed in any area in order to meet performance expectations

It should be noted that if the state has determined in advance that certain rules or methods for combining performance across indicators or groups will be used, that may constrain decisions about the SPLDs. For example, if the state proposes to average scores across indicators (or use some other weighted composite) to determine an overall index score for schools, this method assumes that performance is compensatory. If SPLDs are written in a conjunctive manner, the outcomes will not produce classifications that are consistent with the SPLDs. Therefore, SPLDs should be developed in a manner that is consistent with the intended aggregation approach.

Another key decision for crafting SPLDs is how to define imprecise terms like “high rates” or “meeting expectations.” These terms may be used in PDs but need to be “unpacked” at the SPLD stage. Some approaches for defining these terms include:

- Associate performance expectations with the state’s long-term or interim goals (e.g., exceeding expectations occurs when the school or group meets the long-term goal)
- Set performance based on normative thresholds (e.g., a high rate refers to that which is exhibited by the top 20% of schools)
- Use external criteria to inform thresholds (e.g., required performance on a national test is associated with benchmarks endorsed by institutions of higher education in the state).
- Incorporate expectations that are central to federal or state policy in the SPLDs (e.g., a minimum graduation rate of 67% has significance in ESSA).

The point is to produce detailed descriptions of what it looks like to achieve each performance level in the state system. Taken to an extreme, this could be a mechanistic process. That is, if the SPLDs are too detailed, it leaves no room for judgment and producing the overall rating is simply a matter of applying a set of decision rules. However, in most cases, even detailed definitions leave room for judgment, particularly in a system designed to be compensatory.

For example, consider the sample SPLD cited below for a ‘C’ school in a hypothetical letter grade system. This SPLD contains quite a bit more specificity than the PD examples in Table 1. However, imprecise terms like “most” and “typically” allow for the role of judgment. Moreover, this recognizes that the range of possible C schools might include some profiles that are atypical. This is not a flaw in the SPLD. Rather, it is necessary for any system designed to be primarily compensatory.

A ‘C’ school has minimally met the state’s standard for performance by:

- o The all students group typically meets expectations for academic performance with 60% proficient in mathematics and ELA*

OR

The all students group typically meets academic growth targets with a mean growth score of 45 in mathematics and ELA.

- o Most student groups at least partially meet expectations for academic performance with **40%** proficient for academic performance in mathematics and ELA.*

OR

Most student groups at least partially meet expectations for academic growth with a mean growth score of 40 in mathematics and ELA

- o The graduation rate is not less than 75% for all students and not less than 67% for any subgroup.*
- o Approximately 50% of English learners should reach targets for progress in English language proficiency*

Note, also, that this sample SPLD aligns with the sample PD in Table 1. Obviously, SPLDs should be constructed for each performance level and each school type (e.g., elementary, middle, and high school).

Because developing sufficiently detailed and coherent SPLDs require a non-trivial amount of work, a strong approach is to develop them in advance of the standard setting event. This development process should include substantial input from key education leaders, stakeholders, and/or advocates. Doing so not only offers the benefit of expert advice from a variety of perspectives. It also provides more buy-in by including key stakeholders in the decision-making process.

These SPLDs should be regarded as pliable and will be reviewed and revised as appropriate by the standard setting panelists, but may not deviate from the policy definitions.

Convene Standard Setting Panel

Next, the state may impanel a broad-based panel of leaders and stakeholders to evaluate information and make recommendations regarding performance expectations for the accountability system. Members of the panel may include: senior policymakers at the department,

leaders from selected districts (e.g. one or two district superintendents), leaders from selected schools, representatives from critical agencies or offices (e.g., the governors education office, groups representing parents, business community, students with special needs, etc.). The goal is to assemble a team of leaders, experts, and stakeholders broadly representative of the state’s interests and able to articulate a vision for education in the state. A group of 15-20 panelists should be appropriate. Finally, it is beneficial to have representatives from the SPLD development process also serve on the standard setting panel.

Preparation for Standard Setting

Multiple documents and resources are required to implement the standard-setting process described in this document. In addition to agendas, scripts, presentations, etc., used by the meeting coordinators and facilitators, resources provided to panelists include:

- copy of the PDs and SPLDs
- reference materials about the state accountability system
- school profiles
- a mechanism to record ratings for each round
- an exit survey upon completion of the workshop

The school profiles may take on different characteristics depending on the specific methods selected. However, this should include a description of how the schools being evaluated performed on each indicator in the system and overall. The profiles should also include information about performance by student group. For example, a separate profile sheet could be provided for individual schools or a spreadsheet containing these data may be provided to panelists.

The overall or composite outcome is important to order the schools and to calculate the final rating. If the final composite or index is not available, a proxy value that orders schools in a manner that is consistent with the expected final composite could be used.

Standard Setting Event

There are a number of approaches for conducting a successful accountability standard setting event. The process will likely include the activities described in this section, but should be adapted as might fit a state’s individual context.

It should be noted that facilitation plays a central role in implementing the process effectively. The facilitator (or facilitators) should be very familiar with all aspects of the state system and the state context. Ideally, the facilitator(s) worked closely with the teams responsible for developing the PDs and SPLDs and can address the content, process, and rationale for the decisions in detail. Of course, the facilitator(s) should be skilled with working with groups to elicit productive participation and help bring together perspectives to find consensus where possible. Finally, the facilitator(s)

should both operate and be perceived as independent. That is, he or she should not advocate for a particular result; rather the emphasis should be placed on using the PDs and SPLDs in making judgments. If the facilitator is viewed as having “a thumb on the scale” this will impede the process and/or introduce bias.

Review and Elaborate SPLDs

Following an orientation and training process, the first task is to closely review and elaborate the SPLDs. For example, this might involve:

- Independent review of the SPLDs
 - o note of key characteristics and features that distinguish schools at each level
 - o note areas that are higher/lower priorities for each level
 - o note of areas that need to be clarified or elaborated
- Discussion in small groups and with the full group about key points from the review
- The full group makes any consensus decisions to refine or elaborate the SPLDs

Ultimately, the goal of this session is to reach group consensus and clarity on the SPLDs.

Independently Identify Threshold Schools for Each Category

The SPLDs provide an operational definition of the threshold or ‘just good enough’ school for each accountability classification. Accordingly, the panelists then use the SPLDs to identify schools that meet these criteria. There are multiple approaches to do this.

One approach involves evaluating a series of anonymous school profiles and assigning a performance level to each school based on which level is judged to be most correspondent to the school’s profile.³

Another approach involves reviewing a list of school profiles ordered by overall score and locating the school judged to be the ‘threshold’ school that separates each level. As with the previous approach, this involves making a decision about the degree to which each school profile satisfies the criteria outlined in the SPLD.

Whatever the approach, performance on each indicator and by subgroup will vary for schools, even those with the same composite score. Therefore, panelists must use their judgment regarding which school is ‘just good enough’ to exemplify each performance category given the preponderance of the evidence.

Panelists should independently record the school ratings and be prepared to address the rationale for their decisions.

³ Regression analyses can be used to compute the composite score that corresponds with the minimally sufficient rating to be classified in each level.

Establish Group Recommendation

Next, the facilitator shares the outcomes from independent rating including the impact (i.e., percent of schools that would earn each performance level) based on the mean or median rating. Panelists are invited to discuss these results and share their rationale for affirming or revising the median recommendation.

There are multiple ways to reach a decision at this point. Options include:

- Affirm the recommendation by consensus
- Affirm an adjusted recommendation by consensus
- Submit another round of recommendations and discuss results.

More than two rounds of ratings are likely impractical. If a second round of ratings is conducted, the mean or median rating for each threshold may be recorded as the groups' recommendation.

Evaluation and Documentation

Each panelist should complete an evaluation of the process, which includes questions about the process (e.g., Was your role clear? Did you have ample opportunity to share your perspective?) and the results (e.g., Do you believe the final recommendation for [level] are appropriate and defensible?) The results of the evaluation are an important piece of validity evidence for the standard-setting process.

Finally, full documentation of the process and results is prepared. The documentation should include the SPLDs, the range of independent recommendations, the final group recommendation and impact, a summary of the rationale provided by the panel for the recommendation, and a summary of the evaluation results.

CONCLUSION

In light of the importance of overall accountability classifications, the process for establishing performance thresholds should be based on a well-defined, defensible procedure that reflects the state's vision for the system. By doing so, the validity of the system for supporting the intended interpretations and uses is bolstered.

In this paper, a rationale and set of principles to guide the establishment of accountability performance recommendations were presented. Additionally, a high-level framework for a process was presented to illustrate a potential approach. This document may be helpful to education leaders seeking to develop a process for establishing accountability performance classifications that fit their system characteristics and policy priorities.

APPENDIX A – CASE STUDY: NEVADA

The approach the Nevada Department of Education (NDE) adopted for establishing recommended standards for the Nevada School Performance Index (NSPF) incorporated a systematic judgment process consistent with the framework described in this document.

The NDE started by convening advisory groups to develop and refine PDs and SPLDs for their state accountability system, which produces a one to five star rating for each school. Thereafter, a broad-based panel convened to evaluate performance profiles for schools judged to be minimally correspondent with the characteristics described in the SPLDs.

To operationalize this procedure each panelist was given a collection of about thirty school profiles. These profiles presented the performance on each NSPF indicator for all students and subgroups at anonymized schools. Panelists were also given a rating sheet with schools ordered from highest to lowest performing based on a placeholder NSPF index score.

Panelists were asked to rate each school from one to five stars based on their judgment regarding which star level SPLD was most correspondent with the school's performance profile. Because the list of schools panelists worked through was ordered, it was explained that panelists would likely apply the same star rating to a series of schools before switching to the next level. For example, if one started with the five star schools, one might decide the first few schools met the definition of a five star school before arriving at a profile that better matched with a four star school. Panelists were not restricted from 'deviating' from the ordering. That is, participants were permitted to place a more highly ordered school in lower star category if this was warranted in his or her judgment.

Importantly, panelists were instructed to closely attend to the SPLDs with each decision. It was understood that the school profiles evaluated might present a variable pattern (i.e. favorable performance in some areas and less favorable in others). Panelists were asked to consider the question, *Can you justify that the school has overwhelmingly met the criteria in the SPLD for the star rating assigned?*

In round one, each panelist worked independently through each of the elementary, middle, and high school ordered lists, submitting their rating sheets when complete.

After round one, the NDE displayed the results for elementary schools. The mean rating for each school in the ordered list was presented along with the standard deviation of the rating and the distribution of ratings. Summary statistics were also presented that revealed the number and percent of schools classified in each star level and the pooled standard deviation of all rated schools in that level. Schools were placed in the star level that corresponded to the rounded integer of the mean rating. For example, all schools with a rating of 1.5 to 2.49 were classified as two star school; schools with mean ratings ranging from 2.5 to 3.49 were assigned three stars and so forth.

After the data were presented from round one, panelists were invited to discuss the results. In particular, the facilitator asked the panelists to consider schools that exhibited the most variability

in ratings (i.e. the schools with the least agreement) and discuss the rationale for the selected ratings. It was explained that the purpose was not to attempt to persuade anyone to change their ratings as much as it was to help inform judgments by benefiting from the insights and perspectives of one's colleagues.

Following discussion, panelists were given time to independently make any adjustments to their round one ratings. Panelists were asked to note only the ratings that he or she wished to change. This second or adjusted rating was termed the round two judgment.

The process described was repeated for elementary schools, middle schools, and high schools, respectively until all round 2 judgments had been submitted.

In order to establish thresholds on the NSPF scale that are correspondent with the judgments of the panelists, the NDE applied the rating thresholds from standard-setting via linear regression to produce the estimated star level cuts.

More specifically, using the panelists' school ratings, the cuts for each level were produced using the equation

Where:

Y' = Predicted NSPF score associated with target cut

X = Star rating threshold

4.5 = 5 star threshold

3.5 = 4 star threshold

2.5 = 3 star threshold

1.5 = 2 star threshold

A = Slope from regression predicting NSPF from mean rating

B = Intercept from regression equation predicting NSPF from mean rating

This procedure is run separately for each of elementary, middle, and high schools to produce all the star cuts for the NSPF. These proposed cut scores were submitted as recommendations to the NDE.

APPENDIX B – CASE STUDY: UTAH

The Utah State Board of Education (USBE) also used an approach similar to the framework outlined in this document to establish performance standards for its new school accountability system. Utah's recently adopted Senate Bill 220 (SB220) provided detailed specifications of the requirements for its school accountability system. The statute defines the performance indicators on which a school's overall rating is based for elementary/middle schools (Section 53A-1-1106), and for high schools (Section 53A-1-1107), and specifically prescribed how points for the various indicators should be awarded in the calculation of a school's overall rating (Section 53A-1-1108 through 53A-1-1110). It also specified the assignment of letter grades (i.e., A to F) as the state's method for determining a school's overall rating and labels for each (e.g. 'A' is an Exemplary school; 'C' is a Typical school) (Section 53A-1-1105).

The standard-setting process builds on the statutory requirements by including three main steps:

1. Establishment of PDs
2. Specification of SPLDs
3. Recommendation of Performance Level Threshold Scores

Over a span of three days, groups of stakeholders from across the state of Utah were convened to participate in each of these steps.

For the establishment of PDs, a committee of key legislators and state board members reviewed draft PDs for elementary/middle schools and for high schools. The draft PDs were written in advanced by USBE as starting points. The committee made several important revisions to the draft PDs and approved them for use by the SPLD committee on the following day.

The SPLD committee consisted of members of Utah's Assessment and Accountability Policy Advisory Committee (AAPAC), which included superintendents and accountability experts from districts and schools across the state. USBE also drafted preliminary SPLDs for this committee to use as starting points. Separate SPLDs were written for elementary/middle schools and for high schools. During the meeting, the approved PDs from the previous day were shared. The committee then examined the SPLDs to ensure that they reflected Utah's vision for the school accountability system in the PDs, described the expectations in the different performance levels, and were clearly articulated for the performance level setting committee. Each committee member reviewed the preliminary SPLDs independently. Committee members were then divided into table groups of 3-5 people to share their thoughts on the SPLDs and recommend revisions to each SPLD. Finally, the committee reconvened as a group to revise the SPLDs and approve them for use by the performance level setting meeting on the following day.

In the final step of the standard-setting process, a committee of AAPAC members, policymakers, parents, educators, association representatives, and technical experts met to recommend performance level threshold scores for the overall school ratings in the accountability system.

The committee followed an iterative judgmental process that included multiple rounds of review, ratings, and feedback to arrive at threshold score recommendations.

Two key tools were used by the committee during the meeting: the ordered school profile lists (OSP) and detailed school profiles. An *OSP* is a list of all schools in Utah, ordered by the percentage of total points earned on the legislatively-mandated accountability indicators. This list also included information about the number and percentage of points earned for each indicator by each school. A *detailed school profile* is a report that includes additional empirical data about a given school, including test participation rates, detailed breakdowns of each accountability indicator by subject area, and historical demographic and performance data, for all students and by student groups.

During the meeting, committee members first reviewed and discussed the PDs and SPLDs approved by the committees on the previous days. They then participate in two rounds of judgments and group discussions to arrive at recommended threshold scores. In Round 1, the goal was to identify probable ranges for each threshold score (i.e., “range-finding”) using the OSP. This step was done independently by each committee member based on his or her interpretation of the SPLDs. In Round 2, the goal was to locate the threshold score for each performance level (i.e., “pinpointing”) within the respective probable ranges identified in Round 1. This step was done in table groups (of 3-5 people) by reviewing the detailed school profiles for every school within the probable ranges. The final committee recommendations were determined by examining the median threshold scores across table groups in Round 2 and evaluating the associated impact. By group consensus, the committee could make adjustments to the Round 2 threshold scores. The committee also made additional revisions to the SPLDs so that they reflected the recommended threshold scores.

The outcomes of the standard-setting process were summarized and shared with the Board of Education and the Utah State Legislature for approval and implementation.



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