OPERATIONAL BEST PRACTICES FOR ACCOUNTABILITY

For the Accountability Systems and Reporting State Collaborative on Assessment and Student Standards

February 2022

Juan D'Brot, Ph.D., Senior Associate
Center for Assessment

Center for Assessment

National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

Prepared with support from the
Council of Chief State School Officers

CCSSO
Council of Chief State School Officers

OPERATIONAL BEST PRACTICES FOR ACCOUNTABILITY

TABLE OF CONTENTS

OPERATIONAL BEST PRACTICES FOR ACCOUNTABILITY

INTRODUCTION

Accountability mechanisms and processes have evolved with changes in statute. State Education Agency (SEA) accountability systems may exhibit characteristics of compliance- and/or improvement-focused systems in light of requirements under the *Elementary and Secondary Education Act* (ESEA) and its reauthorizations under the *Improving Americas School Act, No Child Left Behind,* and State Flexibility from ESEA (aka Waivers). ESEA, as amended by the *Every Student Succeeds Act* (ESSA) provides states another opportunity to consider how federal, state, and local accountability mechanisms contribute to improving student outcomes and facilitating equitable access to high-quality educational opportunities.

The purpose of this document is to provide guidance and criteria for the design, development, implementation, evaluation, and potential revision of accountability systems with which SEAs may interact or operate. These criteria describe practices that are applicable across context-dependent systems while providing concrete benefits to users. These best practices are intended to:

- Establish a common set of operational criteria that correspond with guidance reflected in other areas of measurement in education (e.g., large-scale assessment, evaluation, psychological testing, etc.);
- Reflect 'lessons learned' from state experiences with NCLB and local accountability models;
- Reflect expectations that improve accountability system design, development, and implementation in non-partisan and unbiased language consistent with other professional standards;
- Represents the breadth of the considerations and decisions associated with accountability system design, development, and implementation;
- Provide multiple examples or methods of how to achieve a goal, objective, or criterion; and
- Articulate a range of best practices and criteria that can be approached by SEAs with varying levels of capacity and expertise.

Structure of this Document

This document is organized in a series of chapters that serve to categorize the operational best practices around 9 key chapters, aligned to five distinct stages. Chapters are organized in sequence under each phase. The stages and chapters include:

A. Design Stage
   1. Establishing an Accountability System's Theory of Action
   2. Identifying Stakeholders and their Involvement
   3. Accountability Roles, Responsibilities, and Program Management

B. Development Stage
   4. Selecting and Integrating Measures for Accountability Systems
   5. Establishing Performance Standards for Accountability Systems

C. Implementation Stage
    6. Articulating Operations and Quality Control in Accountability Systems
    7. Reporting and Communicating Accountability Results

D. Evaluation Stage
    8. Monitoring and Evaluating Accountability Implementation

E. Revision Stage
    9. Engaging in Accountability System Change Management

Within each chapter, readers are provided introductory text that clearly describes the importance and role of the information provided and how it is intended to be used. The chapter then presents detailed recommendations and considerations for accountability designers in the form of best practice statements (e.g., 1.1, 1.2., etc.) and supporting actions (bullets). Where applicable, the chapter includes or references illustrative examples or scenarios.

The first portion of this document includes this introduction and the pre-chapter. The pre-chapter is intended to provide readers with an overview of the key considerations influencing accountability system design, development, implementation, evaluation, and revision. The next section of this document is organized into chapters where the best practices and criteria are presented. The final section of this document includes a glossary to assist with language throughout. The following common terminology is provided to clarify how key terms are defined and used throughout this document.

TERM[1]	DEFINITIONS[2]
Theory of action	Also sometimes referred to as a theory of change, defines the mechanisms by which the accountability system will accomplish its goals and identifies the assumptions which must hold in order for the change agents to properly function. While sometimes used interchangeably with a logic model, a theory of action is more outcome focused, causal in nature, and articulates underlying assumptions that are determined by goals.
Claims	Statements or assertions we make about the system, system activities, and their intended impact or outcomes. Monitoring and evaluation efforts would then seek to identify evidence that supports confidence in a given claim.
Assumptions	Those conditions that must hold to confirm that connections between inputs, outputs, and outcomes for each sub-component satisfy the claim. For example, specifying the system vision presents several assumptions that must be proven.
Inputs[3]/ Resources	Inputs, which may also be referred to as resources, include those human, organizational, structural, and procedural resources a state has available to direct toward the activities for accountability systems.
Outputs	The direct products or results of activities and may include types, levels or targets of services to be delivered by the program.

[1] Terms are presented in conceptual order and are helpful to review before reading sections of this guide. These terms and several others presented again in alphabetical order in the Glossary section.

[2] All definitions are new unless otherwise attributed.

[3] Adapted from W.K. Kellogg Foundation (2004).

Outcomes	The specific changes in program participants' behavior, knowledge, skills, status and level of functioning. States should define the amount of time associated with short-term, mid-term, or long-term outcomes. The progression of the timeline from short-term to long-term outcomes should be based on the expected timeframe to build capacity (e.g., 5-7 years for new standards and implementation).
Accountability Stages[4]	The high-level categories of accountability system design. These are intended to group more specific development efforts and include stages like design, development, implementation, monitoring, support delivery, and revision.
Accountability Design[4]	The design stage includes refining the system's overall vision and theory of action, identifying and operationalizing indicators based on intended outcomes, and determining policy weights or decision rules to capture the SEA's values and priorities.
Accountability Development[4]	The development stage includes those activities that operationalize the accountability design. This can include examining indicator measures and relationships among them, identifying potential data gaps or capacity concerns through the use of simulations, and specifying performance expectations over time by setting defensible performance standards.
Accountability Implementation[4]	The implementation stage includes those activities that are associated with the release and publication of accountability data, results, and school identification. This may include activities like releasing reports; helping stakeholders access, use, and interpret information; and defining and delivering supports and interventions. These activities can help inform local inquiries and information use.
Accountability Evaluation[4]	The evaluation stage seeks to evaluate system design, development, and implementation stages. This may include substantiating the claims being made by using relevant evidence.
Accountability Activity	Activities are subsumed under stages and are intended to be more specific in nature. They typically have stand-alone outcomes that support information hand-offs and inform the next activities in the accountability system. They include what the state does with inputs/resources to support the accountability system. Activities are the processes, tools, events, technology, and actions that are an intentional part of the program implementation. These interventions are used to bring about the intended program changes or results.
Program[5]	A set of related measures or activities with a particular long-term aim.
Logic Model[6]	A systematic and visual way to present and share your understanding of the relationships among the resources you have to operate your program, the activities you plan, and the changes or results you hope to achieve. While often used interchangeably with a theory of action, logic models are more detailed and include more specific activities, outputs, and outcomes that are typically based off a program.
Components	A generic term that refers to the activities or programs associated with an accountability system.
Mechanisms	The relationships or connections between or among activities.

[4] From D'Brot (2018).

[5] From Oxford University Press (2020).

[6] Adapted from Kellogg Foundation (2004).

The Use and Limitations of this Document

Federal, state, and local accountability systems are rooted in statutory, regulatory, or policy requirements. A review of the legal issues and requirements governing the implementation of accountability systems is beyond the scope of this document. However, the practices and criteria presented in this document may reference the need to review local laws, regulation, or policies. Interpretations of best practices or their criteria require local contextualization and may benefit from the advice of local counsel regarding any relevant legal requirements.

Furthermore, the practices and criteria in this document are not intended to be a part of statutory or regulatory requirements nor are states expected to implement all of the best practices. They are intended to be reviewed critically and applied thoughtfully in consideration of system design features, known constraints, local requirements and other contextual factors. For example, some practices are framed specifically for a particular accountability system design and are not applicable to all states or systems. The Council of Chief State School Officers (CCSS), the Accountability Systems and Reporting (ASR) State Collaborative, and the workgroup wishes to emphasize that successful accountability systems may not demonstrate every practice or all of the criteria included in a given practice.

The practices and criteria in this document are intended to be reviewed critically and applied thoughtfully in consideration of system design features, known constraints, local requirements and other contextual factors.

Best practices for accountability system design, development, implementation, evaluation, and revision (DDIER) will continue to change as federal and state law is revised and state and local needs evolve. CCSSO, ASR, and the workgroup will adopt revisions to this document consistent with the methods employed by other best practices sponsored by CCSSO[7].

Workgroup Representation and Sponsorship

The Operational Best Practices in Accountability is sponsored by the Accountability Systems and Reporting (ASR) State Collaborative, which is organized by the Council for Chief State School Officers (CCSSO). The ASR collaborative represents accountability leaders from states across the country and partner organizations that support, inform, and work with those states. A workgroup made up of representatives from ASR contributed to and participated in the review of this document. Special thanks to Erika Landl, Senior Associate at the Center for Assessment for her careful review and contribution to Chapter 9.

[7] Please note, following these best practices will not guarantee approval by the U.S. Department of Education. These are meant to serve as guidance that states can implement to improve their own practices and processes.

PRE-CHAPTER: STATE CONSIDERATIONS FOR ACCOUNTABILITY SYSTEMS

The design, development, implementation, evaluation, and revisions (DDIER) of a state's accountability system provide a foundation through which State Education Agencies (SEAs) make decisions about school ratings, identification, support, and service delivery. Readers are encouraged to review *An Introduction to Accountability Implementation* (D'Brot & Keng 2018)[8] which presents three general stages of accountability systems – design, development, and implementation – that can be used to frame accountability related activities. For ease of reference, a brief summary of these stages is provided below. In addition, a fourth stage is defined which highlights the need for ongoing evaluation and is addressed in Chapter 8 of the OBPA. A fifth stage (i.e., revision) is addressed in Chapter 9 and can help states document and implement any proposed changes.

1. The **design** stage includes refining the system's overall vision and theory of action (e.g., policy priorities, educational system goals, role of accountability), identifying and operationalizing indicators based on intended outcomes (e.g., including growth, differentiating between college and career readiness, quantifying engagement) and determining policy weights to capture the SEA's values and priorities (e.g., growth and achievement should be equally weighted).

2. The **development** stage includes examining indicator measures and relationships among them (e.g., descriptive and inferential analyses, qualitative reviews of data and processes), identifying potential data gaps or capacity concerns through the use of simulations (e.g., projections, historical data examinations, mock accountability runs), and specifying performance expectations over time by setting defensible performance standards.

3. The **implementation** stage includes supporting the determination and release of reports; helping stakeholders access, use, and interpret information; and defining and delivering supports and interventions. These activities can help inform local inquiries and information use.

4. The **evaluation** stage includes considerations for evaluating system design and implementation, such as substantiating the claims being made throughout the design, development, and implementation stages of accountability system by using relevant evidence. While beyond the scope of this manual, evaluation efforts can also include examinations of the accountability system's impact and utility.

5. The **revision** stage includes documenting the rationale for proposed changes, determining and defining a change management process, modeling changes to the system, and evaluating the implications of the proposed modifications. These changes should then be documented and communicated to stakeholders. Any revisions to the system should be subjected to the same evaluation and revision strategies presented in this manuscript.

The purpose of this pre-chapter is to establish a foundation that assists staff in departments of education identify and consider appropriate best practices throughout the stages of accountability. It highlights important contextual factors and describes how they might influence the way in which best practices are represented in a state's accountability system. Important state-defined contextual factors include, but are not limited to state regulations, political climate and leadership, historical

[8] This document can be found here: https://ccsso.org/resource-library/introduction-accountability-implementation

factors, state policy priorities, locus of control (e.g., state vs. local), SEA capacity and structure, and funding. Readers are encouraged to use this resource on-demand and review chapters and topics that are relevant to their state's needs.

No two state accountability systems are the same – even if they are built to the same federal requirements and make use of the same indicators. Accountability systems are the end result of a long design process that is influenced by a variety of external forces considered deliberately to reflect a state's unique goals and priorities. The list below highlights key factors that influence accountability system design, development, and implementation and, consequently, how a state addresses the Operational Best Practices in Accountability (OBPA). The bullets do not summarize the chapters in the OBPA; instead, they offer preliminary considerations to assist readers as they plan for the different stages of accountability[9]: (1) Design, (2) Development, (3) Implementation, (4) Evaluation, (5) Revision.

No two state accountability systems are the same – even if they are built to the same federal requirements and make use of the same indicators. Accountability systems are the end result of a long design process that is influenced by a variety of external forces considered deliberately to reflect a state's unique goals and priorities.

Design

1. **The design of a state's accountability system will depend on how it intends to address the accountability requirements defined in federal and state law and, at the same time, consider the role of local accountability (e.g., regional or district requirements, accreditation, district accountability).**
   A state must have a clear understanding about how federal and state statutes can and should interact. As discussed in Chapter 1, a state may seek to develop one system that integrates multiple regulations, develop separate but complementary systems, or design completely distinct systems. This decision will depend on the state's goals for the accountability system (e.g., establish one unified system) and the degree to which there are conflicts, duplications, or parallels between the two sets of laws. For example, state regulations designed prior to ESSA may be retained to ensure longitudinal stability even if they do not support the development of well aligned state and federal systems.

2. **The goals, purpose, and uses of the accountability system will guide its design, development, and implementation. Federal and state requirements may influence timelines for implementation and who is involved in the process.**
   Accountability systems must be intentionally designed to promote and measure progress on outcomes believed to reflect attainment of, or progress toward, the state's overarching goals. In addition, if core outcomes are reflected in previous versions of the accountability system the state must consider whether previous definitions should be maintained or modified given the goals of the system.

[9] Additional examples may be included and updated for the purposes of the Accountability Systems and Reporting State Collaborative (more information about the ASR SCASS can be found here).

Chapter 1 describes the importance of articulating a clear theory of action for the state's accountability system where the goals, purposes, and intended uses both drive the system design and clarify the focus of system evaluation. As discussed in Chapter 2, intended outcomes should be informed and validated through input from a representative set of stakeholders.

3. **A state's vision for school support must be considered throughout the design of the accountability system.**
   A state's vision for school support must be considered during the design of the school accountability system to ensure that school support and identification procedures are coherent. Specifically, they should work together in a manner that both accounts for the resources available and reflects the state's goals and priorities related to identification and its intended role within the accountability system (e.g., auditor vs. technical advisor). A state's system of support is one of the many elements underlying a state's theory of action as addressed in OBP 1.3.

4. **Equity must be a core principle driving the design of the accountability system.**
   Throughout the design and implementation process state leaders must constantly evaluate whether the system design will serve to reduce achievement gaps and incentivize actions and interactions that facilitate equitable outcomes for all students and groups. Each of the operational best practices outlined in this document includes recommendations that reflect the importance of ensuring the accountability system reinforces and extends state efforts to improve equity.

5. **Stakeholders are necessary to validate assumptions regarding the accuracy, relevance, and usefulness of an accountability system and its results.**
   Stakeholder feedback is the way in which the voices of those most affected by accountability system results are incorporated into the system design. As discussed in chapter 2, stakeholders at multiple levels of the educational system (e.g., state, local, and school levels) should be identified and consulted throughout system design and implementation so their role and responsibility in supporting the attainment of intended outcomes is both understood and affirmed.

Development

6. **Business rules should clearly and accurately operationalize the design of the accountability system in a manner that reflects the State's priorities.**
   Business rules should be documented and validated throughout the system design and development process. They must be transparent to stakeholders and detailed enough to ensure consistency in implementation from year to year (or support independent replication). Chapter 4 outlines operational best practices focused on ensuring business rules and other operations are documented, reviewed, and validated throughout the three stages of accountability.

7. **Quality control and operations are critical to taking the accountability system from design to implementation.**
   The procedures used to collect and monitor data quality and integrity, confirm the fidelity of scoring and aggregation, and evaluate the validity of indicator measures must be established

as part of the system design. This information will inform decisions about what is appropriate to include in the state's high stakes accountability system. Chapter 4 discusses the range of factors a state must consider when developing an accountability implementation plan to ensure system results can be interpreted and used in the manner intended. This includes documenting who is responsible for quality control activities at different stages of the implementation cycle, when those activities will occur, and where the data/materials necessary to support them are stored.

Implementation

8. **Accountability reports rely on high quality designs, accurate data, and functional access. Without operational mechanisms that support translating data to information, accountability systems may not communicate performance or be accessible.**
   Accurate reporting of accountability information requires quality control of the accountability data and the reports themselves. When thoughtfully designed, access to high quality reports can help the public and educators make meaning from the data provided. Ensuring that reports are functional, sites are accessible, and revisions assigned to developers and testers are monitored can help facilitate the implementation of the accountability system.

9. **Accountability reports are the primary access point for the public and educators. A thoughtfully designed system will be less effective if results are not reported in an effective manner.**
   A state's theory of action should describe how different stakeholders are intended to use results from the accountability system to improve outcomes and meet the state's goals. System designers should be mindful of the information they want to communicate, how people access that information, and potential interpretations (and misinterpretations) of accountability reports. Chapter 7 highlights best practices underlying the development, evaluation and revision of accountability reports to ensure they provide stakeholders with useful, accurate information. Chapter 1 discusses the elements of a comprehensive theory of action, including assumptions underlying the intended use of system results.

Evaluation

10. **Evaluating accountability systems requires examining the individual decisions and assumptions underlying the system design. This can help clarify where and why decisions, dependencies, or information handoffs within the system may break down or function differently than intended.**
    Accountability systems are based on a complex series of steps, each with their own decisions, rationales, and constraints. The soundness of any one decision is contingent on the soundness of each preceding decision. Documenting these decisions and compiling evidence that the decisions facilitate desired outcomes bolsters the SEA's validity argument for its accountability system. Comparing the evidence gathered against the assumptions defined in the SEA's theory of action (and the rationales provided for different design decisions) may further support the SEA's argument that the accountability system is working as intended. This is described in more detail under Chapter 8.

Revision

11. **Ongoing, critical evaluation of the accountability system design is required to identify where/if changes to the system may be necessary.**
    Accountability systems are designed thoughtfully and attempt to use as much information as available and appropriate. However, they are typically built on a series of design assumptions that may not be testable until the system is operational or until a sufficient amount of time has passed (i.e., to ensure issues or anomalies are not due to idiosyncrasies in performance or data collection). Systematizing monitoring, evaluation, and metadata collection can help delineate whether or when changes should be made. Chapter 8 focuses on the evaluation of the accountability system, while Chapter 9 addresses best practices for making and managing changes, when required.

CHAPTER 1: ESTABLISHING AN ACCOUNTABILITY SYSTEM'S THEORY OF ACTION

Theories of action specify how programs or efforts are intended to function in order to bring about desired outcomes. They provide a road map that serves to explain how and why the design of a system is intended to provide for the attainment of a clearly specified set of goals. As theories of action are tested through system implementation, they must be continually evaluated to confirm that underlying assumptions hold. A broad theory of action for a state's school accountability system should be conceptualized and articulated as part of the system design stage using a process which includes at least the following steps[10]:

1. Describe the goals of the accountability system, including the desired impact of the system on districts, schools and stakeholders.
2. Articulate the purposes and intended uses of the accountability system and how they align to the state's goals.
3. Define the intended outcomes of the system (e.g., increases the rates of students that are college and career ready upon graduation).
4. Identify mediating outcomes or intermediate steps necessary to achieve the ultimate outcome(s) (e.g., increasing student access to AP, dual/concurrent enrollment, or CTE courses).
5. Create an initial "high-level" (large grain size) theory of action which summarizes the state's hypothesis as to how the desired outcomes will be achieved and supports the specification of major system components (e.g., inputs, indicators, ratings, supports, etc.) and how they are intended to relate to one another.
6. Build off the "high-level" theory of action and add enough details to articulate how the major components will be defined, operationalized and prioritized to reflect the state's goals.
7. "Zoom-in" on each component of the system and add the detail necessary to clarify the accountability system design and support implementation and validation (e.g., indicator measures, aggregation procedures, business rules).
8. Complete the chain of logic by articulating the assumptions which must hold in order for the system to function as intended (e.g., intended impact of the system on stakeholders, districts and schools; quality and appropriateness of program measures given the outcomes they are intended to inform; and the fidelity of system implementation quality and usefulness of score reports).

Detailing the goals, outcomes, and theory of action driving the design of a state's accountability system allows a state to determine, in advance, whether it will interact with other state and local educational initiatives as intended. If desired interactions are not possible, the theory of action serves to explain to policymakers and practitioners why this is the case and can identify potential alternatives. Furthermore, it is the integrity with which

Perhaps most importantly, the theory of action serves to ensure equity is established as a core design principle and the system is critically evaluated in light of this principle throughout development and implementation.

[10] Marion, S. M., Lyons, S., D'Brot, J. (2016). Developing a theory of action to support high quality accountability system design. Dover, NH: National Center for the Improvement of Educational Assessment. Retrieved November 15, 2017, from https://www.education.nh.gov/essa/documents/theory-of-action.pdf

the elements underlying a state's theory of action are operationalized and implemented and the extent to which those elements provide for intended and unintended outcomes that drive system evaluation (Landl, Domaleski, Russel, & Pinsonneault, 2016[11]). Perhaps most importantly, the theory of action serves to ensure equity is established as a core design principle and the system is critically evaluated in light of this principle throughout development and implementation. Three examples of theories of action are presented below.:

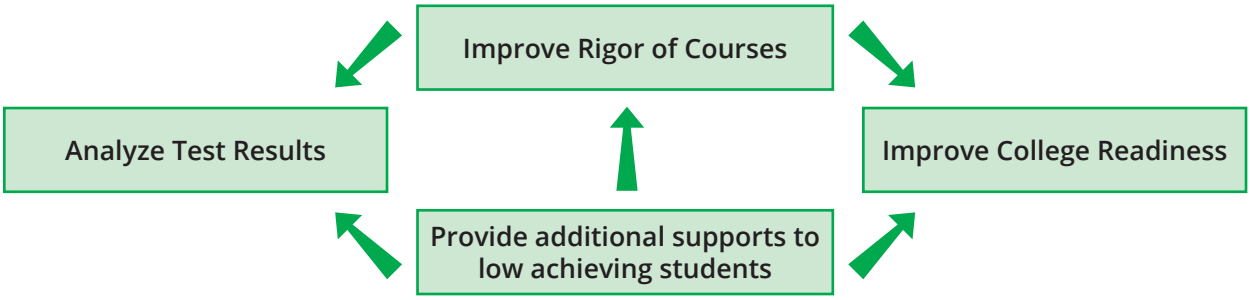

Figure 1. Example of a High-Level Theory of Action.

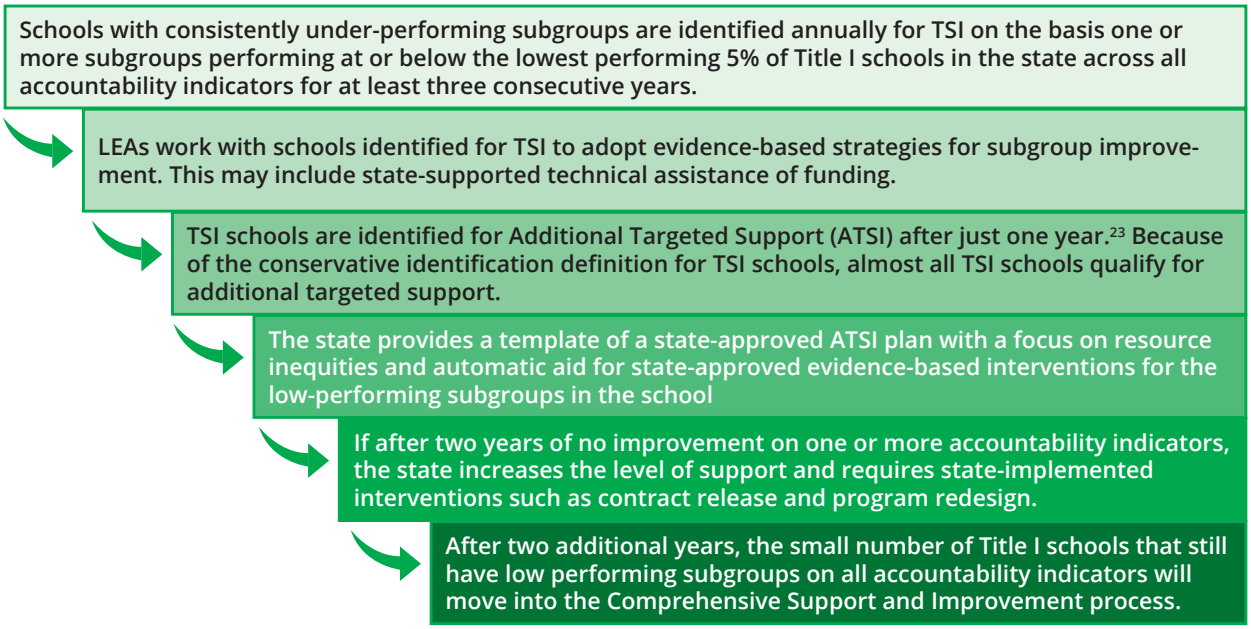

Figure 2. Example of a Theory of Action Prioritizing Moderate Identification with Strong State Support[12].

[11] See "A Framework to Support Accountability Evaluation" (Landl, et al., 2016) for a framework and example demonstrating how a theory of action can be specified to reflect a state's accountability goals and inform system design and validation. A Case Study based on Wisconsin's theory of action is provided in Appendix A. https://www.nciea.org/sites/default/files/publications/A%20Framework%20to%20Support%20Accountability%20Evaluation.pdf

[12] See Lyons, D'Brot & Landl (2016).

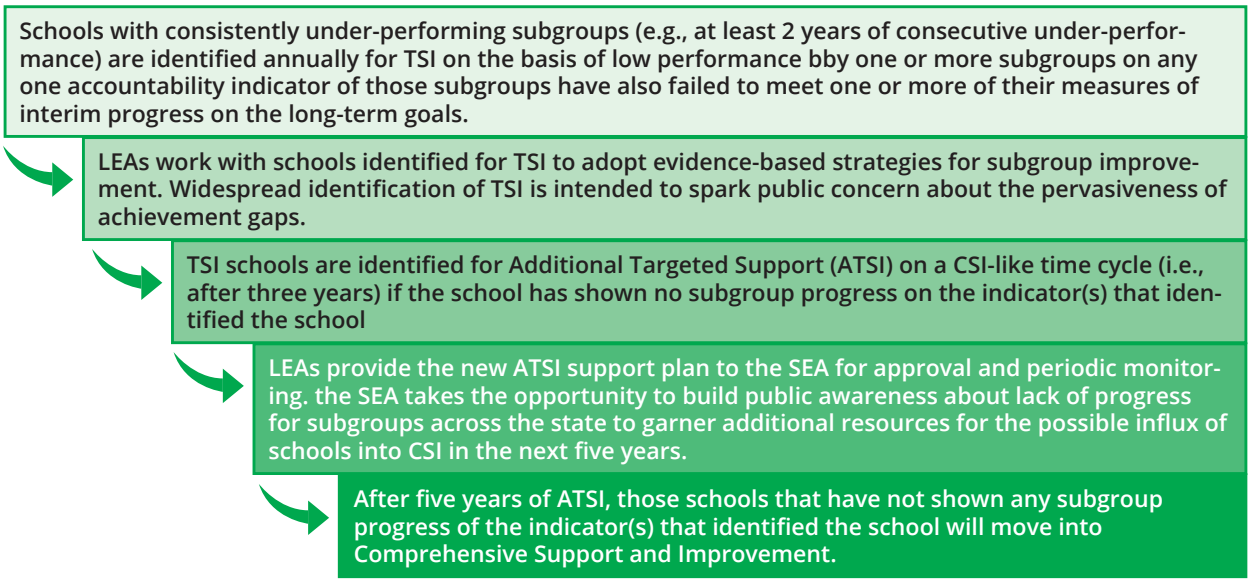

Figure 3. Example of a Theory of Action Prioritizing Widespread Identification with Strong Local Support[13].

The goal of this chapter is to highlight how considerations listed in the introduction might influence a state's theory of action or the factors that inform it. For additional detail about how to develop a theory of action for accountability systems, the reader is directed to Marion, Lyons, & D'Brot (2016). For examples of how a theory of action can be specified to reflect a state's accountability goals and inform system evaluation the reader is directed to Landl, et al. (2016).

1.1 Clarify the requirements underlying federal and state accountability systems as a means of determining how the systems can and should interact.

The intended scope of a state's accountability system defines whether and how much it will integrate, supplement, or supersede other educational accountability mechanisms. It also serves to constrain (or expand) the manner in which goals, intended purposes and uses, mediating outcomes, and other components of a state's theory of action will interact. While policy makers and stakeholders may have clear ideas regarding the desired scope of a state's accountability system (e.g., one unitary system), state and federal law represent the guardrails within which that scope can be defined. Through a clear understanding of what is required at the state and federal level, accountability designers can determine the degree to which these systems can be designed in a coordinated manner.

a. **Review federal legal requirements.** Federal statute and regulation will dictate the any requirements that must be part of the accountability system (or exist outside of the system). The state will need to identify which state priorities supplement or parallel federal requirements and which are in conflict. Depending on the state's priorities and vision for the accountability system, federal requirements could hinder or promote integration with state accountability systems and local initiatives.

[13] See Lyons, D'Brot & Landl (2016).

b. **Review state legal requirements.** State statute and regulation reflect the accountability goals and priorities of the state and may specify the inclusion of certain provider's indicators, measures, or procedures.

c. **Understand local initiatives and constraints.** Large local systems or groups of local systems may have existing accountability initiatives in place that may duplicate or be in conflict with state-level accountability efforts. Being aware of major local accountability initiatives can help states avoid duplication of efforts in their accountability system design stages.

d. **Cross-walk federal, state, and local requirements.** Once states have a thorough understanding of how federal, state, and local requirements interact, they should identify key elements that may support or constrain the design of complementary systems.

1.2 Clearly describe the primary goals of the accountability system. The system will need to be designed in a manner that both meets the state's goal and complies with state/federal mandates.

a. **Define the SEA vision and priorities.** As part of accountability design efforts the state should identify or clarify its goals and priorities *in consideration of any constraints or synergies identified through 1.1.* Priorities may be linked to larger governmental or educational initiatives in the state or across districts. Goals are driven by the state's vision and priorities and may be aspirational.

b. **Incorporate representative feedback.** Likely, the state will start with a set of priorities that are informed by gubernatorial, educational, or legislative leadership. The state should also reflect the views and opinions of key stakeholder groups in developing the goals of the accountability system. By integrating the state's and stakeholder's priorities, it is more likely that designers can promote buy-in for the accountability system overall. A discussion of stakeholder representation and involvement is presented in Chapter 2: Stakeholder Identification and Involvement.

c. **Establish a reporting or update structure.** As the accountability system matures through design, development, and implementation, the state should establish a schedule and structure to update key stakeholders that are included in defining the state priorities. This should be incorporated when addressing the best practices and criteria under Chapter 2: Identifying Stakeholders.

1.3 Articulate the purpose of the accountability system and the ways in which results are intended to be used.

a. **Clarify required purposes and uses based on federal and state legislation.** The state should articulate the purpose of the accountability system and how results from the system will be used. This may range from compliance with identifying the lowest performing schools to driving thoughtful improvement using a comprehensive set of data directly and indirectly related to the accountability system. Clearly articulating the purposes and uses can help minimize scope creep and define the role of the state's accountability system or how it interacts with federal requirements.

b. Clarify additional purposes and uses identified by state leaders and/or stakeholders. By clarifying additional purposes and uses, the state can determine whether it is reasonable to expect the accountability system design to include those additional purposes and uses, or if they are more appropriate for supplemental systems or support structures.

1.4 Specify the elements of the theory of action, indicating which can be measured, monitored, and evaluated.

a. Define key accountability outcomes. The state should identify the long term outcomes that provide evidence the accountability system is working as intended to meet the state's goals. Key outcomes may include increased graduation rates for all student groups, reduced gaps in academic achievement, greater participation in CTE courses or work-based learning, or higher scores on school climate or student satisfaction surveys. To ensure common understanding, accountability outcomes must be defined and operationalized clearly and consistently.

b. Define potential unintended outcomes. As part of the design phase it is useful to consider outcomes that are not intended, but occur as a result of the design, business rules, or procedures driving implementation. Potential unintended negative consequences (e.g., increased drop-out rates; declining satisfaction of educators; decreased access to courses for particular sub-groups) should be mitigated during the design phase, but also flagged for evaluation throughout development and implementation. Procedures should also be put in place to identify potential unintended positive consequences (e.g., increased rates of parent engagement), so the factors providing for those results can be identified and reinforced.

c. Specify mediating outcomes. To monitor and support the attainment of state goals, states will need to identify mediating outcomes that signal or influence progress toward desired long-term outcomes. For example, if the long-term outcome is improved rates of college- and career-readiness, mediating outcomes might include accelerated academic growth, greater educator access to high quality professional development, or a more uniform deployment of instructional support resources.

d. Specify the individual components of the accountability system, their role, weight, and intended relationship.

Considerations include the following:

- Identify the indicators and measures that reflect attainment of or progress toward mediating and long-term outcomes (e.g., weighted proficiency measure, graduation rate, percentage of students earning a credential, median school growth) and how they will contribute to the system (e.g., inform meaningful differentiation, informational only).
- Specify the relative emphasis each indicator should have within the system based on state priorities and discussions with stakeholders.
- Specify the hypothesized relationship among indicators in the system (e.g., decreases in chronic absenteeism should be accompanied by increased growth on the state assessment because students are receiving more instruction).

- Articulate inputs that are necessary to drive and support the accountability system (e.g., a new longitudinal data system, dashboard reporting tools, updated CTE content standards, professional development on the standards). What supports/resources/tools will be put in place in order for the accountability system to function as intended and/or to allow for the collection of required data?

e. **Define underlying assumptions that link the elements within the theory of action.** At this level, designers should specify the hypothesized mechanisms through which mediating and final outcomes are expected to occur (e.g., improved communication among educators; use of reported information to identify relevant, evidence-based improvement practices; teacher participation in professional development) and who is responsible for making them happen. Once these mechanisms are identified, one can determine the assumptions that must hold for the system to work as intended, and the evidence necessary to demonstrate those assumptions are being met.

   Assumptions about the technical quality (e.g., reliability and validity) and intended relationships among indicators; the quality, utility, and fidelity of implementation of specified inputs; and the use and/or perceived utility of system data and results should also be detailed to support evaluation.

f. **Specify the evidence necessary to confirm assumptions.** The outcomes, elements, and mechanisms that make up the theory of action will dictate what evidence should be collected to confirm assumptions. By confirming (or denying) assumptions, designers can determine where the accountability system is working as intended and where it may be yielding unintended consequences. For example an assumption may be that educators attend high quality professional development related to developing and administering performance based assessments. Supporting evidence may include rates of teacher participation by school and feedback regarding the perceived quality and utility of the training.

g. **Identify the programs related to the assumptions and elements in the theory of action.** Accountability designers should identify the individuals, offices, or programs that are associated or responsible for different elements in the theory of action (e.g., indicator measures and inputs). This will help determine the degree to which certain elements can be improved, influenced, or supported to reinforce accountability outcomes. For example, accountability systems typically rely on lagging indicators that are a function of school data collection efforts, varied LEA programs, and multiple SEA offices. The theory of action should specify not only who is responsible for collecting that data, but the assumptions which must hold in order for it to serve its intended role within the accountability system (e.g. the data is verified for quality and accuracy and meets the requirements defined within business rules). Similarly, if a new dashboard is put in place to support reporting and use of results, state and local responsibilities related to training, access and use of the dashboard by educators must be clearly defined.

CHAPTER 2: IDENTIFYING STAKEHOLDERS AND THEIR INVOLVEMENT

To be useful, accountability systems must provide information and incentivize actions which improve teaching and learning. Including stakeholders at key points throughout the stages of accountability ensures those influenced by the accountability system have a say in its design and what is reported. This type of participation is necessary to garner stakeholder support, establish face validity, and create a shared sense of responsibility for successful implementation. It also supports equity by ensuring representatives with varied backgrounds and perspectives have an opportunity to comment on the appropriateness and fairness of the system given its intended purpose. Finally, it provides a means of validating that defined roles and expectations (e.g., for districts, schools, teachers) are appropriate and feasible. The operational best practices within this section were developed to ensure the contribution of stakeholders is a thoughtful, significant part of state's planning and design efforts.

To be useful, accountability systems must provide information and incentivize actions which improve teaching and learning.

2.1 States should establish a stakeholder inclusion plan that identifies stakeholder groups and establishes their roles (e.g., advisory versus regulatory) and degree of involvement (e.g., informed, consulted, or included) throughout the stages of accountability.

a. Clarify state and federal requirements related to the inclusion of stakeholders. The level of stakeholder involvement can vary depending on the scope of the accountability effort. For example, federally required accountability systems have substantial stakeholder feedback requirements. Similarly, state accountability systems and related policies will likely require the SEA to collect public comments. These requirements can be leveraged to promote buy-in, inform the public about changes to systems, or inform design efforts.

b. Identify and establish an accountability task force, structure, roles, and responsibilities. Determine the appropriateness, size, and role of accountability task forces. These may include internal and external structures that serve different roles. For example, an external task force may be used to solicit ideas, feedback, and public concerns. An internal task force may then be responsible for pressure testing those ideas and determining what is feasible and appropriate based on data, policy, and privacy constraints. Additional considerations include the following:

- Determine how many stakeholder groups are necessary/desired and define their intended and differentiated roles (e.g., advising vs. decision-making) and composition (e.g., parents, educators, higher education, workforce, State Board, etc.).
- Assign a task force member(s) to each stakeholder group and determine the outreach structure (e.g., listening tours, town halls, computer mediated information gathering).

- Identify procedures and mechanisms for collecting, sharing, and responding to stakeholder feedback throughout the design process so involvement is not viewed as a compliance activity.
- Determine how and when materials will be revised based on stakeholder feedback. How will different recommendations be resolved and prioritized?

2.2 Select the members of each stakeholder group in light of its intended role and to reflect the composition, political spectrum, and diversity of the state population. Each citizen should be considered a consumer of both the educational system and the information it provides, which will influence how deeply and frequently stakeholder groups are involved, consulted, or informed.

a. Ensure the task force members understand their roles and responsibilities. Internal task force members should work together to determine the specific knowledge, skills, perspectives and backgrounds that should be represented in each group given its role and the type/manner of information desired. These may include

- advocacy representatives,
- relevant educational partners within the K-12 education system (e.g., union representatives, regional agencies, State Board of Education),
- relevant educational partners beyond the K-12 educational system (e.g., early childhood, post-secondary),
- appropriate business and community representatives,
- Local Education Agency and school-level representatives, and
- parent/family representatives.

2.3 The state should prepare the message, structure, and forum to share its initial SEA vision and priorities with each identified stakeholder group. While the detail, complexity, and timing may differ, the core of the message should remain the same.

a. Establish and communicate the overarching goals, purposes, and uses for accountability. The state's overarching goals, purposes and uses for accountability (as discussed in 1.1-1.3) should be established and agreed upon so it can be clearly and consistently represented in all types/levels of stakeholder correspondence. Differentiate between

- foundational elements that are non-negotiable, such as the state's priorities and core initiatives (e.g., equity, personalized learning, requirements); and
- components of the accountability system that are the focus and target of stakeholder feedback (e.g., reporting elements, weighting, design elements).

b. Test the SEA's vision and message. Consider presenting the initial SEA vision and priorities to an internal task force for comment on clarity and comprehensiveness. Use this as an opportunity to pilot the overarching message and materials and identify potential areas of confusion/concern.

c. **Develop supporting materials.** Determine how the internal vision and priorities will be presented to each stakeholder group given their intended roles and functions. Determine how messages and materials should differ for stakeholder groups. These groups may include
    - advocacy representatives,
    - relevant educational partners within the K-12 education system (e.g., union representatives, regional agencies, State Board of Education),
    - relevant educational partners beyond the K-12 educational system (e.g., early childhood, post-secondary),
    - appropriate business and community representatives,
    - Local Education Agency and school-level representatives, and
    - parent/family representatives.

2.4 Engage stakeholders as intended to refine/extend the vision, priorities, intended outcomes and theory of action consistent with the roles and expectations defined in 2.1 and 2.2.

a. **Design meetings with clear outcomes in mind.** Communication materials and meeting structures should be designed to elicit the type of feedback desired from a particular stakeholder group. For example, it may be more beneficial to have focused, rather than open discussion on statutorily required components for accountability. Consider pinpointing the questions/issues to be discussed and clarify why feedback from that stakeholder group is so important. If consensus recommendations are desired, this expectation should be established at the onset of the meeting.

b. **Prepare a communications plan in advance, specifically targeting stakeholders.** Establish a plan for summarizing and reporting meeting results back to participants, as appropriate. Make sure the way in which stakeholder feedback and recommendations will be used is clear from the onset.

2.5 In addition to stakeholder groups the state should identify technical advisors to help substantiate claims about the accountability system's validity, reliability, and fairness and contribute to the development of an overall validity argument.

a. **Establish and communicate the goals, roles, and responsibilities for a technical advisory group.** Determine the appropriate representation of the technical advisors based on state capacity and technical need. As accountability systems expand in scope or include components with a narrow focus, subject matter experts can help identify best practices or unintended consequences aligned to their expertise areas. Consider
    - Whether existing technical advisory committees should include accountability experts or whether a separate accountability TAC is necessary.
    - How to leverage cross-state collaboration opportunities when possible through professional networks.

b. **Establish the structure for the technical advisory group and how their feedback will be incorporated.** It is important to not only define the role of the technical advisory group, but also how frequently they will meet with the SEA and what the expectations are surrounding their feedback. The impact of their recommendations will have implications on the amount of information and lead up time the group will need and how they may frame their recommendations.

CHAPTER 3: ACCOUNTABILITY ROLES, RESPONSIBILITIES, AND PROGRAM MANAGEMENT

Due to the complexity of accountability systems, effective program management is a necessity. Programs are groups of related projects that are coordinated to yield benefits, control, and outcomes that are otherwise unavailable if managed individually (PMI, 2011[14]). Program management as a discipline is applicable across contexts, but requires specific content knowledge and experience to be implemented well. Therefore, well defined roles and responsibilities across accountability system activities and their related processes (e.g., data collection, data processing, policy development, etc.) should be defined and monitored throughout the stages of accountability. The operational best practices within this section define the responsibilities, decisions, and outputs associated with program management of accountability systems. Criteria in this section should be used in conjunction with the criteria outlined in chapters 4 and 7 which address operational quality control and report production, respectively.

3.1 State Education Agencies manage a substantial number of educational programs that address both compliance and improvement initiatives. SEAs should have a clear definition of the scope of their accountability system and specify how divisions and offices are organized in relation to the system.

a. Review the SEAs organizational chart. Having a clear understanding of the existing organizational structure within an SEA can help identify interactions across offices and divisions to support accountability processes and mechanisms.

b. Determine the level of intra-SEA collaboration necessary to support the system. In conjunction with reviewing the theory of action, a review of the organizational chart can help promote shared ownership of activities, programs, or policies supporting and resulting from accountability.

c. Determine the scope and ownership of accountability system components. There is a difference between indirect and direct ownership of accountability components, processes, and mechanisms. It is important to identify and recognize staff groups that have indirect impact on the accountability system but are still critical to effective operation. For example, accountability and school improvement staff groups will have direct oversight of identification and support efforts, whereas information technology or data management offices will have indirect responsibilities to support data collection, processing, and publishing. Similarly, assessment information plays a large role in many accountability systems, so it will be important to identify how changes to the assessment program or processes affect accountability. A responsibility assignment matrix, or RACI matrix, may be useful to help define roles and responsibilities[15].

d. Establish work plans and timelines that help manage expectations related to accountability implementation. In conjunction with 3.1.a – 3.1c, the use of project management techniques can help SEAs formulate work plans that can be easily monitored. For example, establish comprehensive project schedules that include

- accountability components and their associated outputs (e.g., assessment system data, data collection and verification of student demographics, administration and collection of student engagement data, etc.),

[14] The Standard for Program Management, Second Edition. Project Management Institute. 2011

[15] See the Project Management Institutes (2017) Guide to the Project Management Body of Knowledge.

- interdependencies across offices,
- non-negotiable milestones or deadlines,
- high priority activities, and
- additional sub-activities that contribute to activity completion or data processing and collection.

3.2 Identify specific roles and responsibilities attached to project milestones to enable project monitoring and help identify where project slippage may have occurred.

a. Review components of accountability systems to identify the need for responsible parties. As an extension to OBP 3.1, it will be important to assign roles and responsibilities to specific accountability activities and processes. This can be used in conjunction with a project plan that lists a comprehensive set of accountability activities to determine if similar tasks are grouped by office or if responsibilities are scattered across the SEA.

b. Review and crosswalk current offices/staffs with accountability components to determine workload and risk areas. Evaluate the availability and capacity of staff—especially for those with indirect impact on the accountability system—as a means of identifying areas where responsibilities could be expanded or should be constrained. This can help SEAs identify potential points of failure and determine whether revisions to existing work plans are needed.

c. Review and assign roles for production, development, and quality assurance based on system activities and needs. While this practice is explored in greater deal in chapter 4, it is important to create a work plan of assignments and roles by individual to help map out and navigate workload. The following table is an example of how specified activities might be organized and assigned using a cross-walked organizational chart.

ACTIVITY[16]	OFFICE RESPONSIBLE[17]	CONTACT PERSON
Accountability Design and Plan Development	Accountability, School Improvement, Information Technology, Leadership	
Quality Assurance of Assessment Data and Processes	Information Technology; Assessment; Accountability	
Assessment Policies and Administration Requirements	Assessment	
Analysis Plans and Accountability Calculations	Accountability	
Accountability Reporting	Accountability, Information Technology, Communications	

[16] These are sample activities and will differ based on the goals, objectives, and priorities of an agency.

[17] A responsibility assignment matrix (i.e., RACI matrix) can also be used to identify those who are responsible, accountable, consulted, or informed about specific activities or projects.

3.3 There may be aspects of the accountability system that are beyond the capacity of the SEA to develop, manage, or implement. In these cases, it will be important to identify whether external service providers are needed and to define their roles, responsibilities, and contractual obligations.

a. **Evaluate the need for service providers.** Based on a review of the design, development, and implementation activities, the SEA should determine whether it has sufficient capacity to successfully operationalize the system. Depending on the size, resources, or capacity of the SEA, it may be necessary to leverage external partnerships or secure contracts with service providers to help operationalize the accountability system. Examples of this might include web hosting services, reporting and visualizations, quality assurance, or replication services.

b. **Specify the role and scope of service providers (i.e., support the goals of the accountability system and balance costs).** For any external needs, the SEA should make clear the scope of services needed by external partners/service providers and how the scope supports the goals of the accountability system. Based on the scope, SEAs can specify the qualifications needed and make those available when soliciting efforts. Qualifications might include
    - background and expertise in relevant areas,
    - required or preferred certifications,
    - experience and references with similar projects,
    - experience developing documentation, processes, or reports consistent with the SEA's needs, and
    - any other types of experience or requirements (e.g., experience handling and storing secure data, familiarity with SEA protocols and requirements, a demonstrated preference for local service providers, program or project management expertise).

For any external needs, the SEA should make clear the scope of services needed by external partners/service providers and how the scope supports the goals of the accountability system. Based on the scope, SEAs can specify the qualifications needed and make those available when soliciting efforts.

c. **Based on the scope, define the contractual requirements for any external partner of service providers.** For any externally provided needs, the SEA should adhere to its local requirements for soliciting services or partnering with other agencies or organizations. This may include developing requests for proposals or information, establishing sole source agreements, establishing inter-agency agreements, or hiring contractors directly. It will be important to ensure the SEA's needs are met and services are obtained through means coherent with state requirements. Contractual or procurement considerations may include
    - history and context of the accountability program to contextualize SEA needs,

- security, legal, or regulatory specifications, which may include privacy, security, auditing, or other legal or procedural requirements,
- mandatory and optional requirements of the partner or service provider,
- ownership of data, documentation, code, technical specifications, or business rules,
- clear specifications for each activity and deliverable needed by the SEA, and
- cost requirements or cost proposals required to evaluate external efforts.

d. Define protocols for communication and data management (e.g., extractions, transfers, and loads) as necessary. Expectations for external partners or service providers should be established in any contractual negations. Because of the plan-oriented and data-heavy nature of accountability system management, protocols for communication and data management should be a major part of planning. These may include

- data production, extraction, transfer, and load schedules,
- interoperability expectations,
- handling personally identifiable information (PII),
- data elements, definitions, and usage rules, and
- data storage, destruction, purging, or disposal.

e. Determine an agreed upon format for any data transferred between an external partner or service provider and the SEA. While the operations and quality control for accountability system data are addressed in Chapter 4, working with an external partner or service provider will require well documented and agreed upon data formats and layouts. Files and specifications may include

- data layout specifications and codebooks, if applicable,
- technical specifications or business rules to develop any data files or products,
- the types of data collected, both in aggregate or individual, where applicable,
- data protection specifications, and/or
- any requirements or expectations to support a system that facilitates data storage, management, or review.

f. Specify how change requests are handled for any contractual agreement. Although the accountability system itself may require revisions or changes, it is important that the processes for managing changes to external partnerships or contracts with service providers are specified in advance. Ensure that expectations for both changes to delivering services and documenting those changes are available. Considerations may also include

- changes in staff or requirements in expertise,
- the timing of any changes and how or when the SEA should be notified in relation to work,
- changes to the scope of the work, whether expanded or contracted due to additional SEA needs or unforeseen complications, and
- modifications due to any other new federal, state, or local requirements.

CHAPTER 4: SELECTING AND INTEGRATING MEASURES FOR ACCOUNTABILITY SYSTEMS

Accountability systems are designed to communicate school performance coherent with federal or state requirements. However, the way in which information is prioritized and communicated varies depending on a state's theory of action. The best practices presented in this chapter should be reviewed in light of a SEA's theory of action and the relevant practices should be considered.

SEAs should think about the policy goals and intended behaviors driving the selection of system measures and how performance on those measures is expected to change over time.

To accurately represent the focus of accountability systems and the key inferences they intend to support (e.g., school quality, monitoring academic disparity, monitoring progress toward success in college and careers), states should carefully consider how measures are selected and combined. Specifically, what measures reflect the state's theory of action, the policy goals of the accountability system, and the SEA's intended outcomes? SEAs should think about the policy goals and intended behaviors driving the selection of system measures and how performance on those measures is expected to change over time. Additionally, it is important to consider how external stakeholders perceive the inclusion of certain data elements, which can affect how they interpret and use the results.

The operational best practices within this section were developed to help SEAs recognize and clarify constraints assumptions, and design decisions associated with the selection of measures that promote system goals. Criteria in this section should be used in conjunction with the theory of action and quality control best practices in Chapters 1 and 4, respectively.

4.1 Identify how/if measures in the state accountability system are intended to interact with federal requirements for accountability systems. This may reflect distinct, blended, or fully integrated federal and state accountability system designs.

a. **State systems should comply with federal requirements as per the state's theory of action.** In some cases, state accountability systems are separate from federal identification systems, whereas in others they are a single integrated system. While this is typically informed by the SEA leadership's vision and key stakeholder input, such decisions will have implications on what measures are appropriate for use. If integration across systems is desired, federal requirements can help define the minimum requirements to address for measures in state accountability systems.

b. **Based on federal requirements, identify relevant constraints for candidate measures and their associated data.** Depending on the types of constraints specified in federal statute (e.g., requirements for disaggregation, comparability, data collection), certain measures may not be appropriate to include in a federal identification system or in a system that integrates federal and state accountability. Depending on the accountability system's focus and design, non-compliant data may be more appropriate for supplemental reporting or as part of a separate state accountability system.

4.2 Identify state accountability requirements that are specified in law, regulation, and policy and how they impact measure selection.

a. Identify any relevant state requirements relevant to accountability design. Depending on the SEA's theory of action, there may be a distinct or supplemental state accountability system that is separate from federal identification. State requirements may be based on statutory, regulatory, or policy-based requirements that dictate what or how certain measures are calculated, used and reported in the accountability system. SEAs should determine whether there are constraints associated with measures selected specifically to support state accountability requirements, and how those requirements dictate data collection, measure inclusion, or reporting.

4.3 Once federal and state requirements are identified, states should determine how school quality will be defined, measured, and communicated based on the SEA's theory of action. This will inform the selection and calculation of system measures and help SEA's determine how those measures should be combined (e.g., through indices, composite scores, or decision rules) to reflect the relative importance of data and priorities defined within the TOA.

a. Ensure design decisions reflect stakeholder input and priorities, as appropriate. As noted in *Chapter 2: Identifying Stakeholders*, SEAs should collect and incorporate stakeholder feedback as appropriate, throughout the system design. This includes feedback about the manner and degree to which system measures provide valid information about school quality and the emphasis they should have in making overall school performance determinations. SEA must determine how operational constraints (e.g., data collection burden, privacy requirements, or inconsistent policy interpretations) will be balanced with stakeholder feedback and consider communicating this back to stakeholders as needed. When evaluating the selection and use of measures for accountability, SEAs should consider whether they

- consulted a representative set of stakeholders,
- included strategic representatives (e.g., advocacy groups or industry organizations), and
- communicated or updated stakeholders with changes to the system based on feedback, operational constraints and state/federal regulations.

4.4 Establish a clear rationale for why measures are included in the accountability system. There should be clearly stated reasons for why measures are included in accountability systems, which can range from compliance to incentivizing certain behaviors. SEA's must be able to clarify and articulate how each measure selected for inclusion in the states' accountability system contributes to the state's policy goals in a manner consistent with the theory of action. Engaging in the following activities can help SEAs articulate the rationale behind a) why specific measures/data are included in the system and b) the relative/weight or emphasis assigned to each. This, in turn, can help an

SEA determine the extent to which its argument will be clear and transparent when shared with stakeholders and the public.

- Provide examples of how people will interpret, use, and act upon data;
- Consider how the public and educators will react to the data included or the prioritized weight/order of data in the system;
- Project how data are intended to differentiate school performance over time;
- Describe why specific data elements will be reported; and
- Determine how data will be presented.

4.5 Determine and document the defensibility of the measures included in the accountability system. This defensibility should be based on whether measures are technically sound (i.e., reliable, fair, and valid) and whether they function as intended.

a. Evaluate the processes used to obtain data for measures to ensure accountability data can be interpreted correctly. Accountability data tend to be output-oriented measures that are dependent on a series of processes, data collection activities, and policies. It is important that the processes used to obtain data for each measure support consistent and comparable interpretations across the state. Consider enacting policy changes, increasing communication, or identifying professional development gaps based on issues like

- inconsistent interpretation of policies related to data in the system,
- inconsistent or inaccurate data entry or collection, or
- inconsistent or inaccurate quality control.

4.6 Evaluate candidate measures to ensure they demonstrate sufficient internal technical quality for use. In addition to the criteria in 5.4.a, data that comprise accountability systems should demonstrate reliability, fairness, and validity when compared to the construct the data are intended to represent. For example, determine the following for each measure:

- Whether the expected consistency in the data is sufficient to support interpretations around improved performance;
- Whether the level of precision is sufficient to differentiate school quality as intended;
- Whether data fairly represent the construct being measured. For example:
  - Are there potential sources of bias in the data?
  - Do schools and districts have an equal opportunity to demonstrate progress on a given measure?
  - Do all students have equal access to the construct with regard to age, race/ethnicity, language, gender, or other demogrpahic constraints, if applicable?

4.7 Consider whether the SEA's theory of action requires the combination of measures into composites or a series of decision rules. If so, determine and document how measures in the accountability system will be combined to meet the accountability system's goals.

a. **Articulate how combining measures communicates a meaningful grouping of information for the accountability system.** The SEA's theory of action, federal and state requirements, and definitions of school performance will dictate if and how measures should be grouped to form composite indicators. For example, systems prioritizing current achievement may only combine point-in-time school performance in core content areas. System prioritizing a combination of achievement and progress may use a combination of rules that equally weight point-in-time performance with student progress to determine school quality. Consider information interpretation, technical characteristics, and public perceptions where relevant when combining measures.

b. **Document evidence that the measures in the accountability system can be combined without introducing unintended negative effects when interpreting school quality.** Data characteristics within and across measures should be examined to ensure that they are contributing information to the definition of school quality as intended. This may include
    - examinations of within-measure data characteristics (e.g., measures of central tendency, measures of variability, shape, skew, and the presence of outliers),
    - the impact of data-characteristics on within-measure interpretations,
    - comparisons of data-characteristics across measures and across combinations of measures (i.e., indicators), or
    - the impact of combined data characteristics across indicators or overall calculation interpretations.

CHAPTER 5: ESTABLISHING PERFORMANCE STANDARDS FOR ACCOUNTABILITY SYSTEMS

States vary in their approach to producing school ratings. In some states, the accountability system culminates in a state-specific classification such as a letter grade (e.g., A-F), a star rating (e.g., 1 to 5 stars), or some other designation that communicates performance to the public. Other states do not provide an overall rating apart from federally required categories. Whether an overall or composite rating is provided, many states communicate performance using report cards or "dashboards" that often describe indicator performance in terms thresholds or levels (e.g., high/low; met expectations/ did not meet expectations; etc.).

...it is reasonable to require compelling evidence that the [accountability] rating has a high degree of validity for its intended interpretation and uses.

Given the central importance of the accountability rating at the indicator or overall level, it is 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 results from 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 (see Cizek & Bunch, 2007[18]), 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 other cases, states may face the inclusion of, or dramatic changes to, the measures that comprise indicators or the overall accountability system. In this case, states will need to revisit the performance standards that were initially set to determine whether they are still reasonable and communicate performance expectations as intended. States can also expect for schools to improve over time, which requires revisiting performance standards to ensure expectations are still tied to appropriate progress and performance goals.

This chapter provides a series of best practices to guide the establishment of a standard setting-process for accountability systems and references a framework for implementation (see Establishing Performance Standards, Domaleski, D'Brot, Keng, Keglovitz, & Neal, 2017[19]).

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

- Documenting rationale, procedures, and results;
- Ensuring the process allows participants to apply their knowledge and experience; and,
- Including information associated with relevant consequences and criteria.

[18] Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Sage Publications Ltd.

[19] Domaleski, C., D'Brot, J., Keng, L., Keglovits, R., & Neal, A., 2017. Establishing Performance Standards for School Accountability Systems. Council of Chief State School Officer: Washington, D.C. https://ccsso.org/resource-library/establishing-performance-standards-school-accountability-systems

Determining how the accountability system specifies performance standards for schools in an accountability system is a key step in communicating performance, school quality, and progress over time. States may also have limits in their authority for setting school performance standards, which is addressed at the end of this chapter. Operational best practices for defining and establishing school performance standards are provided below.

5.1 Like other performance standard setting processes, states should specify a design in advance of setting standards. This design should be vetted by key stakeholders, technical advisors, and provide the opportunity for states to check processes and decisions against the original design. The following reflect key priorities to consider as readers review 6.2-6.6.

a. Document rationale, procedure, and results. States will need to provide an explanation of the reasons for selecting a given standard setting approach, including justification for how the method fits the context and supports purposes and uses. Standard setting plans should include sufficient detail that reviewers and the public can understand how performance standards were set.

b. Ensure the process allows participants to apply their knowledge and experience. States should select individuals who are well qualified and 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).

c. Include information associated with relevant consequences and criteria. States should ensure that judges have access to relevant and supporting data for a range of schools and subgroups. These data should be used to provide sufficient context to judges regarding the meaning and consequences associated with each performance classification.

d. Ensure that the process, expectations, and activities are sufficiently transparent for stakeholders. Both the required oversight and the role of individual stakeholder groups will vary across and within states. As with other standard setting processes, it is critical that sufficient planning, documentation, and resources are available to inform the necessary stakeholder groups.

5.2 States will need to decide what performance categories should be established through the development of *Policy Definitions* or *Policy Descriptors*.

a. Determine what performance categories should be established and the general Policy Definitions 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. The state should consider the following when developing PDs:

- Develop brief descriptions consistent with the intended outcomes of the system.
- Include the appropriate level of review by key leadership, policy makers, or advocacy groups.

- Leverage the PDs as an anchor for standard setting efforts.
- Ensure the number of PDs reflects the necessary number of school ratings.

5.3 States should develop school performance level descriptors (SPLDs). School performance level descriptors are more detailed descriptions of what it looks like for a school to achieve each performance level in the state system.

a. Base SPLDS on the PDs. SPLDs are written at a finer level of detail than the PDs and should 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, which are defined below.

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

5.4 States will need to define imprecise terms like "high rates" or "meeting expectations." The development of SPLDs will help to unpack vague terminology by defining judgment-driven thresholds. States may consider various approaches to clarifying terms in the SPLDs, which may include the following:

- Associate performance expectations with goals external to the system (e.g., exceeding expectations occurs when the school or group meets a federally related proficiency 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); or
- Incorporate expectations that are central to federal or state policy in the SPLDs (e.g., a minimum graduation rate of 67% has significance in the *Every Student Succeeds Act*).

5.5 Document the development and refinement of SPLDs. Developing sufficiently detailed and coherent SPLDs may require a significant amount of work. States will benefit from collecting evidence that reflects the representative viewpoints, level of effort, and the degree of consensus in supporting sound standard setting practices.

5.6 States will need to identify a representative standard setting panel that can consider both the data and associated consequences of cuts. The state should convene a broad-based panel of leaders and stakeholders to evaluate information and make recommendations regarding performance expectations for the accountability system.

a. **Assemble a representative panel that reflects the state's interests and is able to articulate a vision for education in the state.** The state will need to identify a panel that is both representative of state demographics and strategic groups (e.g., advocacy groups, industry partners, higher education representatives). It will also be important to ensure that representatives have sufficient expertise and experience to understand the impact of their decisions, how to interpret data in the accountability system, and the standard setting process (see Chapter 2 for more detail on stakeholder identification).

b. **Identify an appropriately sized group of panelists for the standard setting.** Standard setting groups should be large enough to be representative and facilitate discussion using diverse viewpoints, but small enough to be managed effectively and efficiently. This will depend on whether a state uses more than one facilitator, if a larger group is split into smaller groups for discussion, or on the method selected for capturing recommendations.

c. **Include representation from prior steps in the process (e.g., PD or SPLD development).** If states have included stakeholders throughout the accountability design process, it may be valuable to include representatives in the standard setting event. The familiarity with the system, buy-in to the process, and contextual knowledge can be leveraged by newer participants and help promote alignment with the accountability system's design.

5.7 States should prepare appropriately for the actual standard setting event. In preparation for the standard setting event, states will need to generate multiple documents and resources that are needed to implement the standard-setting process.

a. **Include resources items like meeting agendas, facilitator scripts, presentations, and handouts.** These documents help ensure that designs, materials, and the intended outcomes can be reviewed prior to the standard setting event. Materials should be developed with sufficient detail that they can be used to conduct the standards setting from beginning to end. States should plan on including materials in their documentation of the standard setting event.

b. **Develop a sufficient number of school profiles to represent both typical and outlier schools.** School profiles are based on the characteristics of schools across the spectrum of school performance levels. These characteristics can include outcome data from the accountability system, demographic information, or additional data that may help participants better contextualize schools. Depending on the design of the accountability system, recommendations may be made at the measure, indicator, or overall system level. The level of granularity of the recommendations will impact the number and types of school profiles that will be needed. For example, the more granular the recommendation (i.e., at the measure level), the larger the number of sample school profiles will be necessary.

c. **Determine whether panelists will use final (i.e., validated) data or will rely on simulated (i.e., proxy) data and inform panelists accordingly. Plan appropriate validation strategies when using proxy data.** For new accountability systems or accountability systems based on significant changes, it may not be possible to use historical or validated data due to operational

constraints (e.g., new assessment program or new data collections). If using simulated data, establish a plan to validate or confirm that performance standard recommendations are appropriate once the accountability system is operational. For example, consider the following:

- Conduct a small-scale preliminary standard setting event using simulated data to determine rules and constraints for the full standard setting event;
- Conduct a full standard setting event using historical and simulated data and convene a standards review using operational data;
- Conduct both a preliminary and operational standard setting event; and/or
- Begin with a preliminary policy-defined set of performance standards that will be revisited when operational data can be used.

5.8 When states conduct the standard setting event, they will need to document the event sufficiently. When convening the standard setting event, the state will need to 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 and unbiased. The event should include the following activities with the panelists:

a. **Review and elaborate SPLDs.** This step requires an independent review of the SPLDs. This review will require participants to note key characteristics that distinguish schools at each level, identify areas that represent higher or lower priorities for each level, and note areas that need to be clarified or elaborated. A combination of small group and full group discussion should be supported that identify key points from the review. The goal of this activity is to help the group to come to consensus on any refinements or elaborations to the SPLDs.

b. **Independently identify threshold schools for each category.** 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. Possible approaches might include
   - Evaluating a series of anonymous school profiles and assigning a performance level to each school based on which level corresponds with the school's profile.
   - Specifying specific indicator or measure thresholds based on the SPLDs and applying participant-specified characteristics to yield a composite score. This will enable participants to review lists of schools based on composite thresholds to identify outliers and necessary adjustments to any rules.
   - Reviewing a list of school profiles ordered by overall score and locating the school judged to be the 'threshold' school that separates each level. This involves making a decision about the degree to which each school profile satisfies the criteria outlined in the SPLD.

The approach will be dictated by the standard setting method that may be adopted or adapted by a state.

5.9 Establish group recommendations. Facilitators during the process should share the outcomes from independent ratings based on participant ratings. This will also include inviting panelists to discuss results and share their rationale for affirming or revising recommendations. For this activity, states can reach a decision through multiple methods. These might include:

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

Similar to the prior practice, the method will be dictated by the standard setting method that may be adopted or adapted by a state.

5.10 Evaluate and document each step of the process. States should ensure that each panelist completes an evaluation of the process, which includes questions about their perceptions of the process (e.g., clarity of role, opportunities to share perspectives) and the results (e.g., recommendations are appropriate and defensible). The results of the evaluation are an important piece of validity evidence for the standard-setting process. Finally, states will need to collect documentation of the process and the results of the standard setting. 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.

5.11 Ensure that the standard setting design, process, and recommendations are incorporated into any state-specific approval or governance processes. SEAs may not be the final authorizing body when setting performance standards for statewide systems. Some states require approval by State Boards, external boards for accountability, or oversight governance boards. Planned design, documentation, and outcomes specified in 6.1.d and 6.6 should be addressed.

a. **When major changes to the system occur, states may need to revisit performance standards for the system.** States should define how much change is enough change to trigger resetting performance standards. Standards may need to be reset when the underlying data characteristics or measures change enough to impact the interpretation of school performance for a given cut score, threshold, or performance profile. Triggers to consider include

   - Changes to performance standards within indicators or measures of the accountability system, if performance standards affect how schools are evaluated;
   - Changes to the measures that comprise indicators in the accountability system;
   - The inclusion of new measures in the accountability system; and/or
   - The reformulation of indicators (i.e., composites of indicators) or business rules that differentiate school performance.

5.12 Upon making changes to the accountability system, conduct an internal review of the accountability system data and determine whether the change is sufficient to warrant a reset of the performance standards. For more minor changes (e.g., changes in data collection, policy, or widespread practice), the same data that would be used for a standard setting should be examined and compared to historical performance. If there are shifts that change the interpretation of measure, indicator, decision rule, or overall performance standards, it may be necessary to reset performance standards.

5.13 In certain cases, states will define performance standards normatively, through legislation, or based on some other policy-defined method. In these cases, the state should document the rationale behind this decision, how it aligns with their theory of action, and any evidence available that informed their decision. Include any corresponding state requirements, stakeholder rationales, and data that justify policy-driven cuts.

CHAPTER 6: ARTICULATING OPERATIONS AND QUALITY CONTROL IN ACCOUNTABILITY SYSTEMS

The implementation of accountability systems requires thoughtful consideration and intentional planning so that the system's annual outcomes reflect its intended purpose. This planning requires attention to operational infrastructures, such as data and reporting systems, processes, business rules, and validation procedures. Therefore, to successfully execute the vision, priorities and goals of the state, accountability systems requires an effective implementation plan. To support the state's validity argument the implementation plan should reflect the theory of action and design of the accountability system with fidelity by ensuring a) the ratings and information presented to schools and districts are correctly computed, b) schools and districts in need of support are appropriately identified, and c) system-based claims about the performance of schools and districts are accurate.

Because each state has unique priorities and requirements for its accountability system, this section does not provide a standardized set of procedures for operational implementation and quality control. Instead, this chapter references a framework—presented in Operations and Quality Control in Accountability Systems (see Keng & D'Brot, 2018[20])—that designers can use to guide the development of an accountability implementation plan and validate the various outcomes and components of their accountability systems.

Accountability system designers and implementers will need to consider a series of best practices (presented in sections 4.1-4.3) to help clarify the state's organizational structure and processes. These practices relate to the what, who, when, where and why of the accountability system. Specifically:

- **What:** tasks data, programs, processes, interactions necessary to ensure data quality;
- **Who:** the people, department, or organizations responsible for those tasks and elements;
- **When:** timelines associated with key tasks and their interdependencies;
- **Where:** sources of and repositories for data, documentation or materials; and
- **Why:** the rationale behind decisions that support the implementation of the system. While best practices are described in this chapter, considerations about the rationale are also raised in Chapter 1: Theory of Action.

Determining how the accountability system will be operationalized is a necessary step in producing intelligible and actionable information. Operational best practices for monitoring operations and quality control are provided below.

6.1 In order to support the soundness of decisions made with accountability results, procedures and timelines underlying the definition, collection and evaluation of input data must be fully specified.

a. Determine the what: Identify the data, information (e.g., business rules), and tasks necessary to compute and verify the accuracy and appropriateness of measures/indicators and populate reports in the system. It may also be important

[20] This chapter of the Operational Best Practices for Accountability is informed by the brief *Where the Rubber Meets the Road: Operations and Quality Control in School Accountability Systems.*

to publish and maintain documentation describing these data elements, information, or tasks. Consider convening stakeholder groups that were identified when accountability systems are designed (see Chapter 2: Stakeholder Involvement). Required data/information may include:

- Demographic information (e.g., gender, race/ethnicity, free and reduced lunch status, special education status, English learner status, full academic year status, etc.),
- Performance data (e.g.., large scale assessment data, postsecondary readiness assessments, English language proficiency assessments, interim assessment data),
- School characteristics (e.g., Title I status, grades served),
- School performance data (e.g., graduation rates, chronic absenteeism, aggregate course grades, school climate etc.), and
- Business rules.

Required tasks fall into two categories: a) those designed to evaluate the accuracy and completeness of input data and b) those conducted to evaluate whether the indicators/measures are appropriate, and fair given current specifications/ business rules. Please note that this is discussed in greater detail in Chapter 8: Monitoring and Evaluating Accountability Systems.

b. **Determine the who: Identify the people, department(s), or organization(s) responsible for supporting the quality of input data and articulate their roles/responsibilities.** It is important that the state also identify representatives who are responsible for maintaining, updating, and securing data and their documentation. Questions to consider include the following:
    - Who is responsible for providing each of the various data files?
    - Who is responsible for validating the accuracy, completeness and integrity of each of the various data files?
    - Who is responsible for establish programs and procedures that serve to merge, validate and assemble the various data files for use in generating accountability system results?
    - Who should be contacted if data files are missing, incomplete or include errors? Note that this can include people outside of an SEA, which would require additional collaboration and coordination efforts.
    - Who should have access to each data system and what type of access should they have?
    - Who is most appropriate to evaluate the quality and appropriateness of indicators, measures, and other elements that influence the accountability system?

c. **Determine the when: Establish calendars that reflect due dates for tasks informing the quality of input data.** States should establish data publishing calendars that are used for managing when schools and districts need to submit, revise, and verify their data. This also includes timeline information for internal processing and external communications. Due dates for the provision of internal and external data files and data processing timelines should be explicit and

transparent. It may also be helpful to include dates when data are actionable or validated for interpretation and use.

d. **Determine the where: Identify and document the source and location of data, documentation, and other relevant materials and resources supporting the quality of input data.** States should identify the content, level of detail, and accessibility of documentation related to accountability system data. This should include the original source of data, as well as any transformations, aggregations, or summarizations where relevant.

e. **Determine the why: Document the rationale for key decisions supporting the quality of input data.** States should be able to articulate why certain data elements, files, or indicators are used for the accountability system. Additionally, designers should be able to explain why data systems (and owners) were chosen for accountability implementation and why different users have access to different levels of detail in the data.

6.2 States will need to ensure that their data systems are working and managed securely. The systems define the process through which data are translated from input to output. Where those data are housed, who manages those data, and how they are managed are critical to operations and system implementation.

a. **Determine the what: Specify the rules for cleaning and processing data for accountability implementation.** States should be mindful of the formulas and logic used to compute indicators, as well as the rules for classifying and identifying students, schools, and districts. Key components to consider include the following:
   - Exclusion rules,
   - Match and merge rules,
   - Minimum n-sizes,
   - Student group inclusion and composition rules,
   - Exception handling, and
   - Indicator, performance-level, or summative ratings.

b. **Determine the who: Identify who is responsible for generating specifications and business rules to support the system's production.** States should identify who is responsible for generating and implementing business rules. In addition, states should identify internal or external partners who will conduct independent replication and validations of data runs to eventually sign-off on data, where appropriate.

c. **Determine the when: Document the timelines associated with processing data and testing business rules to support accountability implementation.** States should specify the timing of testing business rules or conducting dry runs of accountability calculations. This includes specifying when official (i.e., production) runs of business rules should be completed to support validation efforts of accountability files.

d. **Determine the where: Define the location and mechanisms supporting**

version control as edits are made to both the data files and any rules dictating changes to the data files. States will need to identify both storage locations and change processes to data, which may include issue tracking logs, formulas, models, and calculation rationales. These may be storied in documents that include

- technical documentation,
- research articles or internal studies, or
- peer-review evidence.

e. **Determine the why: Clarify the rationale for the business rules in the accountability system.** States should store and make transparent the reasoning behind business rules. Additionally, implementers of accountability systems should specify the rationale for matching across independent verifications of data analyses and signoffs between data sets each year.

6.3 States will need to ensure that their data systems are producing defensible output that yield sound data to its users. This will mean that the reported data are complete, accurate, and can support correct interpretations. As with the input and process stages, the output stages will require a clear definition of what, who, where, when, and why operations support defensible output.

a. **Determine the what: Identify the key data elements that should be reported, their reporting level, and the necessary supporting documentation.** Depending on the purpose and use of the accountability system, states will need to define precisely what accountability outcomes should be reported. It will be important to cross-walk what is reported against who is responsible for validating and who is intended to use the information (more information is provided in Chapter 7: Reporting). This will include reporting outcomes, reporting levels, and supporting documentation. Examples include the following:

- Accountability reporting outcomes might include aggregate data, disaggregated data, summative ratings, normative information, trends, or other metrics.
- Reporting levels may include district, school, building, teacher, student groups, or demographics.
- Supporting documentation may include training, interpretative guides, informational webinars, or social media outreach.

b. **Determine the who: Determine the people who are responsible for validating reporting accountability outcomes.** States should consider identifying individuals who are responsible for reviewing and approving (e.g., "locking down") accountability results. This will include identifying individuals responsible for communicating results, summarizing outcomes, explaining interpretations, highlighting uses of the results, and identifying contact personnel for questions. More information on how information might be used in discussed in Chapter 7: Reporting.

c. **Determine the when: Define the timeframe for processing data to compute outcome data and accountability results.** State will need to establish a timeline and calendar to identify target end dates for outcomes, results, and production of resources. These calendars can be used to work backwards to ensure that process and input timelines are cohesive and inform outcome production appropriately.

d. **Determine the where: Specify the storage location of final data and the location of public reports.** States should define access points, methods, and platforms that each user type can access accountability reports. This will require transparent documentation of final data file storage, reporting data, and documentation and resources that support accountability reporting and access.

e. **Determine the why: Clarify the rationale behind reports, their development, and production.** States should articulate the reasons behind reporting design and decisions. This may include the reasons for highlighting certain accountability outcomes, not including other accountability data, and why data are reported at key dates.

6.4 States should specify their goals, practices, and materials associated with quality control for data, processing, and documentation. This underlying characteristic of commitment to quality is essential to all stages of operational implementation. Well-defined and executed quality control procedures are part and parcel to the annual design, administration, scoring, and reporting processes associated with input and outcome data for accountability systems.

a. **Develop issue tracking logs.** Issue tracking logs can be used as a primary point of reference for regular touchpoints with internal and external accountability staff (e.g., daily calls, emails during processing, implementation concerns or revisions). These usually take the form of a shared spreadsheet or database in which any issues related to the annual implementation of the accountability system are recorded and tracked. Each entry in the log should include fields that might include the following:
    - The date of the issue (*date*),
    - A short narrative of the issue (*description*),
    - The person or group assigned to the issue (*responsible party*),
    - The tasks or processes impacted by the issue (*risks*),
    - The updated status of the issue (*status*),
    - A summary or comment about, each issue and
    - The mitigation approach (*resolution*).

b. **Define specifications in the form of detailed documentation for each portion of implementation.** The state should develop specifications, or "how-to" documents that include step-by-step instructions about how to perform tasks in the implementation plan, such as data cleaning, merging, computations, and classification. Detailed specifications about each task of accountability implementation are vital to documenting to quality and soundness of accountability processes.

c. **Establishing rules, personnel, and procedures for replication.** States should assign multiple people to independently carry out the steps described in the specifications and verifying that they yield the same results. Implementers should identify at least two independent replicators for most scenarios. However, new or substantially different accountability systems or requirements may warrant including additional replicators, potentially outside of the SEA. Factors to consider during replication include
    - Degree of match in data,
    - Software or system systems,
    - Defining acceptable tolerance for discrepancies, and
    - The use of real or simulated data.

d. **Set up conditions to implement test cases that simulate operational analyses.** States should specify the set of conditions or variables to test the accountability system. These conditions should sufficiently simulate real-world conditions to determine whether the system is working as intended. The activities associated with test conditions may include
    - Identifying expected values and outcome data,
    - Empirical simulations and data models, and
    - Test cases based on well-known schools or districts to identify typical results or outlier thresholds.

e. **Clearly establish thresholds for replications as part of a reasonableness review.** While considered under the replication section above, it is important that states engage in a reasonableness review. This review asks, "Do the accountability results make sense?" This requires a wider view of the accountability results by considering the meaning and implications of the outcomes and looking for patterns or trends that are unusual. The reasonableness review is usually more difficult that matching across replications, but is a critical step to avoiding unintended negative consequences.

CHAPTER 7: REPORTING AND COMMUNICATING ACCOUNTABILITY RESULTS

For accountability systems, reporting is often the first and only point of contact for the public. Educators and educational leaders are more likely to engage in deeper examinations of data that are either explicitly part of the accountability system or that inform progress against accountability indicators. For both groups, it is critical to develop reports that are accessible, understandable, and transparent. In conjunction with accountability report development, states will need to consider how reporting and supplementary resources fit into their larger communications plan aligned with their theory of action.

It is critical to develop reports that are accessible, understandable, and transparent.

Several factors should be considered when developing a communications strategy that is intended to help the public and education community understand the state's accountability system. For example, consider the following:

- What you are trying to communicate through accountability reports and supplemental reports?
- How you are facilitating access to reports and resources?
- How you are scaffolding access to and interpretations of accountability reporting and data?
- What do you hope people will do with the information that is reported?
- How you are differentiating information and access points based on the roles and responsibilities of various stakeholders?

The following operational best practices for reporting and communicating accountability results are intended to help SEAs consider not only the content and operations of reports, but also the intended message and behaviors associated with state reporting practices.

7.1 The scope and role of accountability systems vary significantly across states. States should establish a communications strategy that aligns with the state's theory of action and how the accountability system is intended to communicate school performance.

a. Identify the appropriate stakeholders who should have messages tailored to their needs and why. The level of detail needed to meet stakeholder needs will vary. Considering best practice 7.1.a, establish a list of potential stakeholders and group them by the way in which they will consume, interpret, and use reports from the accountability system. Stakeholder may include groups like

- educators,
- principals,
- LEA staff,
- other SEA staff,
- parents and families,
- the community and the public,
- higher education and business leaders, and
- advocacy and minority groups.

b. **Ensure accountability reports and outreach strategies serve to communicate the results and priorities of the accountability system as intended.** If reports and resources are not designed with an audience in mind and/or developers do not solicit user feedback, interpretations of schools' performance may be made using information that is not intended to be the most important. Consider
    - identifying key audiences and consumers of accountability data and their needs,
    - establishing a message tailored to each audience,
    - identifying the possible vehicles and access points for communication, and
    - establishing data expectations for different audiences (e.g., specificity, detail, and breakdown) based on their capacity, interest level and need.

c. **Ensure that stakeholder feedback regarding needs and high priority information are included in the accountability system reports as appropriate.** In many cases, states are required to solicit feedback when designing accountability systems, which can include report and resource design. Each report or resource should include messages and information tailored to any relevant stakeholder groups. When developing reports and resource for targeted groups, consider having representatives from each group involved in the design, review, and confirmation that reports are communicating information as intended.

7.2 Define the design considerations for accountability reporting and establish a clear hierarchy for current and future reporting plans and access points. It is important for states to distinguish between data that is required for accountability system reporting (e.g., proficiency rates) and supplemental data that will be supported by the state (e.g., participation or access to certain course options). SEAs must establish a plan and hierarchy that helps prioritize development efforts and contribute to establishing a data hierarchy that connects state required, state supported, and locally available information.

a. **Establish a hierarchy for how data are reported.** States can consider the following structure for reporting variations as an example
    - State required: Those components that are required for state and federal accountability reporting.
    - State supported: Supplemental information or data that can be used to support deeper explorations of school quality.
    - Locally available: Local Education Agency or school-based information that is conceptually related to accountability outcomes, but not part of the state system.

b. **Specify data that are required to be reported as part of the state accountability system.** State required data will include those data that are explicitly part of the accountability system's measures, indicators, or overall reporting or data explicitly defined in federal or state requirements. This will also include any relevant disaggregations or demographic information.

c. **Identify additional data that are not required (e.g., extends beyond high stakes accountability system data) but are supported by the state.** State supported data may include additional measures that are not used to calculate school performance. Instead, they may include data that contextualize school performance, are linked to high stakes outcome data, or are candidate measures for accountability systems that are not yet ready for high stakes use.

d. **Determine whether the reporting system will support any locally collected or reported data.** State systems will vary in their ability and desire to report locally collected data. Locally collected data may supplement accountability reporting by providing schools and districts the opportunity to communicate qualitative information about unique programs or rewards. Similarly, district- or school-specific progress monitoring data (e.g., early warning systems or related data) may be reported to systematize data exploration across districts.

7.3 Based on the overall communications strategy, focus of reports, and comprehensiveness of supplemental resources, the number and complexity of work plans will vary. States should establish clear and trackable work plans that support the production of accountability reports and associated resources as intended.

a. **Develop a systematic method to manage the report and resource production cycle.** As discussed in Chapters 3 (i.e., contractual change management) and 5 (i.e., operational timeline and data management), ensure that reporting is monitored through timelines, production schedules, monitoring tools, and transparent oversight processes. This will not only facilitate strong project management behavior but can also serve as a strong foundation for transparency. These work plans can be used to define what information and processes should be made available to the public to help promote trust in the accountability system.

b. **Leverage operations and quality control procedures to confirm data used to populate reports are accurate.** The quality of accountability reporting begins with the quality of source data. Ensure that the data used to populate accountability reports are accurate, free of errors, and appropriate for their intended use. Leverage the best practices in Chapter 4 to support the use of accurate data in accountability reports. Consider the following when preparing data for reports:
    - tracking business rules and their changes,
    - identifying individuals who are responsible for calculations and data transfer, and
    - providing sample data, code, or calculators to help districts and schools better understand the quality and meaning of data that are included and reported.

c. **Leverage operations and quality control procedures related to the development, hosting, and publication of reports.** There are many ways through which the public and educators can now access data and information. Ensure that the reports and any associated sub-reports or resources are hosted,

rendered, produced/printed, and made available as designed and without issue. Consider the following when developing and testing reports:

- Reports and hosting sites have been pressure tested for various rates of traffic;
- External IP addresses can access relevant reports without error; and
- Access formats (e.g., tablet, mobile, and computer access across operating systems) do not impede report generation or rendering.

7.4 While the development of accountability reports is an important step to making information accessible to the public, SEAs also try to make information interpretable and actionable for educators. It is important to consider the overall communication strategy to define the scope of the work plan (see OBP 7.1). For any additional resources, identify key stakeholders that play an integral part in improving outcomes in the educational system and design supplemental reports that align with the SEA's communication strategy.

a. **Identify high priority stakeholders that include or expand upon those identified in OBP 7.1.b.** Not all stakeholder groups benefit equally from increased information. Identify those high impact stakeholder groups (e.g., power data users, school improvement leads or teams, district improvement teams, principals, etc.) who would benefit most from additional reports or resources that serve to connect data across reports or systems in meaningful ways. Prioritize the SEAs limited resources accordingly to help support explorations of data meaningfully.

b. **Given the roles and responsibilities of high priority stakeholders, determine the data or strategies that would be most valuable to include or highlight in supplemental reports.** This may include data more frequently monitored and used at the district or school levels. Consider identifying the coherent data elements (e.g., leading and lagging indicators) for each measure or indicator in the accountability system to help expand the perceived relevance of school performance data in the accountability system.

c. **Establish specifications for reports and planned access points for key stakeholders that can provide for enhanced access based on role.** This will in part be informed by the state's overall communication strategy and supported data hierarchy (OBP 7.2). The types of data, purpose of the reports, and levels of access will vary based on a given state's student information system infrastructure. For example, robust information systems integrated learning management systems can leverage data connections to provide student-level reports to schools, school leadership teams, or teachers. In other cases, power users of data in districts and schools can be provided more detailed reports focusing on accountability data and other high priority information.

d. **Identify additional resources targeted to key audiences that can help educators and educational leaders use and interpret accountability reports and data.** In addition to clarifying the goals of the accountability system and specifying how certain data or measures should be interpreted, additional resources can also highlight sample data that may be available at the district level that can be used for progress monitoring or continuous improvement efforts.

These resources might focus on topics that include the following:

- Recommended examinations of data,
- Specific resources that can inform data-driven decision making,
- Evidence-based intervention-specific resources that leverage accountability data, or
- Suggested antecedent, process data, or recommended processes (e.g., interim assessment data, local improvement strategies, school leadership team strategies) that link local data to high-stakes accountability data.

7.5 As an extension of developing and implementing a reporting and communications plan, consider examining the degree to which users of the accountability system and its data understand and use data in a meaningful way.

a. **Determine ways to measure and track impact to determine whether your communication strategy is working.** Even the strongest messages and outreach plans should be monitored to be sure they are reaching target audiences. Identify potential measures of use (e.g., web hits, video analytics, resource downloads, application downloads, etc.) to establish trend data to determine if the public is accessing reports and resources.

b. **Collect feedback from targeted stakeholder groups.** An important step in understanding the impact of a reporting strategy includes understanding whether targeted audiences (e.g., improvement teams, LEA staff, school staff) see accountability reports and data as credible and usable. This should be considered in light of the overall evaluation plan discussed in Chapter 8. Consider the following to better understand where any revisions or supplemental resources may be necessary:
    - Determine whether targeted audiences understand accountability reports and data.
    - Identify whether data are "packaged" or presented in such a way that users will know what related data they should review for needs assessment or school improvement discussions.
    - Identify what resources help targeted audiences understand how to review accountability data in relation to needs assessment information.
    - Determine whether professional development or support efforts are designed to build the capacity of regional or district leaders to better understand accountability data and connect it to school improvement practices. If not, engage with stakeholders to define the role the SEA should play in supporting those efforts.

CHAPTER 8: MONITORING AND EVALUATING ACCOUNTABILITY IMPLEMENTATION

From a measurement perspective, validity refers to the ability of an instrument to measure what it purports to measure—however, validity is dependent on a comprehensive set of coherent evidence (AERA, APA, & NCME, 2014). An accountability system can be thought of as a measurement instrument that helps the public understand the degree to which schools and districts meet the state's educational goals and priorities (see Keng & D'Brot 2018). Therefore, establishing a validity argument for accountability system design entails identifying and connecting the pieces of evidence so SEAs can be confident that the schools' ratings are accurate, fair, and can be interpreted and used as intended. A validity argument is necessary for SEAs (and LEAs) to support the claim that school identification supports struggling schools and recognizes high performing schools.

An accountability system can be thought of as a measurement instrument that helps the public understand the degree to which schools and districts meet the state's educational goals and priorities.

An accountability system's complexity stems from the dependencies among activities through its design, development, and implementation. Documenting these decisions, and compiling evidence supporting outcomes for each decision, helps SEAs make a validity argument for their accountability and improvement systems. However, SEAs must first recognize key activities and their associated decisions, such as defining a system vision, identifying candidate measures, or setting performance standards for the accountability and improvement system.

The development of accountability systems begins with specifying a theory of action (See Chapter 1), which explains how the accountability system will affect desired outcomes. At a high level, the flow of information across components of an accountability system's theory of action (e.g., data, reporting, identification, support, and monitoring) can be examined and confirmed. At a more detailed level, states can focus on the major activities within the design, development, and implementation stages to determine whether assumptions about their impact and interaction with other activities in the system hold true. Each proposed relationship or interactions between activities (e.g., identifying measures through empirical analyses; empirical analyses informing simulations; simulations being used to inform performance standard setting) offers an opportunity to monitor information hand-offs that can clarify the claims and assumptions that must be supported.

Articulating the claims and assumptions underlying the design of the accountability system is necessary to developing a validity argument, as it informs the evidence that must be collected. States should consider identifying and documenting validity evidence using the following steps (D'Brot, 2018[21]):

- Define the state's theory of action for accountability systems;
- Articulate the claim(s) each activity or component of the accountability system was designed to support;
- Clarify the key assumptions that must hold to support each claim;

[21] D'Brot (2018). *A framework to monitor and evaluate accountability system efforts.* Dover, NH: Center for Assessment.

- Determine the appropriate evidence necessary to confirm assumptions and support each claim;
- Identify how/if evidence should be connected across claims to support a larger validity argument; and
- Document the examination and collection of evidence in a meaningful and transparent way.

The following operational best practices are intended to help states think through the identification, collection, and documentation of key pieces of evidence in support of establishing a validity argument for their accountability systems. Readers are also invited to review the *Evaluating State Accountability Systems Under the ESEA*[22] toolkit to support a comprehensive research and evaluation plan for an accountability.

8.1 Review the theory of action to identify the intended purpose and use of the accountability system. It will be important to identify what the SEA considers to be the accountability system, which will include its scope, how it connects to support structures, and how it supports continuous improvement. This will enable states to create a clear logic model that articulates how wide or narrow the accountability system should be (see Chapter 1: Theory of Action).

8.2 Once the scope of the accountability system is identified, the theory of action and accountability logic model can be used to identify the major components of the accountability system. This will enable SEAs to identify the claims that they are making through the accountability system.

a. Clearly define each activity necessary to operationalize the accountability system based on the theory of action. The activities of an accountability system will include meaningfully related tasks that enable SEAs to support the development, design, and implementation of accountability system. Activities can include many of the OBP chapter areas, like establishing a theory of action, identifying stakeholders, selecting and integrating measures, establishing performance standards, and reporting.

b. For each component, identify the claim(s) being made in the accountability system. Claims are the statements or assertions SEAs make about the system, or system activities, and their intended impact or outcomes. For example, a reporting claim may include, *Reporting is designed to communicate the goals and results of the accountability system with multiple users in mind.* This claim makes an assertion that likely includes several associated behaviors, like how people will interact with the system, what people will interpret from reports, and how widely the reports are available. Monitoring and evaluation efforts would then seek to identify evidence that supports confidence in this whether this claim can be supported.

c. Identify the assumptions underlying each claim and the evidence necessary to support them. Assumptions are those conditions that must hold in order to confirm that connections between the inputs, outputs, and outcomes for each activity satisfy the claim. The assumptions will help clarify the claim and identify

[22] D'Brot, J., LeFloch, K., English, D., Jacques, C. (2020). *State Support Network: Evaluating state accountability systems.* Washington, DC: American Institutes for Research.

relevant evidence. Once assumptions are clarified, the state can confirm that the activities associated with each claim will likely result in the intended outcomes.

8.3 A clear understanding of the accountability system's scope, key activities, and associated claims will help define the scope the SEA's research and evaluation agenda. It will be important to focus research and evaluation efforts on collecting evidence that supports design, development, and implementation activities that are of the highest priority to promote confidence in the accountably system's operations.

a. **Prioritize what research and evaluation activities are most important to promoting confidence in the accountability system.** There are a variety of research and evaluation strategies states might consider to validate the claims of their accountability system activities. It is likely through the design and development stages that states have already engaged in data collection and documentation that may only require review and would not need to be replicated. Prioritizing high need information can help confirm assumptions and validate claims with more nuanced activities (e.g., alignment of identified schools with intended performance profiles, report interpretation, intended behaviors associated with data).

b. **Determine the internal capacity and availability of external partners to support the SEA's research and evaluation agenda.** The required experience, expertise, or resources will vary depending on the types of research and evaluation strategies are necessary to collect evidence. SEAs should consider how to use limited internal resources most efficiently. For high priority activities that are resource intensive for activities where the SEA is restricted, identify high impact partners for support. For example, consider outsourcing complex analyses to local research partners and universities or soliciting support from regional or district networks to assist with comprehensive feedback plans.

8.4 The claims associated with each activity in the accountability system will require different types of evidence. It will be important to identify what evidence is best associated with each claim and how to collect the data or information that meets evidentiary needs. The evidence collected will vary by the analyses or methods applied. These pieces of evidence will be used to test and confirm assumptions for activities of the accountability system.

a. **Determine the most appropriate sources of data to inform the evidence collection against a specific claim.** The appropriate data will vary and will not always require quantitative analyses or complicated data collection. Consider the following as information is considered:
    - The necessary methodologies to collect relevant information (e.g., document reviews, checklists, focus groups, interviews, descriptive analyses, or inferential analyses);
    - Whether any specific empirical analyses meet evidentiary needs (e.g., path analyses, cluster analyses, factor analyses, correlations, descriptive statistics, thematic analyses); or

- Whether analyses focus on processes (e.g., business rules, checklists, or policies) or outcomes (e.g., simulations, descriptive counts, observed results).

8.5 Document the activities associated with the design, development, and implementation of the accountability system. Extend this documentation to include the progress of collecting evidence and how evidence supports claims across the accountability system.

a. Determine high priority documentation that helps to confirm the state's validity argument for the accountability system. It is likely that states will have a lot of existing documentation in support of a validity argument (e.g., the state's Consolidated State Plan, critical review documentation and reaction to the system, stakeholder feedback, etc.). In conjunction with the other best practices in this chapter, it will be important to delineate between what has already been collected and what still needs to be collected to support sufficient documentation.

b. Compile documentation in a meaningful and organized manner and ensure that it is updated appropriately. Documentation in support of the state's accountability system will likely be a combination of internally-facing and externally-facing information that serve different purposes. Organizing documentation in a way that highlights its purpose and use may be helpful. Consider included the following types of information:

- Internal documentation (e.g., business rules, syntax, historical data, operations and quality control documentation),
- External documentation and resources (e.g., annotated business rules, public releases, public write-ups or documentation),
- Federal- or state-required documentation (e.g., consolidated state plans, state policy, compliance with regulation checklists, guidance documents), and
- Resources to support interpretation and use of the accountability information.

c. Ensure that complied information and documentation detailing the design, development, and implementation of the accountability system is accessible and available to the public.

CHAPTER 9: ENGAGING IN ACCOUNTABILITY SYSTEM CHANGE MANAGEMENT

As a result of the ongoing process of system monitoring and evaluation a state may determine that modifications are necessary. This may especially be the case after the initial year of implementation, which is often the first time operational data is available for all components of the system. Evidence may suggest that one or more indicators are not functioning as intended, effective weights differ significantly from nominal/policy weights, and/or procedures for identification over-flag certain types of schools. The state may decide to add new indicators of system quality, revise business rules to improve inclusion rates, or revise state and federal accountability reports based on feedback from stakeholders.

Design changes must be vetted, evaluated, well documented, and communicated to the public in a clear understandable manner that highlights the importance of ongoing evaluation and improvement.

While changes are often necessary, they cannot be taken lightly as the implications can be significant (e.g., changes to a school's identification status, frustration/backlash from stakeholders) and the potential for error large (e.g., changes to system code, data layouts). Design changes must be vetted, evaluated, well documented, and communicated to the public in a clear understandable manner that highlights the importance of ongoing evaluation and improvement.

9.1 Document rationale for a proposed change to the accountability system.

Each proposed change to the system should include the following:

- Describe the data/evidence/background highlighting the need for the change.
- Discuss the potential negative implications of not addressing this issue.
- Describe how the proposed modification addresses the issue of concern
- Determine whether changes to the accountability system are within the scope of business rule changes or if they more globally impact the design of the system, which may require internal, external, or official review and approval.

The following table provides an example of how a state may walk through this change management process.

PROPOSED MODIFICATION	EVIDENCE/BACKGROUND FOR CHANGE	NEGATIVE IMPLICATIONS OF NOT MAKING CHANGE	HOW TO ADDRESS
Adjust indicator weights	There is a significant difference between effective and nominal weights for some indicators due to range restriction.	Index scores cannot be interpreted as representing the intended emphasis across indicators.	By adjusting indicator weights, the policy weights will be represented in the overall index scores.
Add an additional measure of readiness to the CCR index	Stakeholders wanted to include ASVAB performance as a measure in the state's CCR index but the data was unavailable for evaluation until this year.	Schools do not receive credit for student performance on the ASVAB which was highlighted as a priority by stakeholders.	Include attainment of a score of X on the ASVAB as one measure of readiness within the CCR.

9.2 Determine and define a change management process.

a. Identify relevant elements and their associated change. For each proposed modification, identify each element of the system that is impacted (e.g., input files, coding, business rules, reports, etc.), the type of change that will be needed, who is responsible for making the modification to each element, and how it will be evaluated (e.g., from a quality control and technical quality perspective).

b. Determine if the change requires a formal amendment to the existing state plan for accountability. The formality of an amendment should include the following considerations:

- Identify the constraints based on how the current state plan is written and the type/degree of change under consideration.
- Determine whether the change requires stakeholder input and approval prior to submission to USED.

The following table outlines a series of examples under which states may need to consider the degree of change to their system.

PROPOSED MODIFICATION	ELEMENTS OF SYSTEM IMPACTED	WHO IS RESPONSIBLE	DOES THIS REQUIRE AN AMENDMENT?
Adjust indicator weights	Program used to determine overall school scores/ratings	State	Yes. Modification to the state's system of Annual Meaningful Differentiation requires an amendment.
Add an additional measure of readiness to the CCR index (based on ASVAB performance)	State-developed documents describing the components of the system Input data file and record layout supporting school accountability calculations Programs used to calculate overall school score/rating Business rules documentation (e.g., score that indicates meeting/not meeting readiness; who is eligible, etc.) Data files provided by district	State- Person District data lead	Yes, if changes are being made to the system operationally. No if the plan is being evaluated, but will not yet be implemented.
Revisions to business rules	Processes and input data files to the accountability system Identification criteria of students to be included in school calculations Assignment of students to schools	State – person	No, if modifications to business rules do not have a substantive impact on the components that make up the accountability system. Yes, if modifications to business rules are significant (e.g., N-size).

9.3 Model candidate changes in the system.

a. **Use operational data to evaluate how each proposed modification should be operationalized to address the issue at hand.** When possible, it will be valuable to use operational data to understand the implications of potential changes to the system. While changes often stem from adjustments intended to address issues or unintended consequences, modeling or simulating changes can help identify other potential unintended consequences.

b. **Compare and contrast the implications of different options identifying the pros and cons of each.** With something as complex as accountability systems, it is likely that many options will be somewhat appropriate to implement. Understanding the strengths, weaknesses, opportunities, and threats associated with each option can help designers determine the option with the least risk that also meets the state's goal.

c. **Once all options have been evaluated, model the combination of all proposed candidate changes simultaneously.** As with any complex system, a single change can have a chain reaction with other interrelated components of the system. Therefore, the state should examine the candidate changes using operational "practice runs." This will enable the state to engage in an evaluation of the decisions or revisions to ensure changes better reflect the state's theory of action.

9.4 Evaluate the implications of the proposed modifications (independently and as part of the full system).

a. **Determine the impact of the change on the relationships among indicators.** While OBP 9.3 focused on the impact to data elements and indicators, OBP 9.4 takes a wider view of the impact changes may have on other indicators. States should confirm that changes to one indicator do not have an adverse effect on the relationship with other indicators in the system.

b. **Determine the impact on final school ratings or judgments.** As a result of changes to how indicators interact with one another, the state may find that changes will result in impacts to how schools are rated or judged. It will be critical to determine if these are intended or unintended changes to school-level determinations.

c. **Finalize proposed modifications including calculation procedures and business rules.** Once a state has identified the most appropriate set of options, change scan be incorporated into calculation procedures. This will require documentation, updates to business rules, and changes to all necessary communications and resources. Maintaining a tracking log or repository of accountability-related documents can be helpful in supporting change management efforts (see *Chapter 6: Articulating Operations and Quality Control in Accountability Systems*).

9.5 Develop and implement a communication plan for proposed and accepted changes.

a. For each accepted change, indicate the type of communication and outreach that is necessary. Depending on the scope and focus of the change, developers, implementers, and users of the accountability system and its results will need to be informed. The state should consider whether communications need to focus on either of the follow, or potentially both:

- Internal communication and documentation only or
- Widespread public communication that may include a formal press release describing the change, the rationale behind the change, and implications on outcomes or understanding.

9.6 Engage in and adjust any monitoring and evaluation activities as a result of the enacted changes.

a. Evaluate the operationalization of any proposed change in the system. In conjunction with the OBPs and criteria suggested in Chapter 8 (*Monitoring and Evaluating the System*), states will need to evaluate whether any changes are having the desired impact.

GLOSSARY

Accountability Activity	Activities are subsumed under stages and are intended to be more specific in nature. They typically have stand-alone outcomes that support information hand-offs and inform the next activities in the accountability system. They include what the state does with inputs/resources to support the accountability system. Activities are the processes, tools, events, technology, and actions that are an intentional part of the program implementation. These interventions are used to bring about the intended program changes or results.
Accountability Design[23]	The accountability stage includes refining the system's overall vision and theory of action, identifying and operationalizing indicators based on intended outcomes, and determining policy weights to capture the SEA's values and priorities.
Accountability Development[4]	The accountability development stage includes those activities that operationalize the accountability design. This can include examining indicator measures and relationships among them, identifying potential data gaps or capacity concerns through the use of simulations, and specifying performance expectations over time by setting defensible performance standards
Accountability Evaluation	The evaluation stage seeks to evaluate system design, development, and implementation stages. This may include substantiating the claims being made by using relevant evidence.
Accountability Implementation[4]	The implementation stage includes those activities that are associated with the release and publication of accountability data, results, and school identification. This may include activities like releasing reports; helping stakeholders access, use, and interpret information; and defining and delivering supports and interventions. These activities can help inform local inquiries and information use.
Accountability Stages[24]	The high-level categories of accountability system design. These are intended to group more specific development efforts and include activities like design, development, implementation, monitoring, support delivery, and revision.
Assumptions	Those conditions that must hold in order to confirm that connections activity between the inputs, outputs, and outcomes for each sub-component satisfy the claim. For example, specifying the system vision accurately assumes.
Business Rules[25]	Business rules provide the foundation for automation systems by taking documented or undocumented information and translating it into various conditional statements.
Claims	Statements or assertions we make about the system, system activities, and their intended impact or outcomes. Monitoring and evaluation efforts would then seek to identify evidence that supports confidence in a given claim.
Components	A generic term that refers to the activities or programs associated with an accountability system.

[23] From D'Brot & Keng (2018).

[24] From D'Brot (2018).

[25] From IBM (2020).

Compensatory Performance	When considering how indicators interact to inform school judgments or ratings, higher performance in some areas can offset lower performance in other areas.
Conjunctive Performance	When considering how indicators interact to inform school judgments or ratings, minimum performance must be observed in all areas in order to meet performance expectations.
Disjunctive Performance	When considering how indicators interact to inform school judgments or ratings, minimum performance can be observed in any area in order to meet performance expectations.
Evaluation[26]	Systematic investigation of the value, importance, or significance of something or someone along defined dimensions (e.g., a program, project, or specific program or project component)
Inputs[27]**/ Resources**	Inputs, which may also be referred to as resources, include those human, organizational, structural, and procedural resources a state has available to direct toward the activities for accountability systems.
Logic Model[28]	A systematic and visual way to present and share your understanding of the relationships among the resources you have to operate your program, the activities you plan, and the changes or results you hope to achieve. While often used interchangeably with a theory of action, logic models are more detailed and include more specific activities, outputs, and outcomes that are typically based off a program.
Mechanisms	The relationships or connections between or among activities.
Outputs	The direct products or results of activities and may include types, levels or targets of services to be delivered by the program.
Outcomes	The specific changes in program participants' behavior, knowledge, skills, status and level of functioning. States should define the amount of time associated with short-term, mid-term, or long-term outcomes. The progression of the timeline from short-term to long-term outcomes should be based on the expected timeframe to build capacity (e.g., 5-7 years for new standards and implementation).
Program[29]	A set of related measures or activities with a particular long-term aim.
Quality Control[30]	Ensuring that all stages of an operational implementation workflow have the highest commitment to quality. Threats to quality (e.g., errors in data, misspecifications or misunderstandings of business rules, lack of stability in data or systems, etc.) are identified, acted upon, monitored, and documented. Quality control processes include issues tracking logs, replication, test cases, and reviews of reasonableness.

[26] From Yarbrough, Shula, Hopson, & Caruthers (2010).

[27] Adapted from W.K. Kellogg Foundation (2004).

[28] Adapted from W.K. Kellogg Foundation (2004).

[29] From Oxford University Press (2020).

[30] From Keng & D'Brot (2018).

Replication[31]	Assigning multiple people to independently carry out the steps described in the specifications and verifying that they yield the same results. Two independent internal replicators are usually sufficient for most implementation scenarios, especially if there are staffing or resource constraints, such as available software licenses.
Simulation[32]	The imitation of the operation of a real-world process or system over time, which often require the use of models
Theory of action	Also sometimes referred to as a theory of change, defines the mechanisms by which the accountability system will accomplish its goals and identifies the assumptions which must hold in order for the change agents to properly function. While sometimes used interchangeably with a logic model, a theory of action is more outcome focused, causal in nature, and articulates underlying assumptions that are determined by goals.

[31] From Keng & D'Brot (2018).

[32] From Banks, Carson, Nelson & Nicol (2009).

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Banks, J., Carson, J., Nelson, B., & Nicol, D. (2009). *Discrete event simulation.* (5th ed.) Prentice Hall: Upper Saddle River, NJ.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage Publications Ltd.

D'Brot, J. (2018). *Accountability validation: Evidence as a function of design.* National Center for the Improvement of Educational Assessment: Dover, NH.

D'Brot, J. M. & Keng, L. (2018). *An introduction to accountability implementation: A preface to the operations, performance standards, and evaluation resources.* Washington, D.C.: Council of Chief State Schools Officers.

D'Brot, J., LeFloch, K., English, D., Jacques, C. (2020). *State Support Network: Evaluating state accountability systems.* Washington, DC: American Institutes for Research.

Domaleski, C., D'Brot, J., Keng, L., Keglovits, R., & Neal, A., 2017. Establishing Performance Standards for School Accountability Systems. Council of Chief State School Officers.

IBM (2020). Business rules. Retrieved from https://www.ibm.com/cloud/learn/business-rules

Keng, L. & D'Brot, J. M. (2018). *Where the Rubber Meets the Road: Operations and Quality Control in School Accountability Systems.* Washington, DC.: Council of Chief State Schools Officers.

Landl, E., Domaleski, C., Russell, M., & Pinsonneault, L. (2016). *A framework to support accountability evaluation.* Washington, D.C.: Council for Chief State School Officers.

Lyons, S., D'Brot, J., & Landl, E. (2017). *State Systems of Identification and Support under ESSA: A focus on designing and revising systems of school identification.* Washington, D.C.: Council of Chief State School Officers.

Marion, S. M., Lyons, S., D'Brot, J. (2016). *Developing a theory of action to support high quality accountability system design.* Dover, NH: National Center for the Improvement of Educational Assessment.

Oxford University Press (2020). Program definition. Retrieved from google.com.

Project Management Institute (2011). *The standard for program management* (2nd ed).

Project Management Institute. (2017). *A guide to the Project Management Body of Knowledge* (PMBOK guide) (6th ed.).

W.K. Kellogg Foundation (2004). *Logic model development guide.* Kellogg Foundation: Battle Creek, MI.

Yarbrough, D.B., Shula, L.M., Hopson, R.K., & Caruthers, F.A. (2010). *The Program Evaluation Standards: A guide for evaluators and evaluation users* (3rd. ed). Thousand Oaks, CA: Corwin Press.

APPENDIX A:
THE WISCONSIN ACCOUNTABILITY SYSTEM CASE STUDY:[33]

To provide additional clarity and context for applying the evaluation framework, we present a case study based on Wisconsin's current school accountability system. We start with a review of the process that established the foundation for the system and clarified the goals and theory of action. Then, we focus on the intended outcomes and the program measures selected to track these outcomes. Finally, we illustrate some potential sources of evidence as part of an ongoing monitoring and evaluation process. Throughout the case study narrative, we highlight linkages to the evaluation framework in shaded boxes.

Background: Determining Goals and Theory of Action

The design of Wisconsin's current accountability largely reflects a shift in accountability priorities resulting from discussions in the state in 2011. That year, heeding calls for a Wisconsin-specific accountability system, the State Superintendent of Public Instruction, Governor, and chairs of the Senate and Assembly education committees convened a School and District Accountability Design Team. This group, comprised approximately 30 education stakeholders representing various education entities, school and district roles, and student populations, discussed key goals and principles of an accountability system "of and for" Wisconsin.

According to the Accountability Design Team, a quality accountability system will:

- Support high-quality instruction in all publicly funded schools and districts;
- Include all publicly funded students in accountability calculations;
- Measure progress using both growth and attainment calculations;
- Make every effort to align this work with other state educational reform initiatives;
- Align performance objectives to career and college readiness;
- Focus on and include multiple measures of student outcomes that can be used to guide and inform practice and for accountability purposes;
- Use disaggregated student data for determinations and presorting to facilitate the narrowing of persistent achievement gaps;
- Make valid and reliable school and district accountability determinations annually;
- Produce reports that are transparent, timely, useful, and understandable by students, parents, teachers, administrators, and the general public;
- Provide differentiated systems of support to the lowest performing schools and districts including professional development targeted to their deficits;

PRIORITY GOALS

The goals for Wisconsin's school accountability system are evident in the focal areas determined by the Accountability Design Team. The system is designed to promote: student achievement, academic growth, equity of outcomes, and readiness for post-secondary success.

[33] Excerpt from "A Framework to Support Accountability Evaluation" (Landl, et al., 2016) for a framework and example demonstrating how a theory of action can be specified to reflect a state's accountability goals and inform system design and validation. A Case Study based on Wisconsin's theory of action is provided in Appendix A. https://www.nciea.org/sites/default/files/publications/A%20Framework%20to%20Support%20Accountability%20Evaluation.pdf

- Recognize the highest performing schools and districts, and disseminate their best practices to schools serving similar populations to help scale up high performance statewide;
- Have reasonable and realistic implementation goals that ensure the state, districts, and schools have the capacity to fully implement the accountability system and act on the results; and
- Remain open to feedback and findings about potential system improvements through implementation to ensure maximum effectiveness of the system.

Ultimately, the Design Team identified the four key areas of focus for the accountability system:

1. Student achievement
2. Student growth
3. Closing gaps
4. On-track to graduation and postsecondary readiness

These came to be known as the report card's Priority Areas and reflect the systems' goals.

The group also felt that the accountability system should engage multiple measures that reflect a value placed on varied postsecondary outcomes. They wanted the system to focus not only on English language arts and mathematics assessment performance, but also science and social students and 21st century skills as appropriate data become available. It also stated that college and career readiness should be measured differently for elementary and middle schools than high schools.

The principles and recommendations laid forth by the Accountability Design team provided an initial framework for more detailed design of the accountability measures and reports. The Design Team discussions also informed the high level Theory of Action (TOA) for how the system was intended to promote the identified goals. The TOA posits that designing and producing school and district report cards that treat every school as fairly as possible, are valid, reliable, and transparent, will inform local improvement planning and highlight actionable areas of performance that reflect key values in the educational system. Moreover, appropriate supports and interventions that are based upon a continuum of levels of support, directly linked and adjusted according to accountability ratings, will help support the intended goals.

HIGH LEVEL THEORY OF ACTION

Report cards inform local improvement planning and highlight actionable areas of performance that reflect key values in the education system. This influences a continuum of support initiatives linked to accountability outcomes.

Measures and Design Features[34]

The school and district report cards include the four priority areas identified by the Accountability Design Team, as well as three Student Engagement Indicators, which reflect individual measures of importance that, to some extent, reflect on the validity of the priority area measures. Supplemental data play a key part in the report cards, in an effort to encourage those viewing the report card to "drill in," ask further questions, and ultimately attend to other, related data sources not captured in the report cards, such as local data. For members of the public that view the report cards, the data therein are intended to provide an understanding of overall performance in key areas.

[34] For additional information about Wisconsin's accountability system see:
- Report Card Technical Guide: http://dpi.wi.gov/sites/default/files/imce/accountability/pdf/School%20Report%20Card%20Technical%20Guide%202014. pdf
- Report Card Interpretive Guide: http://dpi.wi.gov/sites/default/files/imce/accountability/pdf/Interpretive%20Guide%202014.pdf
- Additional Resources: http://dpi.wi.gov/accountability/resources

Priority Areas

The priority areas were listed in the previous section and serve to clarify the intended outcomes. To track these outcomes, the following program measures are produced.

Student Achievement

Purpose: to show how the students' level of knowledge and skills at a specific district or school compares against state academic standards.

Measure(s): a composite English language arts (ELA) and mathematics performance of all students. The score is based on how students are distributed across the four WSAS performance levels, and it takes three years' worth of test data into account.

Supplemental data: performance by subgroup.

Details:

- The method for calculating each content area score is based on assigning points to each of the district or school's students in each of the three measured years according to the student's performance level in that year. A student is assigned no points for being at the Minimal Performance level, one-half point for being at the Basic level, one full point for Proficient, and one-and-a-half points for Advanced.
- ELA and math are equally weighted, comprising 50- points each of the 100-point priority area score. For each year, students' scores are pooled to produce a district or school average. From those yearly averages, a three-year average is calculated. The averaging processes used in the calculations give greater weight to more recent years' data and also reduce the effect of year-to-year enrollment variability on aggregated test data. The score for each content area reflects this three-year average.

PROGRAM MEASURES

To track the prioritized outcomes program measures for Wisconsin's system include:

- Weighted index of ELA and mathematics performance on state tests
- Academic growth based on achieving target Student Growth Percentile (SGP) values
- Gap closure for identified groups based on improvement in test scores and/or graduation rate that exceeds comparison group
- Graduation rate
- Attendance rate (selected schools)
- Other academic measures associated with readiness or on-track to readiness
- Test participation
- Drop-out rates

Student Growth

Purpose: to give schools and districts a single measure that summarizes how rapidly their students are gaining knowledge and skills from year to year. In contrast to Student Achievement, which is based on the levels of performance students have attained, Student Growth focuses on the *pace of improvement* in students' performance. Student Growth rewards schools and districts for helping students reach higher performance levels, regardless of a student's starting point.

Measure(s): the heart of this measure is a point system that rewards schools and districts for students' progress toward higher performance levels from wherever they started. The point system also penalizes for student performance that regresses below the proficient level. The measure also rewards schools and districts that are already doing well by maintaining the high performance of their students, thus recognizing that very high performing students may not be able to grow as much or as quickly as other students as demonstrated by results on the state assessment.

Supplemental data: growth by subgroup

Details:

- Unlike Student Achievement, the Student Growth Priority Area only reflects the progress of students taking the general education assessment because the scoring scale of the alternate has not permitted growth calculations.
- This score reflects the degree to which students are on target to move from their starting scale scores to higher (or lower) performance levels within a three year period, based on their Student Growth Percentile (SGP). Students' starting scale scores are taken from the year prior to the current year of test results and an individual SGP is calculated for each student. Points are assigned to students based on a comparison of their SGPs with target SGPs for higher or lower performance levels.
- Target SGPs represent the pace of growth a student would have to exhibit to be considered on target to reach a different performance level within the three-year measurement period. Usually, this reflects growth to a higher level within three years or decline below Proficient within one year. Target SGPs are calculated using data about the growth track records of preceding groups of students who shared a similar achievement history with the student in question.
- Separate scores are calculated for ELA and mathematics and then combined.

Closing Gaps

Purpose: The purpose of this Priority Area is to provide a measure in sync with the statewide goal of having all students improve while closing the achievement gaps that separate different groups of Wisconsin students. It reflects the fact that achievement and graduation gaps are a statewide problem, not something limited to a small number of individual schools. The Closing Gaps Priority Area is designed to reward schools and districts that help close these statewide achievement gaps.

Measures: For this Priority Area, target racial/ethnic groups (Black students, Hispanic students, Asian/Pacific Islander students, and American Indian students) within a district or school are compared to White students statewide, their complementary comparison group. Students with disabilities, English language learners, and low-income students within a district or school are also compared to their complementary, statewide comparison group. A composite group (aka 'supergroup') is formed to meet the group size requirement (N=20) by combining at least two of the three above target groups when they do not meet the size requirement on their own. The Report Cards give credit for raising test scores and graduation rates for target groups faster than their statewide comparison groups. As a result, this measure encourages performance that lifts the performance of traditionally lagging groups, contributing to closing the statewide performance gaps.

Details:

- There are two components in the Closing Gaps priority area: Achievement Gaps and Graduation Gaps. If both apply for the district or school, each component score counts for half of this Priority Area score. If only one applies, the score for that component is the score for this Priority Area.

- The calculations for each of the two components follow the same basic procedure: Change in performance over the most recent three to five years is measured for each target group in the district or school and compared to the change in performance of the statewide comparison group. Change in performance is determined by finding the overall trend in performance, while also taking into account yearly fluctuations in enrollment. A minimum of three years of performance data are considered, and up to five years are included when available. The difference between the group change and the statewide change is then calculated, producing the closing gaps indicator for each target group. The indicators from all target groups are then combined to produce an overall Closing Gaps score for that component.

- For the Closing Achievement Gaps component, performance means achievement in reading and mathematics, measured in the same way as for the Student Achievement Priority Area, except that students are pooled by group and not the entire district or school.

- For the Closing Graduation Gaps component, performance is measured with the four-year cohort graduation rate. Because Wisconsin began reporting cohort graduation rates in 2009-10, graduation data prior to 2009-10 are not available.

On-Track to Graduation and Postsecondary Readiness

Purpose: The purpose of this Priority Area is to give schools and districts an indication of how successfully students are achieving educational milestones that predict postsecondary readiness. Measures: This Priority Area has two components. The first component is either a graduation rate—for schools that graduate students (i.e. high schools)—or an attendance rate for schools with no 12th grade. For most districts, both attendance and graduation scores will be included. The second component is a set of measures that include third grade reading achievement, eighth grade mathematics achievement, and ACT participation and performance, as applicable to the school. The scores for these two components are added to produce the Priority Area score.

Supplemental Data: subgroup performance

Details:

- Calculations for this Priority Area are based on an "all students" group.

- Component 1: Graduation Rate or Attendance Rate.
  - For schools that graduate students, a graduation rate is used as the indicator. For other schools, an attendance rate is used. Districts use both the graduation rate and attendance rate. Graduation rates and Attendance rates are highly correlated and have virtually identical distributions.
  - The graduation rate is the average of the four-year and six-year cohort graduation rates.
  - The attendance rate is the number of days of student attendance divided by the total possible number of days of attendance. The attendance rates of the "all students" group and the student group with the lowest attendance rate are averaged to produce the report card attendance rate.
  - The performance on this component accounts for a fixed 20 percent

- Component 2: Other On-Track Measures.
  - A school and district may have up to three 'Other On-Track' measures contributing to the score for this component: a third grade reading achievement indicator, an eighth

grade mathematics achievement indicator, and a combined ACT participation and ACT performance indicator.
- Third grade reading achievement and eighth grade mathematics achievement are measured in the same way as in the Student Achievement Priority Area.
- The ACT Participation and Performance score is the average of five rates for twelfth-graders: the ACT participation rate and the college readiness rates for all four ACT subject areas.
- A composite score for this component accounts for a fixed five percent of the weighted average priority areas score, regardless of, overall, how many Priority Areas apply to the school.

Student Engagement Indicators

Three performance indicators measuring student engagement are vital indications of school and district effectiveness. Low test participation reduces the validity of any comparisons and conclusions that can be drawn from assessment data. High absenteeism and dropout rates point to other educational shortcomings. Because of the significance of these three indicators, districts and schools that fail to meet statewide goals marking acceptable performance will receive fixed deductions from the weighted average priority areas score.

Approaches to Ongoing Monitoring and Evaluation

We conclude this case study with some suggestions for potential evidence that may be collected to evaluate selected elements of the Wisconsin school accountability system. The sources of evidence shown in the following table are not intended to be comprehensive. Rather, this is intended to illustrate elements of the evaluation framework.

COMPONENT	POTENTIAL SOURCES OF EVIDENCE
Expected Impact	• Trends in student performance on state tests overall and by subgroup • Annual changes in magnitude of achievement gaps for academic measures and graduation rate • Percent of students enrolling in credit-bearing college courses • Increased student engagement as measured by attendance and absenteeism • Use of data to inform local decisions increases • Local decisions related to behavioral supports, curriculum, or staffing (for example) are adjusted based on, in part, performance as measured by the accountability system
Program Measures	• Indicators are stable (e.g. year-to-year growth outcomes are positively correlated) • Outcomes are not correlated with unrelated factors (e.g. correlation between growth and prior-year status is low)
Fidelity of Implementation	• Focus groups reveal that reports are clear and helpful Surveys show that educators use results in planning and improvement efforts

We stress that ultimately the value of an accountability system is tied to the extent to which it both incentivizes the desired behaviors and produces information that stakeholders can and do use to improve student achievement. In the best case, these claims are made clear in the theory of action and are put to the test in the evaluation process. For example, if the theory of action holds that high

school educators will provide instruction on more challenging academic content to prepare students for college, evidence to support this claim might include: review of syllabi or focus groups with teachers. As another example, if the theory of action in Wisconsin holds that support strategies triggered by the system, such as providing supplemental educational services, will be effective, a study designed to compare similarly performing students who do and do not receive the services will help the state determine if these strategies are producing the desired result.

Ideally, evidence is collected, evaluated, and documented each year and the model will be refined as needed. In this manner, states improve the likelihood that the accountability system works to support the intended goals.

APPENDIX B:
ABBREVIATED OPERATIONAL BEST PRACTICES FOR ACCOUNTABILITY

Pre-Chapter: State Considerations for Accountability Systems.

1. The design of a state's accountability system will depend on how it intends to address the accountability requirements defined in federal and state law and, at the same time, consider the role of local accountability (e.g., regional or district requirements, accreditation, district accountability).
2. The goals, purpose, and uses of the accountability system will guide its design, development, and implementation. Federal and state requirements may influence timelines for implementation and who is involved in the process.
3. A state's vision for school support must be considered throughout the design of the accountability system
4. Equity must be a core principle driving the design of the accountability system.
5. Stakeholders are necessary to validate assumptions regarding the accuracy, relevance, and usefulness of an accountability system and its results.
6. Business rules should clearly and accurately operationalize the design of the accountability system in a manner that reflects the State's priorities.
7. Quality control and operations are critical to taking the accountability system from design to implementation.
8. Accountability reports rely on high quality designs, accurate data, and functional access. Without operational mechanisms that support translating data to information, accountability systems may not communicate performance or be accessible.
9. Accountability reports are the primary access point for the public and educators. A thoughtfully designed system will be less effective if results are not reported in an effective manner.
10. Evaluating accountability systems requires examining the individual decisions and assumptions underlying the system design. This can help clarify where and why decisions, dependencies, or information handoffs within the system may break down or function differently than intended.
11. Ongoing, critical evaluation of the accountability system design is required to identify where/if changes to the system may be necessary.

Chapter 1: Establishing an Accountability System's Theory of Action

1.2 Clarify the requirements underlying federal and state accountability systems as a means of determining how the systems can and should interact.

1.3 Clearly describe the primary goals of the accountability system. The system will need to be designed in a manner that both meets the state's goal and complies with state/federal mandates.

1.4 Articulate the purpose of the accountability system and the ways in which results are intended to be used.

1.5 Specify the elements of the theory of action, indicating which can be measured, monitored, and evaluated.

Chapter 2: Identifying Stakeholders and their Involvement

2.1 States should establish a stakeholder inclusion plan that identifies stakeholder groups and establishes their roles (e.g., advisory versus regulatory) and degree of involvement (e.g., informed, consulted, or included) throughout the stages of accountability.

2.2 Select the members of each stakeholder group in light of its intended role and to reflect the composition, political spectrum, and diversity of the state population. Each citizen should be considered a consumer of both the educational system and the information it provides, which will influence how deeply and frequently stakeholder groups are involved, consulted, or informed.

2.3 The state should prepare the message, structure, and forum to share its initial SEA vision and priorities with each identified stakeholder group. While the detail, complexity, and timing may differ, the core of the message should remain the same.

2.4 Engage stakeholders as intended to refine/extend the vision, priorities, intended outcomes and theory of action consistent with the roles and expectations defined in 2.1 and 2.2.

2.5 In addition to stakeholder groups the state should identify technical advisors to help substantiate claims about the accountability system's validity, reliability, and fairness and contribute to the development of an overall validity argument.

Chapter 3: Accountability Roles, Responsibilities, and Program Management

3.1. State Education Agencies manage a substantial number of educational programs that address both compliance and improvement initiatives. SEAs should have a clear definition of the scope of their accountability system and specify how divisions and offices are organized in relation to the system.

3.2. Identify specific roles and responsibilities attached to project milestones to enable project monitoring and help identify where project slippage may have occurred.

3.3. There may be aspects of the accountability system that are beyond the capacity of the SEA to develop, manage, or implement. In these cases, it will be important to identify whether external service providers are needed and to define their roles, responsibilities, and contractual obligations.

Chapter 4: Selecting and Integrating Measures for Accountability

4.1. Identify how/if measures in the state accountability system are intended to interact with federal requirements for accountability systems. This may reflect distinct, blended, or fully integrated federal and state accountability system designs.

4.2. Identify state accountability requirements that are specified in law, regulation, and policy and how they impact measure selection.

4.3. Once federal and state requirements are identified, states should determine how school quality will be defined, measured, and communicated based on the SEA's theory of action. This will inform the selection and calculation of system measures and help SEA's

determine how those measures should be combined (e.g., through indices, composite scores, or decision rules) to reflect the relative importance of data and priorities defined within the TOA.

4.4. Establish a clear rationale for why measures are included in the accountability system.

4.5. Determine and document the defensibility of the measures included in the accountability system. This defensibility should be based on whether measures are technically sound (i.e., reliable, fair, and valid) and whether they function as intended.

4.6. Evaluate candidate measures to ensure they demonstrate sufficient internal technical quality for use.

4.7. Consider whether the SEA's theory of action requires the combination of measures into composites or a series of decision rules. If so, determine and document how measures in the accountability system will be combined to meet the accountability system's goals.

Chapter 5: Establishing Performance Standards for Accountability Systems

5.1. Like other performance standard setting processes, states should specify a design in advance of setting standards. This design should be vetted by key stakeholders, technical advisors, and provide the opportunity for states to check processes and decisions against the original design. The following reflect key priorities to consider as readers review 6.2-6.6.

5.2. States will need to decide what performance categories should be established through the development of Policy Definitions or Policy Descriptors.

5.3. States should develop school performance level descriptors (SPLDs). School performance level descriptors are more detailed descriptions of what it looks like for a school to achieve each performance level in the state system.

5.4. States will need to define imprecise terms like "high rates" or "meeting expectations."

5.5. Document the development and refinement of SPLDs.

5.6. States will need to identify a representative standard setting panel that can consider both the data and associated consequences of cuts. The state should convene a broad-based panel of leaders and stakeholders to evaluate information and make recommendations regarding performance expectations for the accountability system.

5.7. States should prepare appropriately for the actual standard setting event. In preparation for the standard setting event, states will need to generate multiple documents and resources that are needed to implement the standard-setting process.

5.8. When states conduct the standard setting event, they will need to document the event sufficiently. When convening the standard setting event, the state will need to 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 and unbiased.

5.9. Establish group recommendations.

5.10. Evaluate and document each step of the process.

5.11. Ensure that the standard setting design, process, and recommendations are incorporated into any state-specific approval or governance processes.

5.12. Upon making changes to the accountability system, conduct an internal review of the accountability system data and determine whether the change is sufficient to warrant a reset of the performance standards.

5.13. In certain cases, states will define performance standards normatively, through legislation, or based on some other policy-defined method.

Chapter 6: Articulating Operations and Quality Control in Accountability Systems

6.1. In order to support the soundness of decisions made with accountability results, procedures and timelines underlying the definition, collection and evaluation of input data must be fully specified.

6.2. States will need to ensure that their data systems are working and managed securely. The systems define the process through which data are translated from input to output. Where those data are housed, who manages those data, and how they are managed are critical to operations and system implementation.

6.3. States will need to ensure that their data systems are producing defensible output that yield sound data to its users. This will mean that the reported data are complete, accurate, and can support correct interpretations. As with the input and process stages, the output stages will require a clear definition of what, who, where, when, and why operations support defensible output.

6.4. States should specify their goals, practices, and materials associated with quality control for data, processing, and documentation. This underlying characteristic of commitment to quality is essential to all stages of operational implementation. Well-defined and executed quality control procedures are part and parcel to the annual design, administration, scoring, and reporting processes associated with input and outcome data for accountability systems.

Chapter 7: Reporting and Communicating Accountability Results

7.1. The scope and role of accountability systems vary significantly across states. States should establish a communications strategy that aligns with the state's theory of action and how the accountability system is intended to communicate school performance

7.2. Define the design considerations for accountability reporting and establish a clear hierarchy for current and future reporting plans and access points. It is important for states to distinguish between data that is required for accountability system reporting (e.g., proficiency rates) and supplemental data that will be supported by the state (e.g., participation or access to certain course options). SEAs must establish a plan and hierarchy that helps prioritize development efforts and contribute to establishing a data hierarchy that connects state required, state supported, and locally available information.

7.3. Based on the overall communications strategy, focus of reports, and comprehensiveness of supplemental resources, the number and complexity of work plans will vary. States should establish clear and trackable work plans that support the production of accountability reports and associated resources as intended.

7.4. While the development of accountability reports is an important step to making information accessible to the public, SEAs also try to make information interpretable and actionable for educators. It is important to consider the overall communication strategy

to define the scope of the work plan (see OBP 7.1). For any additional resources, identify key stakeholders that play an integral part in improving outcomes in the educational system and design supplemental reports that align with the SEA's communication strategy.

7.5. As an extension of developing and implementing a reporting and communications plan, consider examining the degree to which users of the accountability system and its data understand and use data in a meaningful way.

Chapter 8: Monitoring and Evaluating Accountability Implementation

8.1. Review the theory of action to identify the intended purpose and use of the accountability system. It will be important to identify what the SEA considers to be the accountability system, which will include its scope, how it connects to support structures, and how it supports continuous improvement. This will enable states to create a clear logic model that articulates how wide or narrow the accountability system should be (see Chapter 1: Theory of Action).

8.2. Once the scope of the accountability system is identified, the theory of action and accountability logic model can be used to identify the major components of the accountability system. This will enable SEA to identify the claims that they are making through the accountability system.

8.3. A clear understanding of the accountability system's scope, key activities, and associated claims will help define the scope the SEA's research and evaluation agenda. It will be important to focus research and evaluation efforts on collecting evidence that supports design, development, and implementation activities that are of the highest priority to promote confidence in the accountably system's operations.

8.4. The claims associated with each activity in the accountability system will require different types of evidence. It will be important to identify what evidence is best associated with each claim and how to collect the data or information that meets evidentiary needs. The evidence collected will vary by the analyses or methods applied. These pieces of evidence will be used to test and confirm assumptions for activities of the accountability system.

8.5. Document the activities associated with the design, development, and implementation of the accountability system. Extend this documentation to include the progress of collecting evidence and how evidence supports claims across the accountability system.

Chapter 9: Engaging in Accountability System Change Management

9.1. Document rationale for a proposed change to the accountability system.

9.2. Determine and define a change management process.

9.3. Model candidate changes in the system

9.4. Evaluate the implications of the proposed modifications (independently and as part of the full system).

9.5. Develop and implement a communication plan for proposed and accepted changes.

9.6. Engage in and adjust any monitoring and evaluation activities as a result of the enacted changes.

National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

www.nciea.org