

# **Accountability Identification**

**is only the Beginning:**

**Monitoring and Evaluating Accountability Results and Implementation**

## THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

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

Accountability Identification is only the Beginning:  
Monitoring and Evaluating Accountability Results and Implementation

### COUNCIL OF CHIEF STATE SCHOOL OFFICERS

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

Juan D'Brot, Center for Assessment  
Leslie Keng, Center for Assessment  
Erika Landl, Center for Assessment

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



© 2018 by CCSSO. *Accountability Identification is only the Beginning: Monitoring and Evaluating Accountability Results and Implementation* is licensed under a Creative Commons Attribution 4.0 International License <https://creativecommons.org/licenses/by/4.0/>.

# CONTENTS

Accountability Identification is only the Beginning .....	2
Executive Summary.....	2
Background .....	4
Assumptions for Accountability Systems .....	4
Establishing Claims to Identify Monitoring and Evaluating Opportunities .....	6
Using a Framework to Monitor and Evaluate Accountability System Efforts .....	8
Claim 1: Indicators provide fair and accurate information that informs the accountability system in the manner intended.....	10
Claim 2: The state’s system of meaningful differentiation provides accurate, reliable, and valid information about the relative performance of schools in a manner that reflects the state’s values and priorities. ....	11
Claim 3: The state’s procedures for identification (Comprehensive Support and Improvement/Targeted Support and Improvement) are fair and identify those schools believed by the state as most in need of support.....	12
Claim 4: Exit criteria are attainable and accurately identify schools that are no longer in need of support.....	13
Claim 5: Identification/ labels and school profile information are signaling school quality in the manner intended to the public. ....	13
Leveraging Partners to Operationalize This Framework.....	15
Conclusion.....	17

# ACCOUNTABILITY IDENTIFICATION IS ONLY THE BEGINNING

## Monitoring and Evaluating Accountability Results and Implementation<sup>1</sup>

### EXECUTIVE SUMMARY

The passage of the Every Student Succeeds Act (ESSA) marked the beginning of a new development cycle for accountability systems. State leaders once again have an opportunity to redesign their accountability systems based on the provisions included in ESSA and to ensure that systems improve outcomes for all students. While accountability systems differ in their theory of action and design, state systems were designed following the Council for Chief State School Officer's (2011<sup>2</sup>) principles that guide the development and improvement of accountability systems.

As states begin implementing and monitoring their accountability systems created under ESSA requirements, the number of stress points across a system becomes more evident. Additionally, effective accountability implementation extends beyond identifying the right schools or obtaining approval for a system that can then be treated as "set it and forget it." Most states' experiences with accountability systems over the past few decades support the view that:

- Merely rating and/or identifying schools will not lead to the desired outcomes; rather, some active supports are required; and
- State, district, and school leaders are still learning what supports really work, especially at the needed scale and desired scope.

The correct identification of schools is a necessary but insufficient condition to build capacity and deliver support to local systems. Systems of accountability, support, and continuous improvement contain a series of feedback loops and information hand-offs that offer opportunities to collect evidence that systems are working as intended. By identifying activities and their relevant evidence throughout the design, development, and implementation of accountability systems, we can begin to develop validity arguments for our accountability and improvement systems. This paper references a framework<sup>3</sup> that can support a systematic examination of the design, development, and implementation stages of accountability identification. This framework can be applied to the activities associated with each stage as follows:

---

1 This is the third of a three-paper set of resources presented at CCSSO's *State Plan Implementation for ESSA* in April 2018. The other resources include <http://www.ccsso.org/resource-library/where-rubber-meets-road> and <http://www.ccsso.org/resource-library/establishing-performance-standards-school-accountability-systems>.

2 CCSSO (2011). *Principles and Processes for State leadership on Next-Generation Accountability Systems*.

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

### *Design Stage*

1. Refining the system's overall vision (e.g., policy priorities, educational system goals, the role of accountability),
2. Specifying indicators based on the system's intended signals (e.g., growth and achievement, college readiness vs. career readiness, engagement), and
3. Defining policy weights that represent SEA values and priorities (e.g., growth = achievement).

### *Development Stage*

1. Clarifying indicator measures and relationships among indicators through analysis (e.g., descriptive and inferential analyses, qualitative reviews of data and processes),
2. Identifying potential data gaps or capacity concerns through the use of simulations (e.g., projections, historical data examinations, mock accountability runs), and
3. Specifying performance expectations over time by setting defensible performance standards.

### *Implementation Stage*

1. Supporting the calculations and release of school designations,
2. Helping people access, use, and interpret accountability data, which in turn informs local inquiries and information use, and
3. Helping the SEA and LEAs deliver support to schools.

Within each of these stages and activities, SEAs can widen or narrow what they monitor to expand or limit system claims. Claims are statements or assertions about the accountability system and its impact. While claims will likely differ in granularity depending on the level of focus, they clarify the kinds of questions state leaders could be asking and the types of evidence they can consider collecting and evaluating.

By developing a set of claims associated with accountability and improvement systems, SEAs can begin developing a logic model that identifies the assumptions, questions, data considerations, and possible evaluation approaches. These claims can help states establish a **validity argument** for their accountability and improvement systems. However, the validity of the full system rests on the confidence states have in the validity of each activity, as well as each preceding step or activity along the way.

The remainder of this paper describes a framework to systematically help states evaluate identification decisions. It first presents example claims associated with each activity listed above (e.g., *Indicators* provide fair and accurate information that informs the accountability system in

the manner intended). For each claim, a series of guiding questions are then provided to help SEAs clarify the intended purpose, use, and process associated with each claim to determine assumptions. These assumptions are then used to help practitioners and designers identify sources of information, methods, or analyses that can be used to collect information to defend each claim. The framework is not intended to be prescriptive, but rather to provide examples of how states can apply this framework to begin establishing validity arguments for their accountability systems.

## BACKGROUND

The passage of the Every Student Succeeds Act (ESSA) marked the beginning of a new development cycle for accountability systems. State leaders had an opportunity to redesign their accountability systems based on the new provisions included in ESSA. Additionally, they have the authority and flexibility in designing and implementing their support structures and defining how they engage with their Local Education Agencies (LEAs), schools, and regional partners. State leaders are deeply committed to improving outcomes for all students and using their accountability and support systems to do so. As part of this effort, state leaders will be monitoring and evaluating these systems to ensure they have the desired impact and will continue to improve them as they learn more about what is working well and where adjustments may be needed.

## ASSUMPTIONS FOR ACCOUNTABILITY SYSTEMS

In 2011, states came together to establish a set of principles to guide the development and improvement of state accountability systems. According to these principles, state accountability systems will: drive school and district performance towards college- and career-readiness; distinguish performance to meaningfully target supports to the students most in need; provide timely, transparent data to spur action; and foster innovation and continuous improvement throughout the system. These principles focus on the entire cycle of the system, including accountability as a driver for school improvement and ongoing continuous improvement. While Congress has passed a new law, these principles continue to signal what states are trying to espouse in their accountability systems today. Readers are encouraged to review the principles, which can be found by following the link in the footnote<sup>4</sup>.

For this brief, we build on the concepts laid out in the principles and assume that accountability systems are part of a larger education system approach that seeks to spark and sustain continuous improvement efforts through the use of accountability *and* improvement systems. “Continuous improvement” involves at least three aspects: a) performance standards that require all groups to improve, b) the search for ways to make processes better, especially supports and interventions more directly linked to improved teaching and learning, and c) the stance that one’s assumptions and views can always be improved, and so must regularly and carefully be monitored. In this document, the focus is on the latter two aspects of continuous improvement.

---

4 CCSSO (2011). *Principles and Processes for State leadership on Next-Generation Accountability Systems*.

As states begin implementing and monitoring their accountability systems created under ESSA requirements, the number of stress points across a system becomes more evident. Additionally, strong and effective accountability implementation extends beyond identifying the right schools or obtaining approval for a system that can then be treated as “set it and forget it.” Most states’ experiences with accountability systems over the past two decades or more support the view that:

- Merely rating and/or identifying schools will not lead to the desired outcomes; rather, some active supports are required; and
- State, district, and school leaders are still learning what supports really work, especially at the needed scale and desired scope.

Thus, with the new law, states are moving to accountability systems that are integrated with support and intervention systems, and they will be evaluating these systems to identify what is working and what needs to be improved. Implementation includes understanding the efficacy of evidence-based improvement practices and determining whether the implementation structures/ approaches support (or at least do not hinder) information hand-offs and capacity building efforts. However, the implementation of identification systems must first be examined to help us understand the quality of support and delivery. The monitoring and evaluation of these identification systems is a key focus of this paper.

The correct identification of schools is a necessary prerequisite to support local systems but is insufficient to actually drive capacity building and the delivery of support. While not always explicitly laid out in a state’s ESSA plan, state leaders have been working to develop or strengthen state-based support systems for schools and districts. Readers are encouraged to review two resources that (1) describe the design and revision of state identification systems<sup>5</sup> and (2) offer suggestions for how to think about evaluating state identification and support systems<sup>6</sup>.

These complex systems of accountability, support, and continuous improvement contain a series of feedback loops and information hand-offs. These feedback loops offer opportunities for self-assessment, self-evaluation, and progress monitoring for both the system processes and the intended outcomes expected to happen in districts and schools. Further, it is useful to identify a common set of steps or processes that can be expanded or contracted based on state-specific implementations to better match context. The figure below offers a visual representation of how the larger feedback loop might exist across the design, development, implementation, and monitoring of accountability and improvement systems.

---

5 Lyons, D’Brot, & Landl (2017). [State Systems of Identification and Support under ESSA: A Focus on Designing and Revising Systems of School Identification](#). CCSSO: Washington, D.C.

6 D’Brot, Lyons, & Landl (2017). [State Systems of Identification and Support under ESSA: Evaluating Identification Methods and Results in an Accountability System](#). CCSSO: Washington, D.C.



Figure 1. The Accountability and Improvement Cycle

As presented in the figure above, each activity within the design, development, and implementation of accountability and improvement systems would also contain its own feedback loop. The next section describes a framework that can help states identify opportunities for monitoring and evaluation of their accountability and improvement systems that use the activities identified in the figure above.

## ESTABLISHING CLAIMS TO IDENTIFY MONITORING AND EVALUATING OPPORTUNITIES

The development of educational programs and systems begins with specifying a theory of action. In the case of accountability, theories of action articulate how the accountability system is intended to function in order to bring about the state’s desired outcomes. As these theories are tested through system implementation, they should be revisited to confirm that the underlying assumptions and processes hold. However, specifying the outcomes that stem from these assumptions and processes is critical to utilizing the framework. Possible outcomes or policy priorities attached to accountability systems might include the following:



- Improve the academic performance of the lowest performing students to proficiency;
- Encourage the academic improvement of ALL students including those already proficient; or
- Broaden the range of skills that students acquire to more fully account for the set of competencies necessary to be ready for college or career success.

By detailing the goals, purposes, activities, processes, and underlying assumptions of an accountability system, policymakers and practitioners will have a strong foundation for monitoring and evaluation.

At a high level, the flow of information across major components of a theory of action (e.g., reporting, identification, consequence and reward provision, monitoring) can be examined and confirmed. At a more detailed level, states can zoom in to the major stages (and activities) to determine whether assumptions hold true. Each of these connections and hand-off points can then be used to define the claims being made at the edge of each hand-off. These series of steps can be separated out into processes that help focus what claims are made where, as presented in the figure below.

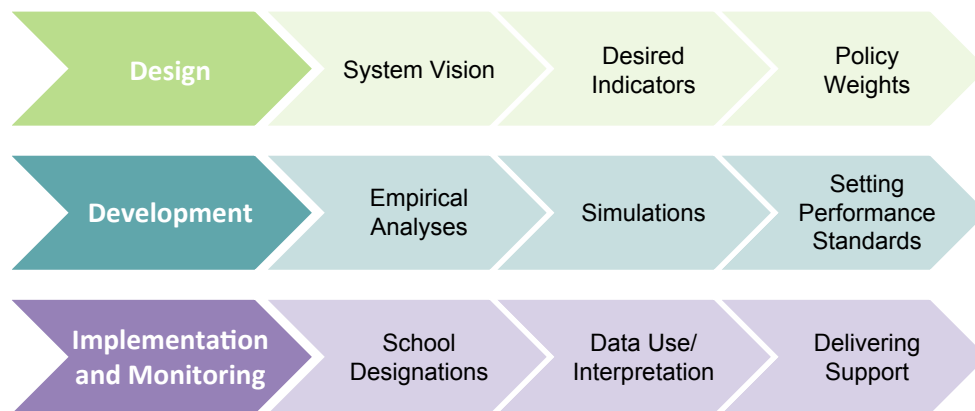


Figure 2. Stages and activities to specify accountability and improvement system claims.

As shown in Figure 2, one can expand or contract the number of stages (and activities) to widen or narrow a state’s monitoring view to expand or limit system claims. Claims are those statements or assertions we make about the system and its impact. While claims will likely differ in granularity depending on the level of system being discussed, claims clarify the kinds of questions state leaders could be asking and the types of evidence they may collect and evaluate.

Clearly defining a state’s system’s claims helps the state to focus on the types of evidence to consider and how it can collect the information that confirms assumptions. Should evidence contradict assumptions, one can then make adjustments within the system to better support continuous improvement efforts and support accountability outcomes. The next section describes how to extend claims using a framework that helps identify information and evidence to help

validate the accountability and improvement system’s goals. While claims can be made for all aspects of the system identified in Figure 1, the focus here is on the activities associated with identification and supporting school designations. Again, these steps are intended to help SEAs determine whether support structures are supporting districts and schools through the implementation of resources and practices, but are only the initial focus in the longer-term monitoring and evaluation effort.

## USING A FRAMEWORK<sup>7</sup> TO MONITOR AND EVALUATE ACCOUNTABILITY SYSTEM EFFORTS

By developing a set of claims associated with accountability and improvement systems, SEAs can begin developing a logic model of sorts that identifies the assumptions, questions, data considerations, and possible evaluation tactics. These claims can help states establish a **validity argument** for their accountability and improvement systems. However, the validity of the full system rests on the confidence states have in the validity of each preceding step or stage along the way. This is similar to the concept raised in the D’Brot, Lyons, and Landl (2017)<sup>8</sup>, which raises the idea of system reliability being dependent on component reliability. Specifically, a state must first verify the reliability of indicators/components before examining the interactions among indicators. That then allows a state to examine the reliability of overall school designations. This idea of how facets of reliability go from the part to the whole (i.e., bottom to top) is presented in the figure below.

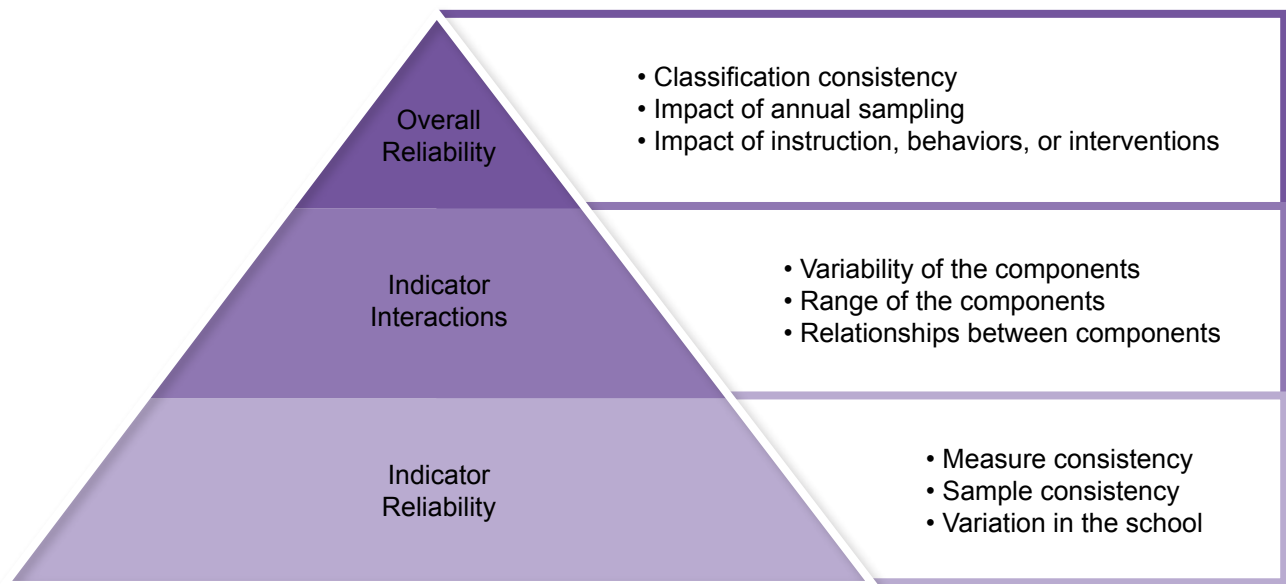


Figure 3. Factors Influencing Facets of Reliability (D’Brot, Lyons, Landl, 2017)

7 D’Brot (2018). *A framework to monitor and evaluate accountability system efforts*. Dover, NH: Center for Assessment.

8 D’Brot, Lyons, & Landl (2017). *State systems of identification and support under ESSA: Evaluating identification methods and results in an accountability system*. CCSSO: Washington, D.C.

The next few pages of this brief present a series of tables that specify example claims associated with the activities illustrated in Figure 2. For each claim, a series of assumptions are made that must be confirmed to support that step in the accountability and improvement system. These assumptions help to steer practitioners and designers toward sources of information, methods, or analyses that can be used to collect information in support of each claim.

It is important to note that the purpose of examining such information is to help SEAs develop a validity argument for their accountability and improvement systems. In the same way that SEAs collect evidence to support the validity of a measure (e.g., statewide summative assessments, attitudinal engagement data, observational rubrics of teaching), it is necessary to define the types of information (i.e., evidence) that allows states to defend the validity of systems. The *Standards for Educational and Psychological Testing* (2014)<sup>9</sup> note numerous ways of making validity arguments, but center on three major ideas:

1. Establish intended uses and interpretations;
2. Issues regarding samples and settings used in validation; and
3. Specific forms of validity evidence.

Thus, it is imperative that states define the uses and interpretations of the overall accountability and improvement system and take into account school and district demographics in their work. This may be done most efficiently by establishing the intended uses and interpretations (i.e., claims) of its composite components to build a comprehensive validity argument from the ground up. This allows states to compare evidence to their theory of action and inform any system revisions.

The tables presented in the remainder of this brief raise a series of questions associated with each of the following claims to narrow the information or approaches necessary to develop a validity argument. The claims include:

1. *Indicators* provide fair and accurate information that informs the accountability system in the manner intended;
2. The *system of meaningful differentiation* provides accurate, reliable, and valid information about the performance of schools in a manner that reflects the state’s values and priorities;
3. The state’s *school identification procedures* are fair and identify those schools believed by the state as most in need of support;
4. *Exit criteria* are attainable and accurately identify schools that are no longer in need of support; and
5. *Identification/ labels and school profile information* are signaling school quality in the manner intended.

These are expanded in further detail below.

---

<sup>9</sup> Standards for testing AERA, APA, & NCME, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

**Claim 1: Indicators provide fair and accurate information that informs the accountability system in the manner intended.**

The first claim highlights the need for states to have confidence in the activities that lead to school-level designations. Whether designations are summative (e.g., index-based) or descriptive (e.g., dashboard, not-identified) in nature, the activities to produce a school score or decision rules used to classify schools are contingent on a series of interactions or branching decisions that depend on how measures within each indicator operate and interact. The following guiding questions should be answered to help inform what sources of information or methodological approaches are most appropriate:

- What is the intended purpose of the indicator within the system?
- How will the indicator be used?
- What inference(s) is the indicator intended to support? Is it an observable count of something or is the measure intended to be a proxy for something else?
- How should the indicator relate to the other indicators in the system?
- What characteristics do you expect for the indicator to display to support its intended purpose?
- What data are required to inform calculations of the indicator?

In answering these questions, states can determine which of the following assumptions, sources, or analyses are most appropriate.

Claim to be Supported by Evidence: <i>Indicators provide fair and accurate information that informs the accountability system as intended.</i>		
Assumptions	Research and Evaluation Sources	Example Analyses or Data <sup>10</sup>
<ul style="list-style-type: none"> <li>• Indicator scores/ratings are reliable.</li> <li>• Data necessary to inform the indicator is accurate and complete.</li> <li>• The indicator relates to other indicators in the manner intended.</li> <li>• The indicator demonstrates characteristics necessary to ensure it is valid and fair.</li> <li>• Data and procedures support the intended interpretation of results (e.g., standard setting, norming).</li> </ul>	<ul style="list-style-type: none"> <li>• Clear tie between policy rationale and indicator</li> <li>• Specification of measures (i.e., data) leading to use for indicator</li> <li>• Quantitative and qualitative examination of reliability, robustness, objectivity of measure</li> <li>• Relationship of measures within and across indicators</li> <li>• Examining business rules and quality assurance procedures</li> </ul>	<ul style="list-style-type: none"> <li>• Descriptive examination of measures (e.g., shape, skew, distribution, range)</li> <li>• Reliability evidence where appropriate</li> <li>• Analyses to determine relationships among indicators (e.g., correlation, regression, factor analyses, path analyses)</li> <li>• Quality assurance checklists and business rules monitoring<sup>11</sup></li> </ul>

<sup>10</sup> Readers are encouraged to review Keng’s (2018) Operations and Quality Control brief for specific suggestions on this approach.

<sup>11</sup> Readers are encouraged to review Goldschmidt’s (2018) handbook on developing and monitoring English Language Proficiency indicators, found [here](#).

**Claim 2: The state’s system of meaningful differentiation provides accurate, reliable, and valid information about the relative performance of schools in a manner that reflects the state’s values and priorities.**

The second claim addresses the degree to which the system of meaningful differentiation is in alignment with the policy goals and priorities of the state. Like the first claim, regardless of whether designations are summative or descriptive (e.g., dash-board, not-identified), they are intended to send a signal to educators and the public and should provide enough separation among schools and their profiles. The following guiding questions can be answered to help inform what sources of information or methodological approaches are most appropriate to confirm how the larger sets of decisions inform the assignment of school decisions:

- Do aggregate data reflect the same differentiation as student level data?
- Are intentionally different signals (e.g., achievement and growth) complementing each other sufficiently within the system?
- Does variability exist for both the overall score/designation and the indicators separately?
- Have simulations been conducted on the system of meaningful differentiation? If so, is the volatility in scores or designations over time reasonable?

**Claim to be Supported by Evidence: The state’s system of meaningful differentiation provides accurate, reliable, and valid information about the relative performance of schools in a manner that reflects the state’s values and priorities.**

Assumptions	Research and Evaluation Sources	Example Analyses or Data
<ul style="list-style-type: none"> <li>• Aggregation procedures are applied accurately.</li> <li>• The intended emphasis of indicators is accurately reflected in the results of the state’s system of meaningful differentiation (e.g., effective vs. nominal weights of indicators).</li> <li>• Overall schools scores/ ratings/ classifications are reliable.</li> <li>• Ratings reflect separation in school profiles/data.</li> </ul>	<ul style="list-style-type: none"> <li>• Applying accountability business rules to prior year data to compare score/school profiles over time</li> <li>Quantitative examination of overall index scores (if applicable) or indicator based profiles</li> <li>• Comparative analyses of indicators to ensure sufficient variability exists to support differentiation (i.e., indicators are not overly constrained).</li> </ul>	<ul style="list-style-type: none"> <li>• Descriptive and inferential analyses of overall index scores or indicator based profiles (e.g., k-means clustering, discriminant analyses, distance comparisons)</li> <li>• Comparing range, shape, and skew of indicators to each other and over time</li> <li>• Quantifying empirical influence of indicators on overall score/ designation variability (e.g., commonality analyses, comparisons of measures of central tendency)</li> <li>• Identify max/min values of overall scores or indicator values and compare to means and standard deviations over time to assess volatility</li> </ul>

**Claim 3: The state’s procedures for identification (Comprehensive Support and Improvement/Targeted Support and Improvement) are fair and identify those schools believed by the state as most in need of support.**

The third claim addresses the federally required school designations and raises questions regarding procedures, intended outcomes, and rates of identification. While all SEAs were required to propose CSI, TSI, and Additional Targeted Support and Improvement (ATSI) school identification processes, states differed in their TSI and ATSI methodologies. While the considerations in the table below cannot address every state design, the following questions are intended to apply to the processes and capacity concerns that have emerged throughout the proposal and revision of CSI/TSI identification:

- Has the SEA developed school policy level descriptors (SPLDs) for CSI/TSI/ATSI designations?
- Does peer review feedback reflect initial design decisions? If not, have SPLDs been appropriately updated and do they still reflect the state’s policy priorities?
- Do identified school profiles reflect underperformance or lack of progress as specified by policy and design decisions?
- Is there clear separation in school data/profiles for CSI or TSI schools compared to others?
- Do identification rates of schools fall within the state’s capacity to support state-led or LEA-collaboration improvement efforts?

**Claim to be Supported by Evidence: The state’s procedures for identification (CSI/TSI) are fair and identify those schools believed by the state as most in need of support.**

Assumptions	Research and Evaluation Sources	Example Analyses or Data
<ul style="list-style-type: none"> <li>• The state’s procedures for identification do not over-identify certain types of schools (N-counts or high free reduced lunch populations).</li> <li>• Schools identified for CSI or TSI demonstrate a profile of performance consistent with that believed by the state as most in need of support.</li> <li>• Procedures used to identify TSI schools capture those schools with sub-groups in need of support.</li> <li>• ATSI and TSI identification approaches do not put a strain on state and local capacity to deliver support.</li> </ul>	<ul style="list-style-type: none"> <li>• Stakeholder feedback sessions, leadership vision, SEA priority articulation for CSI/TSI/ATSI schools</li> <li>• Cross-walk of intended identification design and revised identification business rules for CSI/TSI/ATSI schools</li> <li>• Quantitative examination of data over time</li> <li>• Comparative analyses by school category/designation and school profiles to compare indicators and overall performance</li> </ul>	<ul style="list-style-type: none"> <li>• Qualitative feedback from leadership and appropriate stakeholders through focus groups or interviews</li> <li>• Cluster analyses of school types using their indicators as predictors</li> <li>• Simulations applied to historical and projected data to confirm identification of schools and estimate volatility (e.g., bounce across thresholds)</li> <li>• Distance analyses of indicators and measures by school types (e.g., k-means or proximity analyses) to show differences among TSI, CSI, and other schools</li> </ul>

**Claim 4: Exit criteria are attainable and accurately identify schools that are no longer in need of support.**

The fourth claim addresses the federally required exit criteria associated with the CSI, TSI, and ATSI school designations (given the district-role for TSI exit criteria, the SEA will need to appropriately partner with districts to evaluate expectations statewide and systematically to offer support). In addition to the need to carefully design identification methods, exit criteria are equally important because they signal that schools are no longer in need of the same intensity of support in a focused or schoolwide manner (depending on the designation). The following questions are intended to highlight the considerations states can make as schools are monitored against state-defined requirements:

- Do design decisions for exit criteria reflect the system or Theory of Action’s policy priorities (e.g., equity, universal improvement, etc.)?
- Have simulations been conducted using historical data? Do data reflect possible (i.e., realistic vs. challenging vs. unreasonable) gains over time to exit?
- How do school profiles for CSI or TSI schools (i.e., indicator data within school designations) change over time compared to changes of non-identified schools?

**Claim to be Supported by Evidence: Exit criteria are attainable and accurately identify schools that are no longer in need of support.**

Assumptions	Research and Evaluation Sources	Example Analyses or Data
<ul style="list-style-type: none"> <li>• Exit criteria reflect the outcomes for improvement expected in the system (e.g., overall performance changes, subgroup specific improvements).</li> <li>• Historical improvements suggest exit criteria are attainable with focused improvement efforts.</li> <li>• As schools exit identification status, their profiles reflect separation from other newly identified schools.</li> </ul>	<ul style="list-style-type: none"> <li>• Alignment between policy decisions, design decisions, and school profiles over time</li> <li>• Quantitative examination of simulation data</li> <li>• Comparative analyses of changes in school profiles (overall, by indicator, by subgroup) between identified and non-identified schools</li> </ul>	<ul style="list-style-type: none"> <li>• Qualitative cross-walk between SEA policy priorities and exit criteria design decisions</li> <li>• Simulations applied to historical and projected data to confirm attainability of exit criteria</li> <li>• Change-over-time comparisons between identified and non-identified schools to determine if acceleration can be expected using historical data</li> </ul>

**Claim 5: Identification/ labels and school profile information are signaling school quality in the manner intended to the public.**

The fifth claim addresses the larger impact of school designations and how school performance is communicated to the public and educators. ESSA requires states to report school performance based on their systems of meaningful differentiation. However, states had a significant amount of flexibility in how they reported information. The following questions intend to raise considerations around coherence between performance standards for schools and how that information is communicated:

- Has the SEA conducted and documented a performance standards setting for school designations (while this question is related to the mechanics associated with the second claim, here it is more focused on the signals school designations communicate)?
- Have reports and school profile descriptions been evaluated by groups of stakeholders<sup>12</sup>?
- What are the “hit rates” of online reports? What feedback have local sites provided with regard to report use?
- Have coherent connections between outcome (i.e., lagging) data and leading indicators been established? Have these connections been communicated through resources or tools?
- Are low/no-stakes data provided in reports to support local monitoring efforts?

Is information reported at the school level reflective of expectations identified through performance standards setting?

**Claim to be Supported by Evidence: *Identification/labels and school profiles are signaling school quality in the manner intended.***

Assumptions	Research and Evaluation Sources	Example Analyses or Data
<ul style="list-style-type: none"> <li>• Performance standards for the system are defensible.</li> <li>• Overall school ratings support intended interpretations of school performance.</li> <li>• Reports are easily accessible and used by targeted groups of end-users.</li> <li>• Reports are clear and intuitive.</li> <li>• LEAs and schools understand how to interpret and use the results to inform improvement efforts.</li> </ul>	<ul style="list-style-type: none"> <li>• Representative and informed group engages in standard setting using SPLDs, school profiles, and impact data</li> <li>• Policy review includes a “reasonableness check” for performance standards over time</li> <li>• Quantitative examination of web page views and report access</li> <li>• Qualitative feedback for reporting structure and designed reports</li> <li>• Multi-stakeholder group brainstorming sessions to develop resources and tools</li> <li>• Identification of coherent low/no-stakes data by SEA and LEA experts</li> </ul>	<ul style="list-style-type: none"> <li>• Simulated performance standards applied to historical and projected data to determine change over time to impact data</li> <li>• Descriptive analyses of web and report access by location</li> <li>• Focus groups and feedback sessions for reporting tools and structure</li> <li>• Analysis of brainstorming sessions; focus groups review proposed resources and tools</li> <li>• Focus groups and feedback targeting supplemental information provided in reports not explicitly linked to accountability data or outcomes</li> </ul>

By replicating this approach across a series of claims for an accountability and improvement system, states can begin developing an action plan for whom to identify as partners, which claims to prioritize, and the data required to develop validity arguments for the larger system. Determining what claims to prioritize, what assumptions are most critical, who could successfully lead the examination of evidence, and who may be key collaborators or partners will be critical to supporting a comprehensive validity argument for the system.

<sup>12</sup> States are encouraged to read the <Performance Standards Brief> for more detail on setting standards.



Initially, states should at least address claims 1, 2, and 5 associated with indicators, the system of meaningful differentiation, and the system’s performance standards. The questions and assumptions subsumed within those three claims will have impacts on any subsequent decisions. The considerations raised in these claims can also help identify where there may be shortcoming or concerns associated with other claims. For example, if schools are not sufficiently differentiated, perhaps the school profiles weren’t used sufficiently during standard setting. Alternatively, there may be issues with how indicators interact as a whole, even though individual indicators appear to demonstrate sufficient variability. Even if states are unable to invest time into engaging in data collection efforts or in running analyses raised throughout the claims, answering some of the guiding questions can pay large dividends in developing a strong validity argument for the accountability and improvement system.

**LEVERAGING PARTNERS TO OPERATIONALIZE THIS FRAMEWORK**

Linking the theory of action to implementation efforts can help clarify claims along the entire system’s path. This can then help SEAs determine who the actors are at each point in the system. Some may be internal, some may be external, and some may require a collaborative partnership to examine bottom-up and top-down behaviors and decisions at varying levels of the system. Eventually, improvement efforts will require district and school partnerships to identify whether support structures and improvement activities are successful and useful.

It is likely that early development efforts (e.g., establishing a vision, policy priorities, and a system theory of action) begins with a heavy external focus to accurately reflect the input of a variety of stakeholders. As SEAs transition toward development and implementation, they become naturally more internally focused on ensuring design decisions are coherent with policy, supported by data collection efforts, and reflect data characteristics. However, once states begin implementing accountability systems, it is necessary to consider how external partners play an increasingly significant role in understanding how information is interpreted, data are used, and how behaviors may change throughout the improvement process. The figure below highlights the interrelated nature of these partnerships and how they may interact with the design, development, implementation, and monitoring of the system. Note that the internally focused portions of the process focus on the SEA role, whereas the externally focused portions, especially during implementation and monitoring stages may require partnerships with district, school, and other external groups.

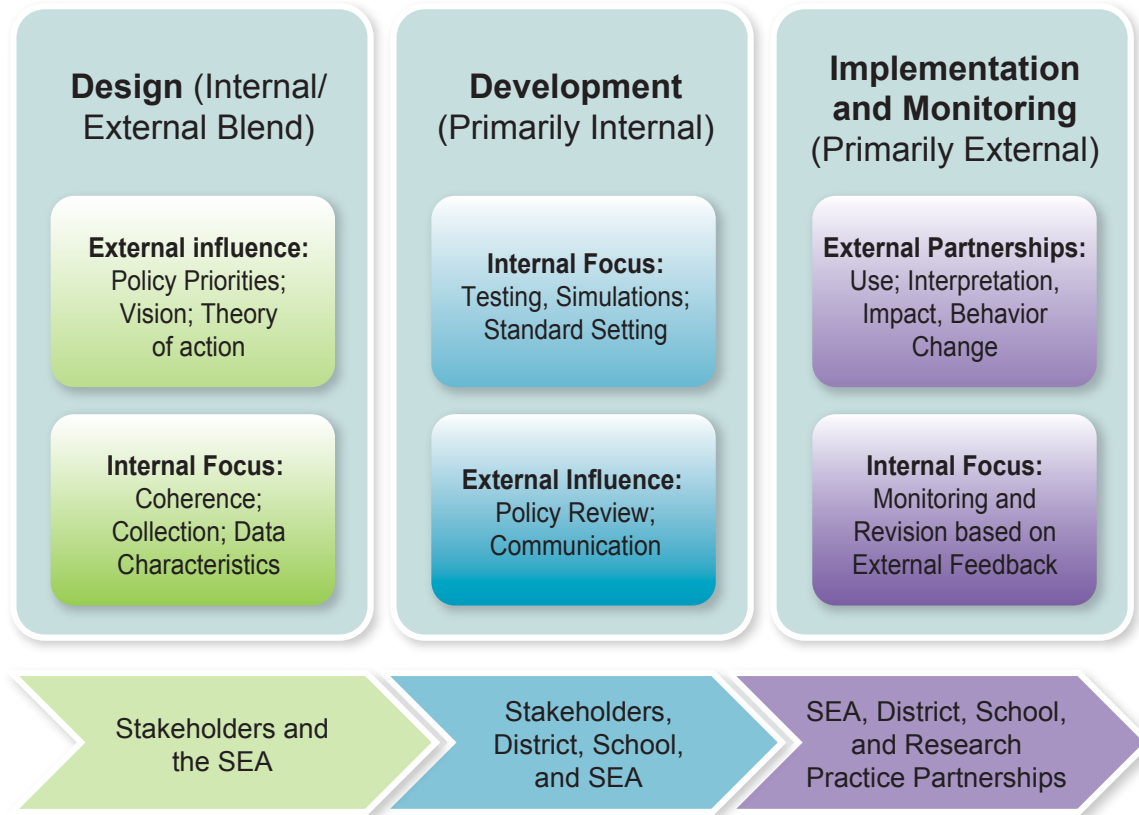


Figure 4. Internal and external influencers in monitoring and evaluation efforts.

Developing resources that target each of the stages of the process can lead to a wide scale benefit for varying accountability and improvement system designs. However, SEAs and their partners must identify whether they have direct, indirect, or upstream responsibility<sup>13</sup> for certain behaviors and actions. Because SEAs have limited resources, they may consider leveraging partnerships that identify external partners who can more closely monitor and evaluate the use of information and delivery of support efforts. Readers are encouraged to review the following resource<sup>14</sup>, which highlights strategies for how states can enable continuous, sustained improvement through Research-Practice Partnerships (RPPs). These partnerships include groups like

- ED funded Comprehensive Centers<sup>15</sup>,
- IES supported Regional Education Laboratories<sup>16</sup>,
- Academic Researchers (often based out of in-state research universities), and
- Research Consortia

<sup>13</sup> Upstream responsibility refers to the notion that SEAs are responsible for earlier parts of the improvement process (e.g., identification and delivery of support, policies that are not superfluous to improvement activities) that are critical to successful implementation at the school and classroom levels.

<sup>14</sup> Advancing School Improvement in SEAs through Research Practice Partnerships: <http://www.ccsso.org/resource-library/advancing-school-improvement-seas-through-research-practice-partnerships>

<sup>15</sup> <https://www2.ed.gov/about/contacts/gen/othersites/compcenters.html>

<sup>16</sup> <https://ies.ed.gov/ncee/edlabs/>

By identifying key partners and developing strong relationships, SEAs can focus on the work of maintaining strong systems (and revising them when necessary) while still leveraging key findings and lessons learned from program evaluation, research, and operational efforts conducted and examined through RPPs.

## CONCLUSION

Establishing a validity argument for accountability systems can be a daunting process. SEAs should prioritize collecting evidence associated with accountability in order to later support claims that are focused on improvement systems and support structures. This process may be more feasible if states identify the types of evidence needed based on a systematic examination of the assumptions attached to each step of the identification and support process. This can strengthen the validity of the system by better supporting the intended interpretations and uses of the accountability system.

In this paper, a rationale, a series of claims, and set of guiding questions to inform the types of evidence needed to build a validity argument for accountability were presented. Additionally, a high-level framework was proposed to recommend an approach to collecting relevant evidence. This paper then closed with considerations to identify potential external partners. This document may be helpful to education leaders and accountability system developers who need to identify evidence that supports accountability system claims that are in line with the state's theory of action.



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