RESTART & RECOVERY

CONSIDERING THE OUTLOOK FOR SCHOOL ACCOUNTABILITY:

STATE GUIDANCE FOR MAKING ANNUAL ACCOUNTABILITY DETERMINATIONS IN SCHOOL YEAR 2020-2021 AND BEYOND

We are grateful to our partners at the National Center for the Improvement of Educational Assessment for their help in developing this guide.

AUTHORS: Juan D'Brot, Erika Landl, Chris Domaleski, and Chris Brandt

CCSSO
One Massachusetts Avenue, NW
Suite 700 Washington, DC 20001

CCSSO
Council of Chief State School Officers™

INTRODUCTION

Disruptions due to the COVID-19 pandemic have impacted school accountability decisions. In spring 2020, the U.S. Department of Education granted waivers to all 50 states, Washington D.C., Puerto Rico, and the Bureau of Indian Education. These waivers focused on the following requirements outlined in the Elementary and Secondary Education Act (ESEA), as amended by the Every Student Succeeds Act (ESSA):

- assessment requirements in section 1111(b)(2) for school year (SY) 2019-2020;
- accountability and school identification requirements in sections 1111(c)(4) and 1111(d)(2)(C)-(D), based on data from SY 2019-2020.

The loss of student assessment results and other accountability data has required states to consider the impact on accountability systems for SY 2020-2021 and beyond. As states grapple with downstream effects, it is now evident that state systems will be affected for at least two years. As difficult as the loss of data is for states, it also presents an opportunity for them to reflect on their existing accountability systems and performance data interpretations.

The challenges facing states are numerous. Accountability systems often rely on a bundle of multi-year data. For example, almost every school accountability system includes estimates of academic growth; these estimates require at least one prior measure and, in some cases, multiple prior measures. Moreover, many systems rely on improvement measures, multi-year averaging, and lagged data. Many states also "bank" test scores for accountability purposes by using test results from previous years. Taken together, the COVID-19 disruptions will have a substantial impact on accountability operations, calculations, performance, comparability, and interpretations—both within and across years.

Given these challenges, the purpose of this paper is to provide:

1. guiding principles to inform states' approaches to restarting accountability;
2. a process to examine key decisions for accountability in SY 2020-2021; and
3. considerations for developing, implementing, and evaluating systems in SY 2020-2021 and beyond, in the era of COVID-19 disruptions.

Note that the focus of this paper is limited to accountability systems under ESEA and takes into account the constraints of the statute, waivers, and any necessary amendments to existing ESEA consolidated state plans. Nonetheless, what follows can be applied to a broader range of accountability, reporting, and support initiatives.

GUIDING PRINCIPLES TO INFORM STATES' APPROACHES TO RESTARTING ACCOUNTABILITY

As state leaders consider alternatives for school accountability in SY 2020-2021 and beyond, the first priority is to reflect on the role of school accountability, in order to promote improved outcomes. Broadly viewed, school accountability is a system that:

- signals which outcomes are valued;
- provides information about school performance with respect to those outcomes; and
- prescribes a system of supports and interventions based on performance (Domaleski et al., 2018).

Accountability systems are not a prescription for improving schools, but they can play an important role in an overall plan to support student success. Therefore, any decision to maintain, suspend, or modify the school accountability system will affect the state's school support initiatives.

With this in mind, this paper offers guiding principles to inform the tough choices to be made ahead. To be sure, there is not one best path forward; accountability decisions should be grounded in an understanding of the state's priorities for school improvement. In the following section, this paper shall provide considerations for clarifying the potential solutions.

REEXAMINE THE ACCOUNTABILITY THEORY OF ACTION IN LIGHT OF STATE PRIORITIES

The theory of action that an accountability system embodies should be revisited periodically (e.g., D'Brot, Keng, & Landl, 2018). As states do so, system designers and practitioners should ensure that the accountability system still aligns with state priorities. For example, priority outcomes might include:

- bringing the lowest-performing students up to proficiency;
- encouraging the academic improvement of all students, including those already proficient; and
- broadening the range of skills students acquire to ensure college or career success.

If a state's priorities have shifted, it will be necessary to determine the extent to which the accountability system's design, processes, and procedures align with those shifting priorities. Depending on the amount of shift, it may be necessary to amend the ESEA consolidated state plans to bring the system back into alignment with the state's priorities. If the state's priorities have not shifted, the state can examine the activities, processes, and procedures for each indicator, as well as the system overall, to ensure that the intended interpretations will hold in light of data loss or other problems associated with school closures in SY 2019-2020. Whether states align their systems with existing or revised priorities, leaders will need to consider tradeoffs and implications, especially regarding identification and performance expectations.

CONSIDER TYPE I & TYPE II ERRORS

Any path forward has thorny tradeoffs. One way to think of these tradeoffs is the consideration of Type I and Type II errors in accountability-related classifications. For example, consider the requirement to identify the lowest-performing 5 percent of schools in the state for Comprehensive Support and Improvement (CSI). The implicit theory of action is that the state has limited school-improvement resources, and a substantial portion of these resources should go to the schools most in need. Imagine that one could know the "true" condition regarding whether a school is among the lowest performing—for example, the state could credibly continue using their legacy[1] accountability model with no reservations; could use it to identify low-performing schools; and could know that their flags for identification accurately reflect the state's evaluation process. In that case, one could evaluate actual classifications in SY 2020-2021 with respect to that true condition (see Table 1).

TABLE 1 ILLUSTRATION OF TYPE I & TYPE II CLASSIFICATION ERRORS FOR CSI

	The school truly is among the lowest performing in the state	The school truly is **not** among the lowest performing in the state
THE SCHOOL IS CLASSIFIED AS CSI	Correct Decision	Type I Error – False Flag
THE SCHOOL IS NOT CLASSIFIED AS CSI	Type II Error – Failed to Flag	Correct Decision

Note that in Table 1, Type I schools are falsely flagged: they are identified for support but are not among the state's lowest-performing schools. On the other hand, Type II schools were not flagged for CSI but should have been. Each error has a cost: Are resources directed to some schools unnecessarily (Type I)? Did the state fail to support the schools most in need (Type II)? Understanding these tradeoffs can help with the tough choices a state will face. For example, if Type I errors are deemed more costly, the state may be more conservative about identifying new schools for support. However, if Type II errors are more of a concern, the state may privilege alternatives that leverage all available information to identify a wider range of schools for support.

LEVERAGE "BIG-A" & "little-a" SOLUTIONS

While most of the attention may go to the state's school-accountability system that fulfills ESEA requirements, the federal accountability system is likely only a small part of the state's overall plan to support school improvement. School improvement initiatives typically rely on a range of information (inputs and outcomes) to determine actions and evaluate outcomes. We use the shorthand "Big-A" to refer to accountability components directly tied to ESEA school classifications and "little-a" for elements outside ESEA (e.g., improvement processes, low-stakes indicators, local data elements, etc.). The role of "little-a" accountability solutions can be amplified to compensate for disruptions to "Big-A" systems.

What does this mean, in practice? Listed below are several examples of initiatives that may help states focus on school improvement in the midst of COVID-19 disruptions.

- Identify components[2] for reporting (internal or external), but do not use them to inform classifications. There may be data elements or indicators[3] from the state's legacy accountability model which are not appropriate in SY 2020-2021. Consider reporting these indicators—but withhold them from the model for determining high-stakes classifications.
- Work with districts and schools to identify new data elements that can inform school improvement. Districts are on the front line, serving the needs of schools and students. The state can help districts use information, such as interim assessments and survey results, to help inform school improvement initiatives without influencing outcomes in the state's ESEA model.
- Share resources and promising practices to enhance school-improvement efforts. Again, accountability is about improvement, which may include helping schools and districts implement initiatives beyond ESEA-specific expectations. For example, the state education agency might support the curation of a bank of exemplary curricular resources, conduct training on best practices for distributed learning, or offer resources to promote assessment literacy.

[1] A legacy accountability system is defined as the accountability system that was in place during school year 2019-2020.

[2] A component is defined as a generic term that refers to the activities or programs associated with an accountability system (D'Brot, in press).

[3] Indicators refer to the required components of an ESEA accountability system, which include academic, other academic, English Learner progress, graduation, & school quality & student success indicators.

CONSIDER RESTARTING ACCOUNTABILITY IN STAGES

By taking advantage of "little-a" accountability initiatives, states can consider phasing in aspects of their existing ESEA accountability system. These stages, which are addressed more fully in the Implementation section of this paper, range from implementing a complete legacy or revised system to implementing a transitional system that will not be complete until SY 2021-2022.

State leaders will need to work with the U.S. Department of Education to determine the impact on their approved ESEA plans and what kinds of short or longer-term changes may be necessary. The U.S. Department of Education is expected to work closely with states to make adjustments to their plans through a streamlined process. If any components of the legacy or revised systems fall short of the state's designed accountability model (e.g., missing data elements, incomplete processes or procedures, threats to data interpretations), it may be valuable for states to leverage reporting or school-improvement initiatives to supplement any missing data in SY 2020-2021 and beyond. For example, if certain changes to how academic growth is calculated could lead to a decision to exclude it from accountability, that would not prohibit reporting academic growth if this data is used in strategic ways to inform improvement, especially during a transitional period with the state's "Big-A" accountability model.

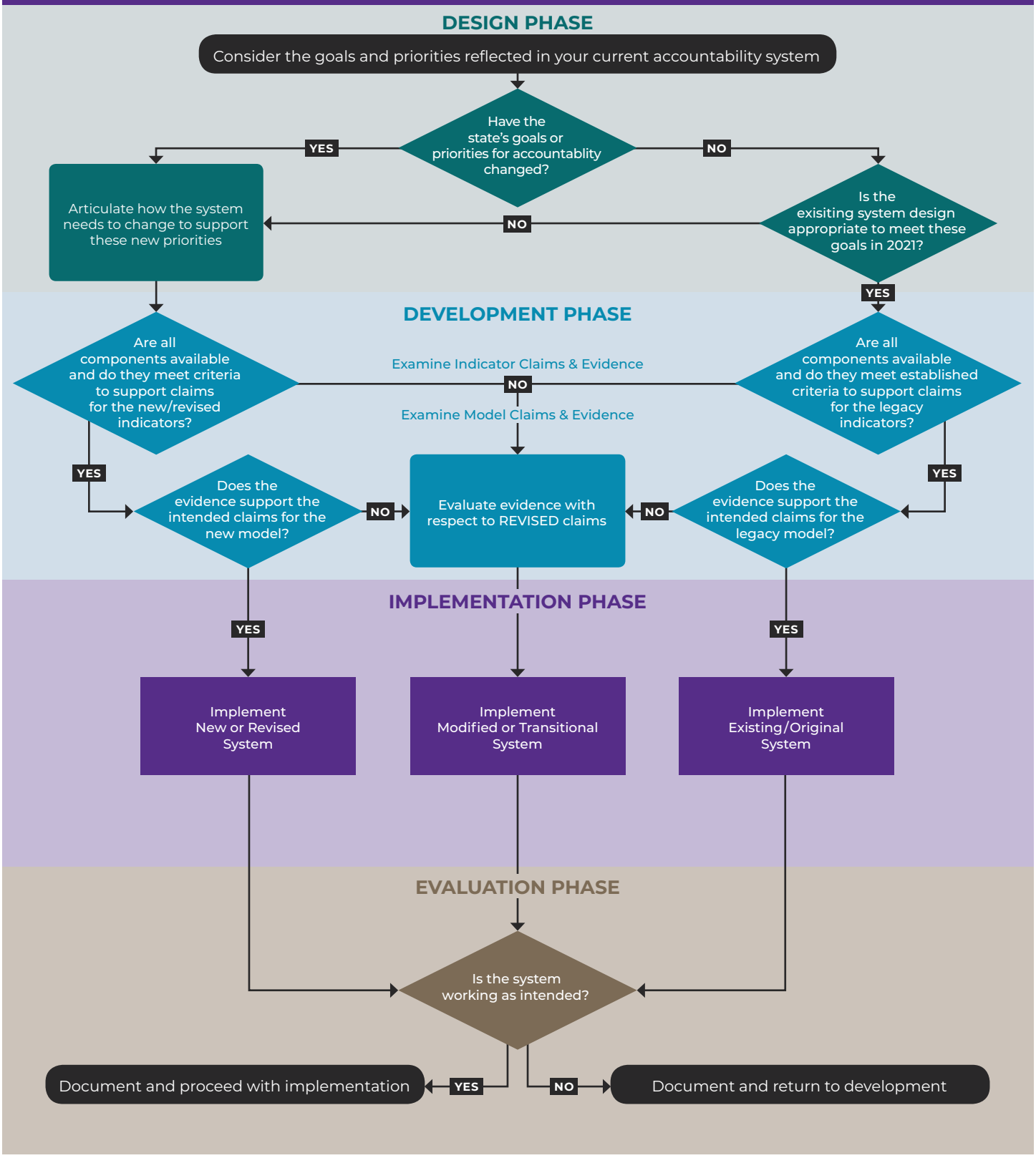
A PROCESS TO EXAMINE DECISIONS FOR ACCOUNTABILITY IN SY 2020-2021

In many ways, making decisions about a state's accountability system in SY 2020-2021 and beyond may be part of a broader initiative to develop a new or revised system. Guidelines for developing and revising systems are well established in the literature (e.g., D'Brot & Keng, 2018; Landl, Domaleski, Russell, & Pinsonneault, 2016; Domaleski, Boyer, & Evans, in press; Perie, Park, & Klau, 2007). In particular, we draw on the D'Brot framework (in press) to describe the phases and tasks that should be addressed throughout the accountability design, development, implementation, and monitoring stages.

The focus of this paper is not to provide a comprehensive guide for the development and evaluation of accountability. Rather, the authors use this process to highlight decisions and actions believed to be the most relevant in the era of COVID-19 disruptions. While the D'Brot framework refers to the development of accountability systems (often without data to test assumptions), many of the stages in this process can be evaluated formatively to collect evidence in advance of finalizing decisions.

Figure 1 (on the next page) illustrates the general phases (and several key decisions) associated with determining the path forward for school accountability.

FIGURE 1 **A PROCESS FOR ADDRESSING KEY ACCOUNTABILITY DECISIONS**



In brief, the four phases in Figure 1 are:

1. **DESIGN**

Review the values and goals underlying the design of the legacy system to determine if they appropriately represent the state's beliefs and priorities. Some states may take advantage of COVID-related break(s) in continuity to rethink their priorities for the accountability system and, and in turn, build a new path forward. Other states may affirm their existing design principles and focus on a course that restores the legacy system.

2. **DEVELOPMENT**

Determine the extent to which the information required to meet the state's goals is available and appropriate. State leaders are encouraged to first take stock of the data elements necessary for individual indicators, including an examination of whether the indicator meets feasibility and acceptability criteria. Then, the state can evaluate the extent to which the overall model can support the claims and the intended use-case.

3. **IMPLEMENTATION**

Determine an implementation plan in SY 2020-2021 that takes into account information gained during the development phase. States may find that the new or legacy system supports the state's claims and priorities and can be implemented as intended in SY 2020-2021. Alternatively, states may discover limitations that will impede the ability to roll out a complete model in SY 2020-2021, suggesting that it would be preferable to employ a transitional system supporting an abbreviated set of claims and uses.

4. **EVALUATION**

Evaluate the system using data from SY 2020-2021 to determine whether the model functions as intended. This is especially important, insofar as analyses based on pre-pandemic accountability data cannot suitably model conditions experienced during the COVID-19 disruptions. By revisiting the criteria and claims with operational data, states can determine whether the model attributes are affirmed or should be revised.

DESIGN PHASE

To plan for accountability in SY 2020-2021, a state educational agency (SEA) should first consider whether the goals and priorities represented in the legacy system have changed. For many states, the objectives, design principles, and intended outcomes associated with the legacy system design will remain the same (e.g., increase graduation rates, improve college and career readiness, promote equity). For other states, factors related to COVID-19 disruptions may cause SEAs to establish new goals, reprioritize or clarify existing goals, or reconceptualize the accountability system in ways that will necessitate making changes to the system design (e.g., the inclusion, weight, and role of indicators; the procedures used to identify and support low-performing schools). Each scenario is illustrated in Figure 1. The right side of this figure shows a decision-making pathway for those states that have not changed their goals and wish to restore the legacy system; the left side shows a pathway for states that have modified their goals, priorities, or theory of action such that the system requires revision. Determining which path best represents a state's intent for accountability in SY 2020-2021 is the first step in the design phase.

STATE GOALS & PRIORITIES REMAIN THE SAME

If a state's vision for accountability has not changed, implementation of the legacy system may be the primary objective for SY 2020-2021. Note that, given the current context, SEAs should pause before moving forward to be sure they have sufficiently evaluated whether the existing system is still appropriate for achieving the state's goals. For example, after watching its districts scramble in spring 2020 to purchase technology in service of remote learning, SEA leaders may decide to add an indicator regarding the schools' ability to provide equitable access and support for learning in a distributed model. The state's goal has not changed in this case, but the system is revised to better reflect relevant inputs for the current environment. This scenario is represented by the route leading to an examination of revised claims in the design phase of Figure 1.

STATE GOALS & PRIORITIES HAVE CHANGED

If a state is using the COVID-19 disruptions to introduce new goals or change the way it defines and prioritizes existing goals (i.e., its theory of action[4]), SEA leaders may propose a revised model. In many states, for example, the pandemic has uncovered significant shortcomings in their schools' readiness to provide students with emergency instructional services. Certainly, some shortcomings were inevitable, but others have highlighted preexisting inequities that differentially affect students' ability to learn and educators' ability to teach. For example, certain input-like data elements focusing on schools' readiness to provide remote or blended services might include:

- student access to digital devices and Internet services;
- teacher access to resources, tools, and professional development opportunities that support online instruction; and
- specialized supports for English learners and students with disabilities.

While state leaders may be aware of these inequities, SEAs may not have had the support to address them in the past. During COVID-19 disruptions, states may decide to modify accountability systems to better provide educators and students with the resources necessary for success in a remote-instruction or blended-learning context. In this case, the newly revised system or a transitional version of that system may be implemented in SY 2020-2021, depending on the degree of change required and the data necessary to support it.

[4]A theory of action, also 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 and causal in nature, and it articulates underlying assumptions which are determined by goals (D'Brot, in press).

STATE GOALS & PRIORITIES HAVE CHANGED (CONT.)

If a state does decide to revise its system, system designers should engage in a thoughtful process that includes soliciting stakeholder feedback and, further, developing a theory of action that communicates the rationale for the system design (e.g., indicators, weights, procedures for identification). Given the required components defined within ESEA, most states will begin with their existing state plan when establishing a theory of action and rationale for their revised state system.

By the end of the design phase, a state should have a clear picture of its desired accountability system for SY 2020-2021 (i.e., legacy or revised), assuming the school year proceeds as intended and there is sufficient evidence to support the intended use of results.

DEVELOPMENT PHASE

The development phase involves evaluating components of the legacy or newly revised system against defined criteria and, in turn, making plans for implementation in SY 2020-2021. This process begins with the examination of indicators.

EXAMINE INDICATOR CLAIMS & EVIDENCE

When developing an accountability system, SEAs carefully select and operationalize indicators that will meet the state's goals. This process includes specifying the conditions, both technical and practical, that must hold for the indicator to function as intended. For example, a state may indicate that the growth measure should (a) differentiate among schools, (b) allow all types of schools to demonstrate scores across the score scale, and (c) remain unrelated to sample size and other school characteristics. For an indicator to be interpreted the same way in SY 2020-2021, evidence must be collected to demonstrate that the claims associated with each indicator still hold. If claims cannot be supported, the SEA may need to change the business rules, calculation procedures, performance expectations, or the decision to include the indicator in the system for SY 2020-2021. This analysis is a crucial first step toward determining whether the proposed system can be implemented as intended and, if not, what modifications are necessary.

When evaluating indicators, one should consider how each indicator might be affected in SY 2020-2021, due to a lack of data and other consequences of the COVID-19 disruptions. To that end, this paper draws on four evaluation criteria outlined by Domaleski, Boyer, and Evans (in press), each of which affects the degree to which data elements or indicators should be included in the accountability system:

COMPLETENESS — Are data elements missing? Do the data capture the full breadth and depth of findings as expected prior to COVID-19? While the criterion for completeness is rarely 100 percent, it is more appropriate to evaluate completeness as the deviation from pre-pandemic standards. Checks for completeness should include multiple disaggregations (e.g., by school, student group, program).

CONSISTENCY — Were data properties altered? Specifically, did COVID-19 disruptions change the ways in which data are defined, calculated, or collected? This will affect both the individual metrics (e.g., the availability of Advanced Placement or International Baccalaureate data) and how they are aggregated (e.g., overall school quality/student success indicators or grade-point averages).

IMPACT — How is the interpretation of performance on the individual data elements or overall indicator impacted? Is it likely that data values (e.g., performance) will change substantially? Do values change based on other circumstances, even if the elements are complete and calculated based on the same procedures? The answers to these questions will inform data reporting.

PRACTICALITY — Is it reasonable to collect and report the data? Will collecting it cause undue burden on or deflect from higher priorities? If the data could be misunderstood, misinterpreted, or misused, it may be advisable to withhold these data from collection and/or reporting.

Table 2 shows an example of review outcomes for each of these four criteria.

TABLE 2 EXAMPLE OF REVIEW FOR COMPLETENESS, INCONSISTENCIES, IMPACT, & PRACTICALITY

THE GAPS IN COMPLETENESS ARE:		
LOW	**MODERATE**	**HIGH**
The indicator is complete. The depth and breadth of data elements are unchanged. When compared to pre-pandemic findings, completeness appears to be sufficiently similar.	There is some incompleteness in the indicator. The depth and breadth of data elements demonstrate some differences. When compared to pre-pandemic circumstances, there is some deviation from the typical completeness of the indicator.	The indicator is incomplete. The depth and breadth of data elements are not comparable to that of pre-pandemic data. There are significant deviations from the typical completeness of the indicator, pre-pandemic.
THE INCONSISTENCIES IN DATA ELEMENTS ARE:		
LOW	**MODERATE**	**HIGH**
There are no inconsistencies in data properties for this indicator. Calculations, definitions, and data collection are unchanged. Aggregations based on these data elements should not be affected by pandemic-related disruptions.	There are some differences in data properties for this indicator. Calculations, definitions, and data collection may reveal some changes. Aggregations based on these data elements may be affected and should be examined to determine whether the affected aggregations have an impact on data interpretation.	There are significant differences in data properties for this indicator. Because of changes or inconsistencies in calculations, definitions, and data collection, aggregations may be significantly affected or not feasible. State leaders should determine whether the current data can be compared with pre-pandemic findings.
THE CHANGE IN IMPACT TO DATA ELEMENTS ARE:		
LOW	**MODERATE**	**HIGH**
There are no known novel sources of impact on the performance of this indicator due to COVID-19 disruptions.	There is some potential for novel sources of impact on the performance of this indicator, due to the pandemic.	There is a strong potential for novel sources of impact on the performance of this indicator; the impact may be substantial.
THE RISK TO PRACTICALLY COLLECTING AND REPORTING THE DATA IS:		
LOW	**MODERATE**	**HIGH**
There are few, if any, threats to the feasibility or reasonableness of collecting and reporting these data. There are few threats to interpretation or use of these data if they are included in the accountability system.	There are some threats to the feasibility or reasonableness of collecting and reporting these data. There may be threats to interpretation or use of these data if they are included in the accountability system.	The threats to the feasibility or reasonableness of collecting and reporting these data are high. There are many threats to interpretation or use of these data if they are included in the accountability system.

DEVELOPMENT PHASE

Following Domaleski et.al. (in press), this paper recommends summarizing the results of a state's evaluation by assigning each indicator an overall rating. For example, the state may decide to bin data elements to capture the likely impact on indicator-score interpretation and use, as in Table 3 and the accompanying definitions.

TABLE 3 EXAMPLE SUMMARY OF OVERALL EVALUATION OF INDICATORS

	COMPLETENESS	CONSISTENCY	IMPACT	PRACTICALITY	BIN
CHRONIC ABSENTEEISM	Low	Low	Low	Low	Red
ACHIEVEMENT	Low	Low	High	Low	Yellow
GROWTH	Moderate	High	High	Moderate	Green

DEFINITIONS

GREEN	The evaluation suggests indicator scores can be interpreted and used as intended.
YELLOW	Additional analyses are necessary to determine the degree to which indicator scores are likely to support intended interpretations and use.
RED	It is unlikely that indicator scores can be interpreted and used as intended, or in any way that is consistent with how they have been in the past.

EXAMINE MODEL CLAIMS & EVIDENCE

Accountability systems are reliant on parts all working well together to support the whole. That is, the individual components must function as intended, both individually and collectively, in order for the design to meet expectations for the state's system of Annual Meaningful Differentiation (AMD). As stated in statute, AMDs must result in meaningful differentiation and enable states to identify those schools most in need of support.[5] A state's evaluation of its design, development, and implementation processes and procedures must include an examination of both individual indicators and the system of AMD overall.

Once a state evaluates its individual indicators, the accountability system as a whole is then evaluated to determine the degree to which the claims underlying each component of the accountability system will likely hold. Claims are statements about the system, system activities, and intended outcomes. As an illustration, Table 4 presents a series of high-level claims, organized by system component, developed as a joint effort by Juan D'Brot from the National Center for the Improvement of Educational Assessment (Center for Assessment), the State Support Network, and the U.S. Department of Education's Office of State Support (State Support Network, in press).

These claims are intended to help states evaluate whether results and evidence sufficiently meet the claims an accountability system should substantiate. Generally, these claims are organized into policy claims, technical/operational claims, and impact claims. While the policy and impact claims are important, they likely can be evaluated through qualitative or conceptual reviews; states should focus primarily on the technical/operational claims under COVID-19 disruptions, focusing on their systems of AMD.

TABLE 4 INDICATOR & SYSTEM OF AMD CLAIMS

SYSTEM COMPONENT	POLICY CLAIM	TECHNICAL/ OPERATIONAL CLAIM	IMPACT CLAIM
INDIVIDUAL INDICATOR WITHIN THE SYSTEM OF AMD	The indicator aligns with the state's overall system theory of action and its policy objectives. The indicator fairly represents the construct as intended.	The indicator supports valid and reliable results. Measures that constitute the indicator can be compared and differentiated appropriately. The indicator contributes as intended to the state's system of AMD.	Data from the indicator are useful to consumers of the system, because these data represent important signals of a school's performance. The data from the indicator are understandable. The indicator provides sufficient information for supporting continuous improvement through reporting and resources to aid interpretation.
INDICATOR INTERACTION FOR THE STATE'S SYSTEM OF AMD	The indicator's weights or decision rules reflect the state's theory of action and stakeholder vision.	The empirical indicator weights reflect the intended state priorities and promote valid, fair, and reliable school ratings. The empirical results of decision-rules reflect the intended sequencing of decision rules to promote valid, fair, and reliable school ratings.	The indicator weights or decision rules do not impede the usefulness or interpretations of how schools are differentiated.

[5] ESEA, as amended by ESSA, requires that states identify schools in need of Comprehensive Support and Improvement (CSI), Additional Targeted Support and Improvement (ATSI), and Targeted Support and Improvement (TSI). Please see Lyons & D'Brot (2018) for a description of the identification requirements in statute.

TABLE 4 INDICATOR & SYSTEM OF AMD CLAIMS (CONT.)

SYSTEM COMPONENT	POLICY CLAIM	TECHNICAL/ OPERATIONAL CLAIM	IMPACT CLAIM
SYSTEM OF ANNUAL MEANINGFUL DIFFERENTIATION	Results from the state's system of AMD align with objectives and policies around subgroups and school size, setting, and demographics.	School rankings and groupings created via the state's system of AMD reflect data as intended. That is, rankings are not skewed or inappropriately distributed, or do not include schools that are unexpectedly low or high performing.	Results from the state's system of AMD reflect meaningful differentiation among schools.
ESEA IDENTIFICATION OF SCHOOLS NEEDING SUPPORT	Identification of schools aligns with the overall system theory of action—i.e., they have subgroups most in need of support. Identification meaningfully captures all grade spans. Identification supports subgroup-specific objectives.	Identification and exit mechanisms for schools reflect meaningful differentiation within and across school classification(s).	Identification of schools in need of support results in districts and schools engaging in meaningful exploration of, and continuous improvements in, these schools in response to indicator results.
REPORTING	Reporting is designed to communicate the objectives and results of the accountability system with multiple users in mind.	State and local report cards and reporting systems provide access to accurate data to support the AMD system.	State and local report cards and resources facilitate meaningful exploration of accountability data and stimulate inquiry into ways to ensure continuous improvement.

To inform decisions about how the system should be modified in SY 2020-2021, state leaders are encouraged to determine which claims are at risk of not being supported in SY 2020-2021 and why. While some claims might be evaluated using historical data and simulation, many will need to be evaluated (or revisited) once operational data are available.

DEVELOPMENT PHASE

States also would do well to evaluate more general, system-level claims by holistically examining the overall impact of AMD on the accountability system. While system-level claims are dependent on the component-level claims above, the system-level claims should be considered in light of the state's priorities and desired impact on behaviors and potential interpretation. Table 5 provides an example of a system-level evaluation of claims.

TABLE 5 SYSTEM-LEVEL EVALUATION OF CLAIMS

DECISION POINT TO CONSIDER ON OVERALL SYSTEM OF AMD	OVERALL IMPACT TO SYSTEM (LOW, MODERATE, OR HIGH)
Impact on aggregated weights or sequence of decision rules to the overall system of AMD	Moderate
Impact on the rankings and groupings created via the system of AMD	Moderate
Impact on the meaningful identification of CSI, TSI,[6] and ATSI[7] schools	Low (due to delay in identification, based on missing 2019-2020 data)
Impact on the timing of identification of CSI, TSI, and ATSI schools	Low (due to delay in identification, based on missing 2019-2020 data)

If a state's examination yields a moderate or high overall impact on its accountability system, it is unlikely that the state will be able to fully implement a legacy or newly revised system in SY 2020-2021. In this case, SEA leaders will need to consider different options for SY 2020-2021, as described in Table 6.

TABLE 6 EXAMPLE OF ACCOUNTABILITY OPTIONS BASED ON REVIEW OF SYSTEM IMPACT

IMPACT TO OVERALL SYSTEM	OPTIONS, BASED ON REVIEW OF SYSTEM IMPACT
HIGH	Explore a transitional system of accountability. A waiver or amendment will likely be necessary, because implementation would require making substantive changes to process, procedures, policies, or data collection.
MODERATE	Explore a transitional system of accountability. Evidence will determine whether a legacy or revised system is feasible. A waiver or addendum may be necessary, if changes to calculations, properties, or procedures could be construed as substantively different (even if changes only seem minor).
LOW	Implement a legacy or revised system. Note that a legacy system should require sufficient documentation justifying that data are complete, consistent, of similar interpretation, and practicable. A revised system should include the same documentation and will require an amendment to the state's ESEA consolidated state plan.

These three options shall be further explained, with examples, in the following section on the implementation phase.

[6]Targeted Support and Improvement (TSI).

[7]Additional Targeted Support and Improvement (A-TSI).

IMPLEMENTATION PHASE

As shown in Figure 1, a state's implementation plan may pertain to a legacy system, a revised system, or a transitional system. It should be noted that, if a state plans to fully implement a legacy or revised system, doing so suggests that all components of the system will be available in 2020-2021 and that the results can be interpreted and used as intended.

A transitional system—i.e., one which deviates in some manner from the intended system design—is proposed when evidence from the development phase suggests some aspect of the legacy or newly revised system cannot be calculated, collected, or interpreted as intended. If full implementation of the legacy or revised system cannot be supported, states will need to reflect on lessons learned and information gathered from the development phase to identify and prioritize options for a transitional system in the spring of SY 2020-2021. Transitional systems may vary with respect to not only what is modified (e.g., indicators, measures, calculations), but also the degree to which the modification deviates from the intended design.

Coming out of the development phase, a state may already have a good idea about what needs to change and how to change it.

In most cases, however, the data necessary to fully evaluate a plan for modification will not be available until the end of SY 2020-2021. For example, many states are planning to calculate estimates of growth in 2020-2021, despite missing summative data from the 2019-2020 school year. Preliminary data may suggest that such a plan is reasonable and that system results can be used as they have in previous years; however, data from 2020-2021 may suggest otherwise (i.e., the claims or assumptions related to this indicator do not hold). Consequently, it would be better for the SEA leaders to develop a plan that incorporates a sufficient number of likely or representative scenarios. Doing so can prepare the state to pivot quickly toward implementing an alternative, if conditions supporting the preferred option should change.

DIMENSIONS OF MODIFICATION

Table 7 outlines five dimensions along which a state may modify its system. For each dimension, since the degree of change and its impact on the overall system can vary significantly, it is wise not to consider one modification as less (or more) significant than another. The last column in this table provides examples of evidence that may cause an SEA to consider a particular modification when determining options for SY 2020-2021. This evidence aligns with findings collected through an examination of indicators and the evaluation of system-level claims, as described above. In fact, many of these dimensions overlap with the system components discussed in Table 2 (see "gaps in completeness"). Thus, some states may have already addressed a number of the potential modifications aligned with these dimensions.

IMPLEMENTATION PHASE

TABLE 7 DIMENSIONS OF MODIFICATION

DIMENSION	DESCRIPTION OF MODIFICATION	EXAMPLES OF WHEN TO CONSIDER
INDICATORS/ MEASURES	Refers to decisions that impact: - business rules used to compute measures, scale scores, or indicator results - inclusion of indicators in the system[8]	Indicator data are unavailable or of insufficient quality (e.g., low reliability). Indicator does not differentiate as intended. Indicator performance is affected by factors compromising the interpretation or utility of results.
SUMMATIVE DETERMINATIONS (ANNUAL MEANINGFUL DIFFERENTIATION)	Refers to decisions that influence how the overall score or rating is calculated and interpreted: - whether an aggregate score is produced for schools and, if so, how; - adjustments to indicator weights (e.g., if an indicator is missing or changed, an adjustment to the weights may be appropriate).	Scores for one or more indicators are missing or unreliable. Evidence suggests that the overall school rating, as calculated, cannot be interpreted as intended or might lead to misinterpretations. The overall school rating does not meaningfully differentiate. The indicator's weights do not reflect the intended state priorities, due to changes in performance characteristics.
PERFORMANCE EXPECTATIONS	Refers to decisions related to long-term goals and measurements of interim progress, rules for entry/exit into support categories, and the standards defining different levels of performance on specific indicators (e.g., does not meet, meets, exceeds expectations).	Evidence suggests that existing timelines or expectations for school or student-group performance are inappropriate because of COVID-19 disruptions. Performance expectations are tied to identification and exit decisions.
IDENTIFICATION DECISIONS	Refers to decisions related to how the system will be used to identify schools for CSI/ATSI or to exit schools from this status.	Results do not support the attribution of overall scores or ratings to school performance (i.e., too many externally related factors). Identification procedures do not accurately flag schools that are most in need of support or have made adequate progress to exit. Data are not available to support defined procedures for annual, meaningful differentiation.
REPORTING DECISIONS	Refers to decisions about what should be reported, and how.	Indicator or overall scores or ratings cannot be produced, or this data may be misused/ misinterpreted if provided.

[8] *Indicators* are the elements of school performance included in the system, such as academic achievement, college career readiness, and growth. *Measures* are the data used to quantify performance on each indicator, such as proficiency or graduation rate.

Identifying what can/should be modified is important, but ultimately it is the degree of modification a state would need to make that determines what to report, how to use the results and, consequently, whether a state will require a waiver or amendment to support a revision to the its plan or system. It is therefore vital for a state to be extremely thoughtful when determining the appropriate modification(s) to be made.

The best path forward will vary, depending on the state's goals and priorities for SY 2020-2021. For example, if a state's primary goal is to identify low-performing schools for CSI but the evidence does not support implementation of the full system, state leaders might limit modifications to only those that reflect the highest priorities for the schools most urgently in need of support. Such priorities may differ from modifications required to differentiate among performance of higher-performing schools. On the other hand, if a state's primary goal is simply to report information about school performance with no implications for classification or comparison (i.e., if identification and comparison to previous years are not a concern), the state will have more flexibility with respect to the types of modifications it can be make and what it can report.

To clarify: the way in which results can be used will vary, depending on (a) the degree to which a state's model deviates from the intended design and (b) the degree of confidence an SEA has that the system's results will be interpreted and used as intended. The less confidence an SEA has in its data, the more modifications it is likely to make to the system in order to meet at least some of the state's goals for 2020-2021. As the number of modifications increases, comparability to previous years' results decreases; in addition, data-uses tied to high-stakes decisions (e.g., identification, exit from support status) are less likely to be supported.

Even if a state has a high level of confidence in the data and has not modified its system, it will be difficult for SEA leaders to substantiate claims that school-improvement efforts were sufficiently implemented to have taken root during SY 2019-2020 and during the start of SY 2020-2021. Therefore, the availability of and confidence in the data are necessary, but insufficient conditions to attribute changes in performance to specific school-improvement initiatives. Stated another way, disentangling 'pandemic effects' from other influences on school performance will be challenging, making it difficult to support high-stakes decisions like entry/exit from ESEA school designations.

Ultimately, when deciding how to use accountability data, state leaders will need to evaluate against the design, intended uses, and risks for misinterpretation—keeping in mind each of the dimensions listed in the table above. As the level of confidence decreases and/or the system deviates from its intended design, data-uses will become more limited and descriptive in nature. If it becomes apparent that a state should not use the data for a legacy or revised system, SEA leaders might consider implementing a transitional system. For example, a transitional system would be appropriate for informing and reporting, but not for identifying schools in the traditional manner.

USE-CASES TO IDENTIFY CLAIMS

Since the way in which a system is modified will influence the uses it can support, SEA leaders need to consider the claims which must hold in order to support a particular use, and what these claims suggest about the needed modification. Table 8 presents the primary claims (with increasing stakes) and provides a brief description of the evidence required to support it. The last column of this table highlights inappropriate modifications—i.e., modifications which should not be considered if a given use is desired (i.e., if making that modification would serve as a barrier to demonstrating the associated claim).

IMPLEMENTATION PHASE

TABLE 8 CLAIMS & EVIDENCE TO SUPPORT A SPECIFIED USE OF ACCOUNTABILITY DATA

USES (FROM LOWER- TO HIGHER-STAKES)	PRIMARY CLAIM	EXAMPLES OF EVIDENCE	INAPPROPRIATE MODIFICATIONS
DESCRIBE A SCHOOL'S PERFORMANCE IN SY 2020-2021	Accuracy & utility: Indicators and overall ratings (if calculated) provide accurate, useful information about a school's performance in SY 2020-2021.	Required data are available; indicator calculations are feasible; conditions required to interpret the indicators as intended hold.	Business rules that may influence the reliability or accuracy of results for some schools (e.g., reducing N-counts, inclusion rules)
COMPARE PERFORMANCE ACROSS SCHOOLS WITHIN SY 2020-2021	Within-year comparability: Scores or ratings that serve as the basis for comparisons across schools can be interpreted similarly and demonstrate sufficient variability to support meaningful comparisons.	Measures and ratings can be calculated similarly for all schools. Evidence demonstrates similar levels of data completeness and accuracy across schools.	Changes to procedures that allow for different business rules or calculation procedures across schools.
EVALUATE TRENDS IN SCHOOL PERFORMANCE ON SELECT INDICATORS	Between-year comparability: Indicators are calculated using the same or similar procedures and can be interpreted as in previous years.	The degree or incidence of missing or extreme data is not significantly different than in previous years. Measures do not relate to school-related factors (e.g., N-size) or student demographics (e.g., % free-reduced lunch) in unexpected ways.	Changes in indicator calculation procedures, business rules (e.g., inclusion criteria) or performance expectations.
FLAG SCHOOLS THAT ARE PERFORMING "FAR BELOW EXPECTATIONS"[9] (E.G., EARLY WARNING)	Within-year comparability & differentiation: Scores or ratings provide useful, accurate information for identifying low-performing schools and demonstrate sufficient variability to support meaningful differentiation.	Measures and ratings can be calculated similarly for all schools. Evidence demonstrates similar levels of completeness and accuracy across schools. Results do not relate to school factors (e.g., N-size; % free-reduced lunch), student-group characteristics, or other factors that threaten the use of results as an early warning for identification in 2022.	Significant changes to indicator calculations, business rules, aggregation procedures, or other design decisions that significantly change what it means to be low-performing, as compared with what is intended (i.e., in the legacy or revised system).
IDENTIFICATION OF SCHOOLS FOR ENTRY OR EXIT FROM CSI/ATSI, PER ESEA	Schools that are identified for CSI, ATSI, and TSI are the ones in need of support. Improvements in accountability data reflect sufficient progress to warrant removal of support.	Entry and exit criteria are consistent with prior decisions and/or prior designs, and data are available to evaluate criteria. All grade spans are meaningfully captured. Identification and exit criteria sufficiently capture differentiation in school performance.	Significant modifications to any of the dimensions reflected in Table 7.

After working through the implementation phase, the state will have examined substantial evidence and decisions to help identify approaches to SY 2020-2021. This paper shall now address the final stage, evaluation, in which the state assesses each proposed option to determine which would be the most reasonable and fair. Evaluation is presented here as the final phase of the process, but states may conduct evaluations throughout the design, development, and implementation phases before the chosen system—be it legacy, revised, or transitional—becomes operational.

[9] It is important to note that the evidence necessary to support this claim depends on whether "far below expectations" is operationalized normatively or against some pre-defined criterion that represents an "expected" amount of change. In this example we are assuming that "far below" is defined in terms the degree of change observed by a school from compared to that demonstrated by similar schools in the state.

EVALUATION PHASE

An ESEA accountability system should support the state's overall theory of action, while:

- meaningfully differentiating schools;
- identifying schools in need of comprehensive, targeted, and additional targeted support; and
- improving student outcomes for all students.

This final phase comprises the evaluation of all preceding phases (design, development, and implementation) by methodically substantiating claims with compelling evidence (D'Brot, in press; D'Brot, Keng, & Landl, 2018). Note that, if a state is using a new accountability system, these evaluative processes would occur after implementation. By contrast, the evaluation of legacy-system revisions and modifications should be ongoing throughout the design, development, and implementation phases. The process of ongoing evaluation is especially relevant for the review of accountability systems in the era of COVID-19 disruptions.

In evaluating their accountability systems to determine the best options for spring 2021 implementation, states need to:

- review the individual components (and their interactions) of the accountability system;
- identify component- and system-level claims that need to be evaluated, using evidence and not just logic;
- document areas where the accountability system is/is not functioning as intended;
- identify results for informing stakeholders about the accountability system, including strengths and limitations of the system to date; and
- identify necessary adjustments, revisions, or adaptations to the accountability system.

By evaluating how a state's accountability system functions in the face of pandemic-related data losses (i.e., examining empirical results from modeling, simulations, or operational results), SEA leaders can determine how well the system components align with policy and state priorities. Given that it would be exceedingly difficult to model the likely impact of COVID-19 disruptions, states are encouraged to develop an evaluation plan based on operational data—i.e., data collected after implementation in SY 2020-2021—in order to reexamine the degree to which criteria were met and assumptions held. This evaluation may entail making additional modifications to the model or associate claims prior to final roll-out of the system.

CONCLUSION

The pandemic-related challenges facing SEAs are numerous and far-reaching. In an effort to help states address these challenges, this paper has presented:

- a set of guiding principles for restarting accountability;
- a process to guide decision-making; and
- considerations for implementation in spring 2021.

While the focus has been limited to accountability systems under ESEA, the authors' argument also applies to a broader range of accountability, reporting, and support initiatives.

As a result of engaging in the process presented here, state leaders will be able to answer these questions:

- How well does our system align with the state's theory of action, policy goals, and priorities?
- Do we need to revise our system to better reflect existing or shifting state priorities?
- Which of our indicator- and system-level claims should be evaluated? What is the impact on our overall system if certain claims cannot be substantiated?
- Given the potential impact of COVID-19 disruptions on the system's claims and data, how should we approach accountability implementation in spring 2021? How should the results be used?

The authors hope that state education leaders can leverage the strategies presented in this paper to evaluate system-specific claims, assumptions, and potential sources of evidence to support defensible identification systems impacted by COVID-19.

REFERENCES

D'Brot, J. (2018). A framework to monitor and evaluate accountability system efforts. Dover, NH: National Center for the Improvement of Educational Assessment.

D'Brot, J. (in press). Operational Best Practices for Accountability. Washington, D.C.: The Council of Chief State School Officers.

D'Brot, J. & Keng, L. (2018). An Introduction to Accountability Implementation. Washington, D.C.: The Council of Chief State School Officers.

D'Brot, J., Keng, L., & Landl, E. (2018). Accountability identification is only the beginning: Monitoring and evaluating accountability results and implementation. Washington, D.C.: The Council for Chief State School Officers.

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

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.

Domaleski, C., Boyer, M. & Evans, C. (in press). Accountability interrupted: guidance for collecting, evaluating, and reporting data in school year 2019-2020. CCSSO: Washington, D.C.

Domaleski, C., Betebenner, D., & Lyons, S. (2018). Promoting More Coherent and Balanced Accountability Systems. Dover, NH: National Center for the Improvement of Educational Assessment.

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

Perie, M., Park, J. & Klau, K. (2007). Key elements for educational accountability models. Washington, DC: Council of Chief State School Officers.

State Support Network (in press). *Evaluating State Accountability Systems*. Washington, D.C.: American Institutes for Research.