July 2019

CCSSO
Council of Chief State School Officers

Managing Changes That Affect Accountability Outcomes

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, Bureau of Indian Education, 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.

Managing Changes That Affect Accountability Outcomes

COUNCIL OF CHIEF STATE SCHOOL OFFICERS
Pedro A. Rivera (Pennsylvania), President
Carissa Moffat Miller, Executive Director

Leslie Keng and Charlie DePascale
Center for Assessment

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

CONTENTS

INTRODUCTION

Heraclitus of Ephesus said that "The only thing that is constant is change." This statement is an apt description of state accountability systems over the past few years. With the passage of the Every Student Succeeds Act (ESSA), many states have significantly refined their existing school accountability systems or completely redesigned new systems in recent years. For example, most states have introduced, eliminated, or redefined the measures and indicators that are part of their systems of annual meaningful differentiation (AMD) of schools under ESSA. Some states have implemented different approaches for aggregating school outcomes, established new performance categories for schools, and defined new rules for identifying schools that are in need of support. More changes can also be expected as various policy initiatives are rolled out and states evaluate and continuously improve their ESSA accountability systems[1].

"Change" also seems to be the operative word for state assessment programs. In April 2018, the Council of Chief State School Officers (CCSSO) gave an informal survey to 21 states attending its Accountability State Plan Implementation Meeting. One of the survey questions asked each state about recent changes to its assessment program. The results are shown in Table 1.

Table 1. Results of CCSSO informal survey on changes in state assessment programs (April 2018)

Summary of Recent Changes to State Assessment Program

• 16 states have changed assessment programs
  o 10 changed from Common Core assessments to state-developed assessments or SAT/ACT
  o 6 are making changes to their existing assessment programs
• 12 states have changed testing vendors
• 8 states have transitioned from paper-and-pencil to online assessments
• 5 states have shortened their tests
• Other states have implemented new science and/or social studies assessments, removed performance tasks, shifted from untimed to timed tests, changed to 100% machine/artificial intelligence (AI) scoring, added writing tasks, or moved away from an end-of-course model

Many of these changes have been motivated by demands from the field for shorter testing time and faster score reporting. States are also feeling the pressure to produce assessment results that can serve multiple purposes including informing instruction, measuring student progress, determining readiness for college and careers, evaluating teacher effectiveness, and supporting federal accountability requirements under ESSA.

Indeed, many of the changes to the accountability and assessment systems are necessary and, if implemented with fidelity, can effectively support the ultimate goal of any educational system—improving learning for all students. However, implementing these changes can be a complicated, multi-layered, and multi-faceted endeavor involving many actors and stakeholders. A change

1 For CCSSO resources on the evaluation and continuous improvement of accountability systems, please check out: https://ccsso.org/resource-library/accountability-identification-only-beginning

to one component or aspect of either accountability or assessment systems can lead to a chain reaction that affects other components or aspects in both systems. Changes also generally have implications and consequences, intended or unintended, on the interpretation and use of accountability outcomes. States therefore need to thoughtfully plan and carefully manage the changes with respect to implementation, reporting, and communication.

The goal of this brief is to describe a logical framework and practical recommendations for states as they wrestle with the continuing demands on their accountability outcomes despite substantial changes to their educational systems. We provide a systematic approach to diagnosing the potential effects a change to a state's assessment or accountability system could have on accountability outcomes. We also provide practical advice for states as they consider additional changes to their systems.

THE COMPARABILITY DEMAND AND THE COHERENCE QUESTION

Even amid changes to a state's accountability or assessment system, there is often a desire or mandate for the new accountability outcomes to be *comparable* to those in the old accountability system to preserve longitudinal trendlines and interpretations of school performance. One key factor in determining whether comparability interpretations or claims can be supported is the *coherence* of the changes to the system. In other words, is the new system logically consistent with the old system?

Consider, for example, states that have selected new testing vendors for their assessment programs. If we assume all design aspects of the assessment system, such as the content standards, blueprint requirements, item types, and task models, remain the same across the vendor transition, then the coherence question asks: to what degree is the new vendor able to administer, score, and report on the state's assessment in a consistent manner as the previous vendor?

Another example is the ESSA requirement to define and incorporate a student success and school quality (SQSS) indicator in the accountability system. Many states have introduced new measures such as chronic absenteeism, access to and completion of advanced coursework, postsecondary readiness, school climate and safety, and student and/or educator engagement. In this case, the coherence question to ask is: how consistent are the classifications of schools into performance categories before and after the inclusion of the SQSS indicator?

Every time changes such as these arise, a state needs to first determine whether it intends to maintain the comparability of outcomes or claims in its accountability system after the changes are implemented. If so, then the state should evaluate the coherence of the assessment changes, along with possible changes to the accountability system, to ascertain the likelihood of supporting comparable interpretations and uses of accountability outcomes. Table 2 shows the relationship between the *coherence* of changes and *comparability* of outcomes in school accountability systems.

Table 2. Relationship between coherence and comparability

Coherence\ Comparability	State intends to maintain comparability	State does not intend to maintain comparability
Changes are coherent	**Scenario 1**. Represents higher likelihood of maintaining interpretation and use of accountability outcomes	**Scenario 3**. Represents a transparent and cautious approach to implementing a new accountability system
Changes are not coherent	**Scenario 2**. Needs additional analysis or processes to evaluate and maintain interpretation and use of accountability outcomes	**Scenario 4**. Represents a 'reset' of the state's accountability system due to substantial changes

In this table, Scenario 1 represents the optimal case in which a state intends to maintain the comparability of accountability outcome across changes, and the state's evaluation process found evidence for the coherence of the old and new systems such that the effects on the various accountability outcomes are minimal. While this is considered the optimal case, it is unusual that changes are coherently implemented across *all* components in the old and new system, thereby minimally affecting *all* accountability outcomes.

Scenarios 3 and 4 are cases in which the comparability is not a goal. A state in Scenario 3 has implemented a thoughtful approach to rolling out a new accountability system. The state has also given itself the option of making comparability claims should that demand arise at a later point. Scenario 4 is typical of a state that is making a clean break from its previous accountability system due likely to a significant shift in vision or priorities for the state's educational system.

Scenario 2 is probably the most common case found in operational settings. In fact, many, if not all, states that responded to the informal CCSSO survey described in the introduction are probably wrestling with the challenge of maintaining comparable interpretations and claims across their old and new assessment systems in the face of changes that are hard to rationalize as coherent. For such states, the framework in this brief can help pinpoint the components or elements in their systems that are most impactful on accountability outcomes, suggest additional steps to take to evaluate the impact, and determine strategies or solutions to mitigate the effects and meet comparability demands.

With this theoretical framing of the relationship between comparability and coherence in mind, the rest of this brief aims to provide a logical approach and practical recommendations for states as they wrestle with the demand for comparability on the midst of changes to their assessment and accountability systems. First, we present a simple framework for organizing the types of changes that can affect a state's accountability outcomes. Based on this framework, we outline an approach for evaluating the potential effects of a change in the state's systems on the accountability outcomes. We conclude with guidelines and best practices to help states anticipate and prepare for changes to their systems.

Types of Changes that Affect Accountability Outcomes

The Educational Ecosystem

Most state assessment and accountability systems function like an "educational ecosystem". Changes to one component or aspect of either system can affect other components or aspects in both systems. Figure 1 is a graphical illustration of the high-level components in most state educational ecosystems.

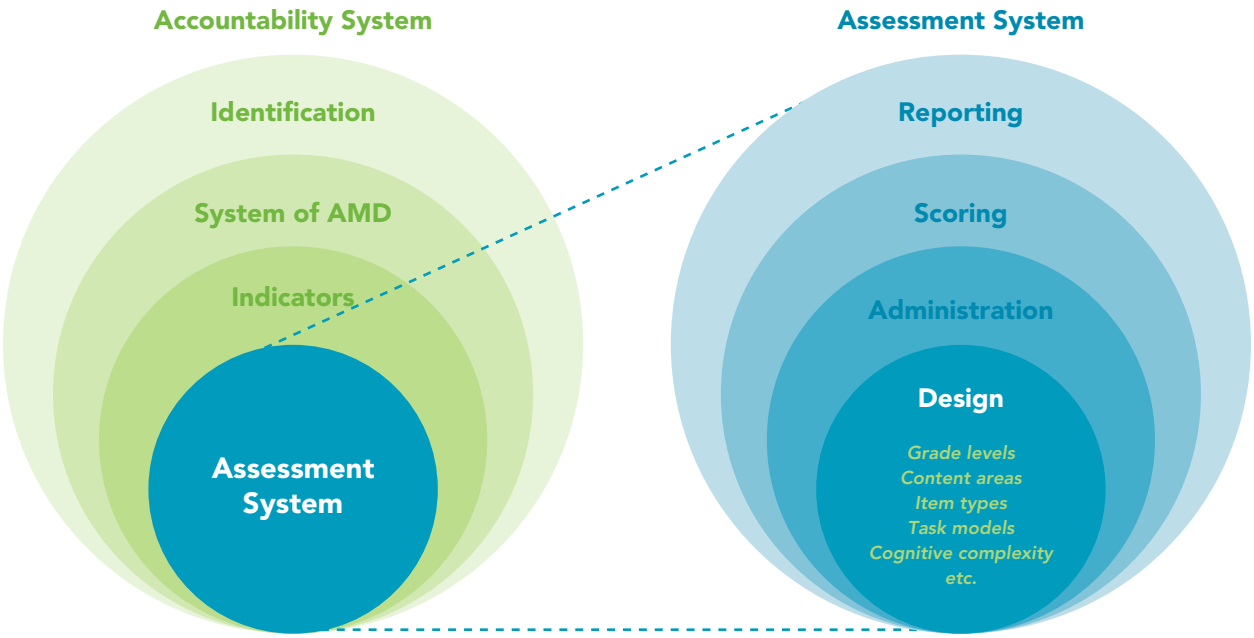

Figure 1. The educational ecosystem

Tables 3 and 4 provides key questions and gives examples of elements within each component of the framework in Figure 1.

Table 3. Components of a state's assessment system

Component	Key Question	Example Elements
Design	What is on the test?	Assessed curricula, content standards, test blueprints, test formats, time limits, allowable accommodations, available accessibility features, etc.
Administration	How is the test given?	Modes of administration, testing interface/platform, test administrator training and instructions, test-taker tutorials, security protocols and procedures, etc.

Scoring	How is test performance determined?	Machine scanning/scoring process, human scoring criteria and protocols, artificial intelligence (AI) scoring approach, psychometrics models, scaling and equating procedures, standard setting method, etc.
Reporting	How are test results communicated and interpreted?	Reported scores and performance classifications, precision reporting, criterion- and norm-referenced interpretations, interpretive guides, levels of reporting, etc.

Table 4. Components of a state's accountability system

Component	Definition	Example Elements
Indicators	What measures are computed for each school or district?	Academic achievement, academic progress, high school graduation rates, progress in English language proficiency, student success and school quality, such as chronic absenteeism rates, measures of school climate, percent enrollment in advanced high school courses, etc.
System of Annual Meaningful Differentiation (AMD)	How are schools or districts classified based on their performance?	Minimum n-counts, definition of student groups, indicator weights, indicator cut scores, baseline years, long-term goals (LTG) and measures of interim progress (MIP), business rules for school/district classification, exceptions, reporting of school classifications etc.
Identification	How is the type of support for low-performing schools or districts determined?	Business rules for comprehensive support and improvement (CSI) and targeted support and improvement (TSI) identification, timelines for identification, communication to identified schools, exit criteria, etc.

Changes in the Educational Ecosystem

Besides providing a systematic view of the components and elements of a state's assessment and accountability systems, we can use this framework to organize the types of changes in the educational ecosystem that affect accountability outcomes. In general, there are three types of changes that can take place under each system:

- Structural Changes: these are changes to the "what" of the systems.
- Process Changes: these are changes to the "how" of the systems.
- Outcome Changes: these are changes to the "so what" of the systems.

Table 5 and 6 show how the types of changes relate to the components in a state's assessment and accountability systems (in Table 3 and 4) respectively and give examples for each type of change.

Table 5. Types of changes in a state's assessment system

Type of Change	Impacted Component(s)	Examples of Change
Structural Changes	Design	• New curriculum • Revisions to content standards • Blueprint changes • Introduction of innovative items
Process Changes	Administration Scoring	• New modes of administration • Transition from human to AI scoring • Changes to psychometric procedures • Updates to computer-based testing interface
Outcome Changes	Scoring Reporting	• Adjustments to performance standards Implementation of a vertical scale • Re-norming or re-scaling • Introduction of growth or progress measures

Table 6. Types of changes in a state's accountability system

Type of Change	Impacted Component(s)	Examples of Change
Structural Changes	Indicators System of AMD Identification	• New statutory requirements • Reorganization of schools and/or districts • Adjustments to subgroup definitions • Change in definition of English language proficiency (ELP)
Process Changes	Indicators System of AMD Identification	• Adjustment to minimum n-count • Updates to LTG and MIP • Changes to business rules for indicators, classification or identification • Adjustments to indicator weights
Outcome Changes	System of AMD Identification	• Adjustments to indicator cut scores • Changes to reported measures and indicators • Revisions to CSI/TSI identification timelines • Changes in consequences or level of support for CSI/TSI identification

Diagnosing the Effects of Changes

Earlier, we described the relationship between the coherence of changes to the assessment or accountability systems and the comparability of accountability outcomes (see Table 2). To summarize:

- Coherence asks: Is the new system logically consistent with the old system?
- Comparability asks: Does the state want to maintain longitudinal trendlines and interpretations of accountability outcomes from year to year? Or, does the state expect to set a new baseline?

The answer to the comparability question is generally driven by the demands of educational stakeholders such as policymakers, administrators, teachers and families. Determining the answer to the coherence question, however, is usually within the state education agency's (SEA) purview to evaluate and address. In this section, we present a principled approach for evaluating the degree of coherence between the old and new systems and diagnosing the potential effect of system changes on a state's accountability outcomes.

EVALUATING COHERENCE

This approach first asks two questions to help the state develop an action plan for evaluating the coherence of their systems before and after the change.

1. What components in state's system does the change affect?
2. What is the nature of the change?

The answer to the first question helps the state determine the assessment or accountability personnel to include in the coherence evaluation process. That is, it suggests WHO should be involved. Tables 7 and 8 provide lists of the specialists, experts, or stakeholders to involve given the system components affected by the change. Note that many of the suggested personnel can come from the SEA, testing vendor, or from an independent third-party organization.

Table 7. Assessment personnel that can help with the coherence evaluation process

Component Affected	Who to involve
Design	• Content specialists • Special education and EL experts • Representatives for educator review committees
Administration	• Test delivery specialists • User interface (UI) design experts • Information technology project manager and technologists • Local test administrators and testing coordinators • Technical advisory committee
Scoring	• Machine scoring/scanning specialists • Performance or AI scoring experts • Psychometricians or data analysts • Technical advisory committee
Reporting	• Assessment reporting specialists • Psychometricians • Technical advisory committee • Report users

Table 8. Accountability personnel that can help with the coherence evaluation process

Component Affected	Who to involve
Indicators	• Implementation specialists, programmers, and/or data analysts • Accountability advisory committee • External consultants who understand and can validate the accountability business rules
System of AMD	• Accountability reporting specialists • Accountability advisory committee • District and school administrators • Legislative or policy representatives • Federal support program (e.g., Title I, Title III) coordinators • Representatives from community stakeholder groups
Identification	• School/district support and improvement specialists • Federal support program (e.g., Title I, Title III) coordinators • Representatives from community stakeholder groups

The answer to the second question (*What is the nature of the change*?) suggests the type of evidence that can help a state evaluate the coherence of its old and new systems. That is, it suggests WHAT evidence should be collected. Table 9 provides examples of the types of assessment and accountability evidence to collect given the nature of the change, using the categorization of changes outlined in the previous section (see Table 5 and Table 6).

Table 9. Evidence to collect for the coherence evaluation process

Type of Change	Evidence to collect
Structural	• Content alignment studies • Blueprints and evidence statement crosswalks • Validation studies for new task models and new item types • Pre/post impact analysis of schools/districts reorganization, new definitions of groups (e.g., disaggregated student groups) or criteria (e.g., ELP), and other statutory changes
Process	• Cognitive labs for testing interface • Mode comparability studies • Interrater reliability/agreement studies • Validation studies for new scoring approach (e.g. AI scoring) • School and district test administrator and coordinator surveys • Pre/post impact analysis of new measure or indicators, updated goals (LTG or MIP), and revisions to business rules
Outcome	• External validity studies • Classification consistency studies • School and district administrator surveys • Pre/post impact analysis of identified districts and schools

Diagnosing Impact

Next, the state should use the evidence collected to diagnose the potential impact of the system changes to accountability outcomes. Table 10 provides an organizer that can help states with the impact diagnosis process. Based on the preponderance of evidence collected during the evaluation process, the state should provide ratings (in the Degree of Coherence column) based on its evaluation of how logically consistent each assessment and accountability component is between the new and old and new systems. The state should then determine, based on the preponderance of ratings and evidence, the degrees to which the comparability of the different accountability components are affected and indicate those in the "Impact on Comparability of Accountability Outcomes" columns. An evidence-based rationale should be given for the state's ratings along with any solutions or approaches to mitigating significant effects on accountability outcomes.

Table 10. Organizer for diagnosing the impact of system changes

System Component	Degree of Coherence	Impact on Comparability of Accountability Outcomes		
		Indicators	System of AMD	Identification
Assessment: Design	☐ Strong ☐ Adequate ☐ Weak			
Assessment: Administration	☐ Strong ☐ Adequate ☐ Weak			
Assessment: Scoring	☐ Strong ☐ Adequate ☐ Weak			
Assessment: Reporting	☐ Strong ☐ Adequate ☐ Weak	☐ Minimal ☐ Moderate ☐ Substantial	☐ Minimal ☐ Moderate ☐ Substantial	☐ Minimal ☐ Moderate ☐ Substantial
Accountability: Indicators	☐ Strong ☐ Adequate ☐ Weak			
Accountability: System of AMD	☐ Strong ☐ Adequate ☐ Weak			
Accountability: Identification	☐ Strong ☐ Adequate ☐ Weak			
Rationale for Ratings				
Mitigation Approach				

An Example

To illustrate the use of this approach, consider the common scenario in which a state has removed writing from all ELA assessments. The state would like to maintain comparability in its assessment and accountability results (i.e., this falls under either Scenario 1 or Scenario 2 in Table 2). Given the fundamental change in the test design (i.e., removing writing), we would classify this as Scenario 2, in which additional analyses are needed to determine whether comparability can be maintained. Tables 11 and 12 summarize the state's findings from evaluating the coherence of the changes, the action plan for collecting evidence, the outcome from the impact diagnosis, and the state's mitigation approach.

Table 11. Evaluation and action plan for removal of writing from ELA

Components Affected	• Design (blueprints, performance level descriptors, or PLDs) • Scoring (psychometric procedures) • Reporting (reporting scale, performance standards)
WHO is involved	• ELA content specialists from SEA and vendor • Psychometricians from SEA and vendor • Technical advisory committee • Assessment reporting specialists
Nature of Change	• Structural (change to blueprints and PLDs) • Process (change to scaling and equating procedure) • Outcome (potential new score scale and cut scores)
WHAT evidence is collected	• Content analysis of old and new blueprints and PLDs • Empirical analysis that evaluates the impact on item calibration, scaling, test reliability, and performance level classifications • Outcomes of standards validation process • Feedback from focus groups on updated individual student reports

Table 12. Organizer for diagnosing the impact of system changes

System Component	Degree of Coherence	Impact on Accountability Outcomes		
		Indicators	System of AMD	Identification
Assessment: Design	☐ Strong ☒ **Adequate** ☐ Weak			
Assessment: Administration	☒ **Strong** ☐ Adequate ☐ Weak			
Assessment: Scoring	☐ Strong ☒ **Adequate** ☐ Weak			
Assessment: Reporting	☐ Strong ☒ **Adequate** ☐ Weak	☐ Minimal ☒ **Moderate** ☐ Substantial	☒ **Minimal** ☐ Moderate ☐ Substantial	☒ **Minimal** ☐ Moderate ☐ Substantial
Accountability: Indicators	☐ Strong ☒ **Adequate** ☐ Weak			
Accountability: System of AMD	☒ **Strong** ☐ Adequate ☐ Weak			
Accountability: Identification	☒ **Strong** ☐ Adequate ☐ Weak			

Rationale for Ratings

- *Comparative analysis of the old and new blueprints, conducted by the SEA and vendor ELA specialists, found that the underlying ELA construct is not substantively affected by the removal of writing prompts.*
- *Empirical analysis found that not including writing prompts in the calibration and scaling processes did not have a significant impact on the item parameter estimates and performance level classification of students, overall and by disaggregated student groups.*
- *Outcomes of the standards validation process, which incorporated the use of the new PLDs, showed that most cut scores remained unchanged after the removal of writing items.*
- *Correlation analysis of academic achievement scores before and after the removal of writing found that while there were slight changes in the values of the indicators for schools (due to changes in ELA proficiency rates), schools were not differentially impacted by the change as evident by high correlations overall and by disaggregated student groups across all school types.*
- *The state's assessment and accountability technical advisory committees reviewed the analysis findings and standards validation outcomes and agreed with the conclusions and recommendations.*
- *Feedback from report users showed that they understood the information on the updated individual student and aggregate group reports, and in the interpretive guides.*

Mitigation Approach

- *Per the advice of the assessment and accountability technical advisory committees, the empirical analysis will be replicated during the upcoming operational administrations.*
- *The state will continue to monitor for unexpected shifts in ELA performance, especially for female students, and schools that previously showed notable improvement in writing.*
- *Suggestions from reporting focus groups and stakeholder outreach meetings were incorporated into the new score report templates and interpretive guides.*
- *Communication resources that summarize the evidence and explain the key findings were prepared and presented at stakeholder outreach meetings and published on the SEA's website.*

Anticipating and Preparing for Changes

A wise person once said, "If you don't want the measure to change, then don't change the measure!" The reality is that changes are inevitable and often necessary as part of the continuous improvement process for assessment and accountability systems. States should therefore be proactive in anticipating and managing changes within its educational ecosystems. Figure 2, adopted from Keng & Marion (in press), suggests a series of guiding questions that can help states be more intentional in planning for changes to its system with a comparability mindset.

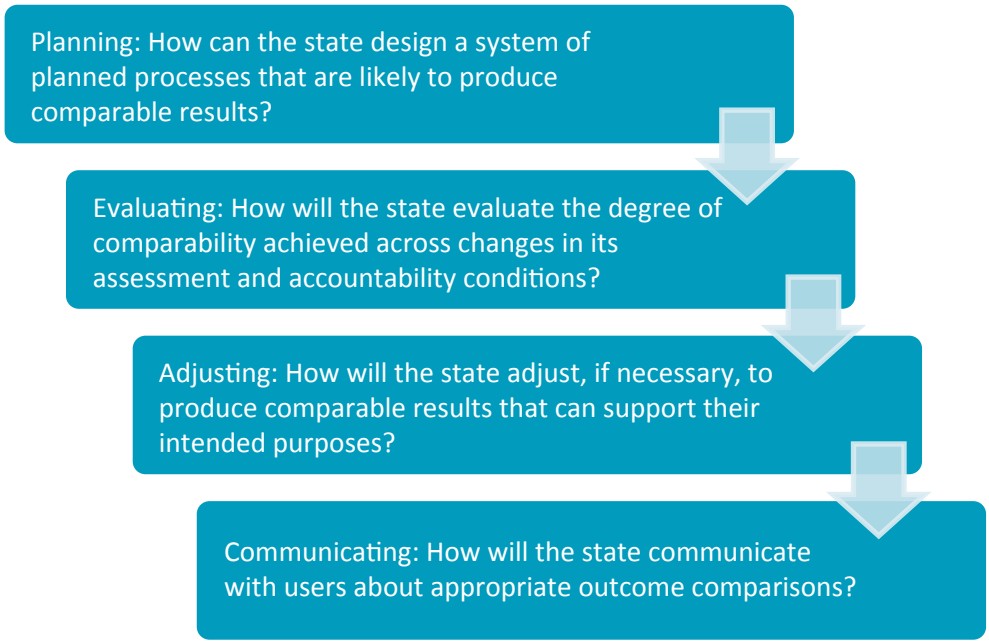

Figure 2. Planning for system changes with comparability in mind

The best way to support comparability is not by gathering evidence from a series of analysis *after* the changes are implemented and deriving *post-hoc* mitigation solutions to "fix" any coherence issues between the old and new systems. Rather, it begins with a deep and thorough understanding of the assessment and accountability systems within the larger educational ecosystem prior to any changes. Comparability is generally established by purposefully planning for it in the system design, evaluating the degree of comparability achieved, and if necessary,

adjusting elements within the systems to account for lack of coherence or logical inconsistencies. Most importantly, the state's role is to engage and communicate with stakeholders about the appropriate interpretations and comparisons that can be made with accountability outcomes[2].

History shows that changes to state educational systems will continue to take place. Maintaining the comparability of accountability outcomes will also be an ongoing demand. We hope that the framework and recommendations described in this brief can helps states establish a sustainable approach or pathway as they navigate through the sea of constant change.

References

Keng, L., & Marion, S. (in press). Comparability of Aggregated Group Scores on the "Same Test". In A. Berman (Ed.), *Comparability Issues in Large-Scale Assessment* by the National Academy of Education.

2 CCSSO has several good resources on stakeholder engagement at https://ccsso.org/resource-library/topic/educator-and-stakeholder-engagement. One particular resource that is helpful in this context is https://ccsso.org/resource-library/lets-continue-conversation.

Council of Chief State School Officers

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