



UNDERSTANDING STUDENT PERFORMANCE ON STATE ASSESSMENTS:

*Leveraging Multiple Sources of State
Data for Strategic Guidance*

May 2025

André A. Rupp & Laura Pinsonneault



Center for
Assessment

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



UNDERSTANDING STUDENT PERFORMANCE ON STATE ASSESSMENTS:

Leveraging Multiple Sources of State Data for Strategic Guidance

The National Center for the Improvement of Educational Assessment, Inc. (the Center for Assessment) is a New Hampshire based not-for-profit (501(c)(3)) corporation. Founded in September 1998, the Center's mission is to improve student learning by partnering with educational leaders to advance effective practices and policies in support of high-quality assessment and accountability systems. The Center for Assessment does this by providing services directly to states, school districts, and partner organizations to support state and district assessment and accountability systems.

This work is licensed under the Creative Commons Attribution 4.0 International License (CC BY). To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

SUGGESTED CITATION:

Rupp, A. A., & Pinsonneault, L., (2025). *Understanding Student Performance on State Assessments: Leveraging Multiple Sources of State Data for Strategic Guidance*, Center for Assessment.

PHOTO CREDIT:

Cover Photo: Allison Shelley/The Verbatim Agency for EDUimages

TABLE OF CONTENTS

SECTION 1 - INTRODUCTION TO FRAMEWORK.....	5
• Motivation for Framework	5
• Framework Components	5
• Framework Development	7
• Organization of Document	8
SECTION 2 - PROFESSIONAL PRACTICES FOR EFFECTIVE DATA INTERROGATION.....	8
• Practice 1: Start Early Through Strategic Planning.....	8
• Practice 2: Engage Experts Across Your Agency.....	9
• Practice 3: Invest in a Culture of Productive Curiosity ...	10
• Practice 4: Ground Interpretations in Local Context Using Mixed-Format Data	10
• Practice 5: Frame Student Outcomes With Multiple Perspectives in Mind	10
SECTION 3 - ESSENTIAL DATA SOURCES	11
• Statistical Analysis Practices	11
- Data Quality Practices	11
- Data Analysis Practices	12
• Source 1 - Population Characteristics.....	13
• Source 2 - Assessment Information.....	14
- State Assessments.....	14
- Interim Assessments.....	14
- College Entrance Exams.....	15
- Educational Surveys	15
• Source 3 - Accountability Information.....	16
- Statewide System	16
- Local Systems.....	17
• Source 4 - Conditions of Student Success.....	17
- State-guided Initiatives	17
- Locally-guided Initiatives.....	18
SECTION 4 - ILLUSTRATIVE USE CASES	19
• Use Case 1: Public State Assessment Results Reporting.....	19
- Report Creation.....	19
- Internal Engagement.....	19
- Implementation Challenges	20



UNDERSTANDING STUDENT
PERFORMANCE ON STATE
ASSESSMENTS:
*Leveraging Multiple Sources of State
Data for Strategic Guidance*

- Use Case 2: Local Outreach on Assessment
and Accountability Results..... 20
 - Context..... 20
 - Data Platform 21
 - Improvement Conversations 21
 - Problem of Practice 1 21
 - Problem of Practice 2 22
- Use Case 3: Internal Strategic Planning to
Support Districts 23
 - Context..... 23
 - Team Structure 23
 - Quantitative Data Interrogations..... 23
 - Qualitative Data Enrichments 24
 - Strategic Planning..... 24

CONCLUDING THOUGHTS 25

EDITORIAL NOTES..... 25

UNDERSTANDING STUDENT PERFORMANCE ON STATE ASSESSMENTS:

Leveraging Multiple Sources of State Data for Strategic Guidance



SECTION 1 - INTRODUCTION TO FRAMEWORK

Motivation for Framework

State education agencies (SEAs) are sitting on rich repositories of quantitative and qualitative assessment data across their various internal divisions or teams. The purpose of this document is to provide an overarching conceptual framework and implementation guidance that can help agency leadership leverage and interrogate these data in systematic ways for reporting, outreach and planning purposes.

The main focus of these data interrogations is typically to understand and contextualize student performance to address important problems of practice. This kind of data interrogation work certainly involves objective description but often takes on exploratory flavors that require iterative, evidence-based reasoning processes. Analysts must draw on multiple variables from diverse sources, adapting their approach based on discoveries made along the way.

Data interrogation work often has explanatory flavors as well, as professionals seek to make sense of the data to craft coherent, compelling narratives for their audiences. In our experience, questions about the “why” of patterns are essentially unavoidable and will inevitably come up, both in internal conversations with senior leadership and externally in public forums.

Principled data interrogation should obviously never be used to “explain away” concerning trends or to highlight only positive results while hiding negative ones. Instead, it should empower SEA leaders and staff by providing trustworthy, contextualized insights that inform a rich understanding of student performance.

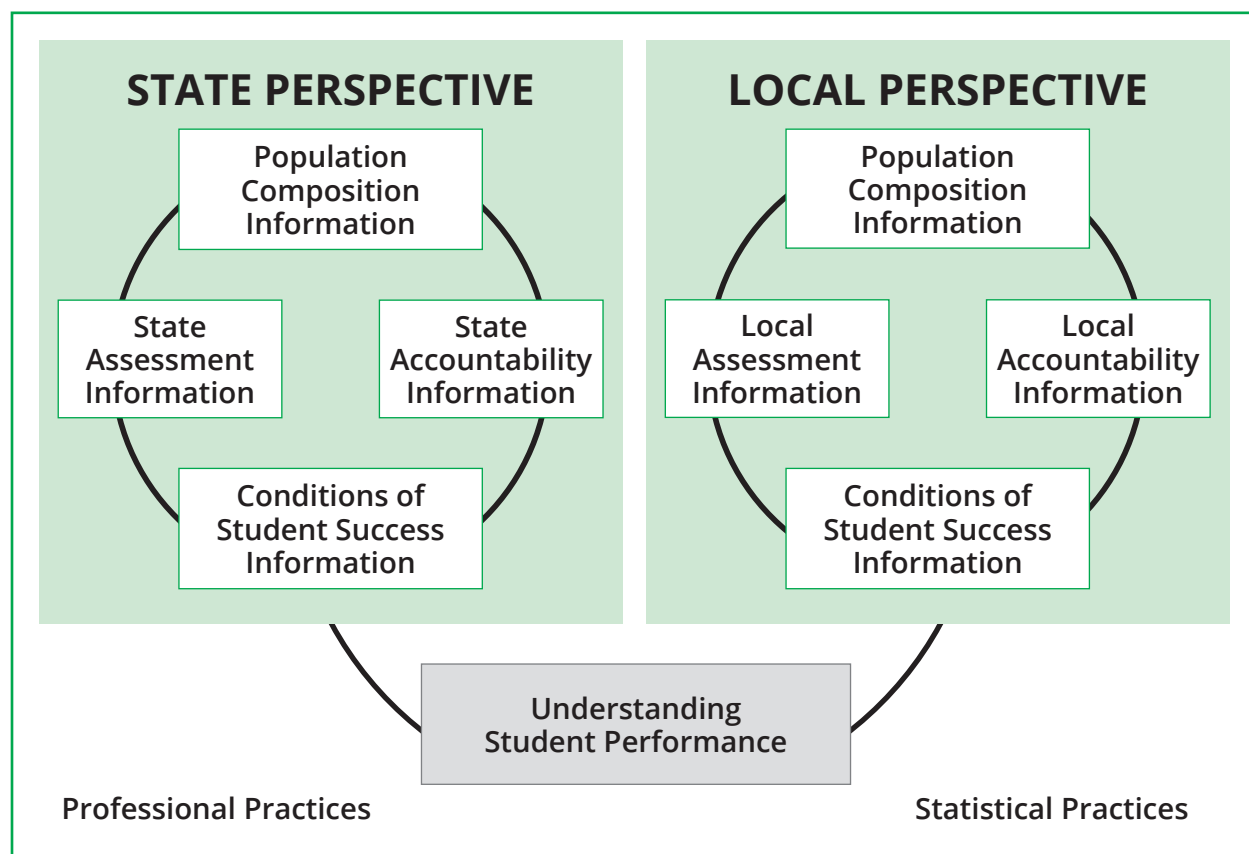
Explanatory data-interrogation work certainly does not support causal claims; such claims require specialized experimental or quasi-experimental research designs and analytic techniques. However, it can move agencies beyond surface-level soundbites and purely descriptive reports.

With thoughtful planning, principled data-interrogation work enables SEAs to foster clear, accurate and constructive conversations—internally, for strategic planning, and externally, with stakeholders such as district leaders, families, advocacy groups and the media.

Framework Components

The framework is anchored in the idea of understanding student performance. As shown through the layers in each of the boxes in Figure 1, this understanding is created by leveraging four core data sources with both state and local components while utilizing professional and statistical best practices. Next, we briefly mention why each source of data can be helpful, and in Section 3, we go into more detail about each source.

Figure 1. Sources of information to aid in interpreting student performance



First, it is valuable to have rich information about important population composition changes. For example, imagine there has been notable mobility in multilingual learner (MLL) populations in the last years in a state. These changes affect the depth and breadth of resources needed to support them across the state, the resulting expectations for performance on state assessments, and the ways in which their group information affects school accountability results. Tracking such information systematically with quantitative and qualitative information can be powerful, as illustrated by [this MLL report](#), [this military students report](#), and this national data portal on [students with disabilities](#).

Second, to understand student performance, it is certainly indispensable to have detailed information about performance on various statewide assessments. These are common measurement instruments that provide consistent, trustworthy and comparable metrics across all districts and schools. Cross-sectional breakdowns and longitudinal trend analyses at various levels of aggregation (e.g., all students vs. student groups, large districts vs. other districts, rural vs. urban vs. suburban districts) using statewide assessment data are critical for monitoring the educational system.

This includes information from regular assessments, assessments designed for learners with the most severe cognitive disabilities such as the Multi-State Alternate Assessment (MSAA) or the Dynamic Learning Maps (DLM) assessment as well as assessments for English language learners (ELLs) such as WIDA Access and WIDA Alternate Access. Information about student performance from educational surveys such as the [National Assessment of Educational Progress](#) (NAEP) as well as statewide and/or major local interim assessments (e.g., iReady, Star, MAP, Navvy) can be helpful as well.

Third, it can also be helpful to leverage core indicators from the statewide accountability system; in general, these are predominantly lagging, rather than leading, indicators of student performance (i.e., indicators that capture outcomes rather than inputs or other predictive information). This information is already shared with the public through a state report card (see this data portal example from [South Carolina](#)) and assessment information is already a core part of the accountability system.

In addition, state accountability systems typically capture additional aspects that help to characterize student performance. This critically includes information about postsecondary readiness (e.g., performance in career and technical education (CTE) courses, dual credit courses, or AP courses) as well as the final student performance indicator, graduation rates. Many systems also include a few valuable leading indicators (e.g., student attendance, disciplinary incidences). Importantly, several additional valuable indicators may be collected in internal databases even though they are not publicly reported.

Fourth, it can be helpful for SEAs to systematically track a broader range of information on educational resources, learning conditions, and experiential opportunities that critically affect student success. State databases may already exist that capture the nature of internships/apprenticeships, out-of-school work programs, and other opportunities (e.g., offerings, uptake, credits/performance) across the state. These may be maintained by teams focused on specialized programs such as CTE rather than the team focused on assessment and/or accountability.

For example, this report from [New Jersey](#) is an example of why a deeper dive into different aspects of CTE matters. Similarly valuable is information about the qualifications of the teacher population across the state (e.g., degrees, certifications, years of experience) or the adoption of high-quality instructional materials (e.g., curriculum adoption status, quality rating, years of use; see [Massachusetts](#), [Louisiana](#)).

Such data sources might even include systematic information about the effectiveness of various educational initiatives (e.g., early literacy or numeracy initiatives). For example, it can be powerful for an SEA to systematically monitor and understand the depth and quality of implementation of these initiatives to develop data-driven, informed hypotheses about likely longitudinal performance trends on statewide assessments for these student populations; see these reports about the work in [Maryland](#) and [Mississippi](#)). We note that some of these reports are commissioned and some are conducted in-house, depending on capacity.

Finally, data-interrogation work with these four core sources at the state and local levels is guided by both best professional and statistical practices. These include ways of working in the agency (e.g., engaging in strategic planning, investing in a culture of curiosity) as well as more technical statistical practices (e.g., ensuring data quality, using multiple measures in conjunction). We discuss the professional practices in Section 2 and the statistical analysis practices at the outset of Section 3.

Framework Development

All descriptions and examples that arise from the framework and the associated implementation guidance in this document were inspired by our own work and expertise, technical documents in the field, and interviews with colleagues working in assessment, accountability and school improvement. We thank our wonderful colleagues for taking the time to discuss earlier drafts with us, share their practices, and offer various inspirational insights; see the editorial notes at the end of this paper for a full list of contributors.

The framework is conceptually situated between two ends of the spectrum for guidance. On the one hand, there are powerful frameworks available that provide high-level walk-throughs of critical organizational practices (see for example this [comprehensive review](#)). On the other hand, there are powerful computational suites available that help with the implementation of the work by data analysts (see for example R or Python environments with their associated packages). Our framework threads the needle by providing mid-level guidance with some concrete examples to complement these resources.

The framework, best practices, and data analysis guidance we present in this document should not be interpreted as a one-size-fits-all approach. Instead, they are intended to be adapted flexibly depending on the use case, the available data, and the SEA's specific priorities. To this end, we would appreciate it if colleagues share new considerations, templates or criteria that they believe should be incorporated into the framework over time. We envision reviewing and updating this document as needed and plan to develop such resources in the future as well.

Organization of Document

In Section 2, we provide a brief summary of the major implementation themes or foundational professional practices that we identified during the creation of the framework. They come from our own professional experience in the K-12 space, industry, and academia, the associated professional literature, and discussions with our colleagues from state agencies.

In Section 3, we walk through the core components of the framework shown in Figure 1 in more detail and provide additional examples of the kinds of variables that are worth considering for data-interrogation work and briefly note how they might be useful.

In Section 4, we present three common use cases in which deeper analysis of student performance patterns and iterative reasoning to explore possible root causes is generally valuable:

- public reporting of state assessment and accountability results
- conversations with school and district leaders around local school improvement
- internal strategic planning for statewide systems and supports

We now turn our attention to the foundational professional practices we identified.

SECTION 2 - PROFESSIONAL PRACTICES FOR EFFECTIVE DATA INTERROGATION

To support meaningful data-interrogation work within SEAs, we outline five foundational practices that integrate both technical and procedural considerations in this section. These practices reflect a blend of strategic planning, organizational capacity-building and intentional culture-setting. Together, they offer a pragmatic scaffolding for agency leaders and cross-functional teams to help with effective data collection, management and utilization.

Practice 1: Start Early Through Strategic Planning

Effective data use begins well before a state assessment is administered or a report is written—thinking ahead in advance of one or even multiple school years across multiple teams is key. This lays the groundwork for thoughtful analysis and coherent messaging when insights need to be

acted upon. As one SEA colleague noted, it is always better to “do the appropriate detective work” than to prematurely declare that “a crime has been committed” when assessment (or accountability) results deviate from expectations.

While we advocate for a comprehensive consideration of various data sources in this document, we are not advocating for a one-size-fits all approach, which would be unrealistic. For example, not all of the data listed in this document need to be brought to every internal or external conversation about student performance. We also recognize that some of the data sources would likely not be available from all districts in the state due to the degrees of freedom that districts have to make local choices (e.g., for curricula or interim assessments) under a local-control paradigm.

Others are also quite cumbersome to collect in the first place. Yet, it can be valuable to invest in a targeted collection of such data for a set of districts that have important contrasting characteristics for an issue at hand (e.g., they are serving different mixtures of MLLs with different kinds of educational programs).

This underscores why it is generally too late to consider which data should be made available for an analysis when that analysis is about to be conducted. Instead, it is important to strategically plan at the beginning of a school year or a particular implementation cycle which data should be systematically collected in the first place.

To help with strategic planning, we recommend asking:

- Which aspects of student performance are most important to understand deeply?
- Which data are most important for understanding student performance trends?
- What local/state capacities are needed/available to collect and analyze these data?
- What legal guardrails facilitate or complicate the collection of these data?
- What levers do we have to act upon any insights garnered from these data analyses?

This kind of planning requires engagement with senior agency leadership, board members, policy makers, and other influential stakeholders who will need to act upon the derived insights. Principled strategic planning also supports improved engagement with a broader set of external stakeholders. For instance, involving members of the media early in the process can foster shared understanding and more constructive public dialogue about student outcomes.

Practice 2: Engage Experts Across Your Agency

Effective interpretation of student outcomes is undoubtedly enriched by diverse, informed perspectives. While the technical staff in assessment and accountability offices are critical, they cannot—and should not—do this work alone.

In most SEAs, offices such as School Improvement, Curriculum and Instruction, Special Education, English Learner Programs, and Early Childhood all have vital contextual knowledge. Although these colleagues work under different timelines, incentives, and constraints, their contributions help ensure that data stories reflect a fuller reality.

When staff across offices understand statewide results, they can better target their supports. Assessment and accountability staff can deepen their understanding of the implementation landscape and tailor their analyses accordingly. This kind of collaboration helps to develop better framings of the “why” behind observed trends and ensures data stories are not siloed or oversimplified.

That said, cross-agency collaboration is clearly not without challenges. Political sensitivities, organizational silos, and public records constraints can discourage transparency. Still, our experience has shown that the benefits of honest internal and external conversations about data far outweigh the risks—especially when handled thoughtfully and with trust.

Practice 3: Invest in a Culture of Productive Curiosity

Related to the previous practice, principled data use depends on an organizational culture that welcomes inquiry. SEA staff should feel encouraged to ask difficult questions, not just highlight successes, and encourage the same in engagement with critical external stakeholders. A culture of curiosity supports reflection, learning and adaptive improvement.

This, again, begins with leadership. When agency leaders model openness, thoughtful questioning, and a focus on student-centered decision-making, they empower others to do the same. Structures also matter. Practices like annual data reflection meetings—with time set aside for noticing patterns, identifying surprises, and raising hypotheses—can reinforce norms that treat data as a tool for learning, not judgment.

The practices we listed in this section are perhaps illustrated best with select use cases, which we describe in Section 4. Before we do that, however, we review the essential data sources in the framework in more detail to further ground those descriptions.

Practice 4: Ground Interpretations in Local Context Using Mixed-Format Data

Student performance cannot be fully understood without reference to the evolving local conditions in which students learn. These include shifts in student and teacher populations, the implementation of new instructional initiatives, changes in assessment participation, and broader social and policy dynamics, to name but a few.

To understand the impact of these changes, SEAs will benefit from drawing on both quantitative and qualitative data sources. This includes administrative data, technical documentation, survey results, interviews, and even logs of implementation-related challenges. Specifically, data interrogation is an [inherently iterative process](#): it often begins qualitatively, moves into quantitative analysis, and returns to qualitative interpretation as teams work toward actionable conclusions. SEA teams that are equipped to navigate this full cycle fluidly and with competence are in a very strong position for actionable, well-informed decision-making.

Some of the richest insights come from mixed-media or "messier" data. These can be burdensome to collect and analyze, although advances in [generative artificial intelligence](#) certainly help. Thus, it is important to prioritize data collection for contrasting contexts (e.g., districts serving different MLL populations) well before the moment an analysis is needed.

Practice 5: Frame Student Outcomes With Multiple Perspectives in Mind

Statewide assessment results are rarely received in a vacuum—they are interpreted through the lenses of multiple audiences. These include district leaders, parents, advocates, journalists and elected officials, each with their own expectations and agendas.

Some stakeholders may seek to spotlight progress and success, while others may focus on performance gaps to advocate for increased attention or resources. In this complex environment, even so-called neutral "just the facts" reporting can be misinterpreted or used out of context.

In our experience, SEA leaders who are able to anticipate this range of responses and prepare for these kinds of challenging conversations are at a decided advantage. Collaborative interpretation

processes—particularly cross-team conversations—allow colleagues to identify and discuss multiple framings in advance. These conversations support a more balanced, thoughtful approach to data storytelling that resists simplistic or overly partisan interpretations.

SECTION 3 - ESSENTIAL DATA SOURCES

We begin this section with a few basic general statistical analysis principles and then walk through each of the data sources in Figure 1. The analysis practices are key examples but should of course be cross-referenced with, and supplemented by, other best practices for data analysis and data-driven sense-making that already exist in an SEA.

We also recognize that SEA resources vary widely. In light of this, we have highlighted foundational, high-leverage practices that can serve as realistic starting points for any SEA team, even those with limited analytic capacity, with a boldface asterisk (*) in the following. We have also sorted the bullet-point lists so that these entries come first.

Statistical Analysis Practices

Data Quality Practices

For each set of analyses associated with the four sources of information in the following sections, a few overarching principles should be considered:

- **Establishing robust data governance policies such as:**
 - aligning practices with legal and ethical standards (e.g., FERPA, state privacy laws)*
 - ensuring the confidentiality of information whenever needed (e.g., student privacy, small cell suppression)*
 - implementing robust security protocols around data (e.g., role-based access, encryption, secure storage)*
 - defining clear roles and responsibilities for data stewards and users*
 - documenting and enforcing data-sharing agreements with external partners
 - setting retention and archival policies for historical datasets
 - conducting regular audits and risk assessments of data systems
- **Verifying the trustworthiness of the data via quality-control checks by:**
 - ensuring accuracy of the received information (e.g., through validation rules or cross-checks with external sources)*
 - ensuring completeness of the received information (e.g., checking for missing fields or submissions)*
 - ensuring reasonableness of the received information (e.g., verifying values fall within expected ranges or patterns)*
 - conducting timeliness checks (e.g., confirming data were submitted within the reporting window)
 - comparing data across years or sources to flag inconsistencies or unexpected changes
 - implementing automated data validation scripts to reduce manual error
 - training staff on data entry standards and error resolution procedures

Data Analysis Practices

Not all of the following data-analysis practices will be possible for every assessment depending on its design and the ways data are shared with SEAs. We discuss these design considerations in the subsection for Source 2 below.

- **Inspecting aggregate performance statistics such as:**

- participation rates*
- means and standard deviations of scale scores*
- percentage of students in each performance level*
- number of students at various percentiles (e.g., 10th, 25th, 50th, 75th, 90th)*
- cohort- and baseline-year analyses of growth (e.g., SGPs)*
- percentage of students below/at/above grade level across time points*
- end-of-year status and within-year growth trajectories*
- accountability indicators related to performance (e.g., graduation rate, postsecondary readiness indicators, chronic absenteeism)*

- **Breaking down performance using structural variables such as:**

- districts (e.g., all districts vs. districts by size, region, or setting)*
- schools (e.g., all schools vs. schools by size, region, or setting)*
- student groups (e.g., all students vs. students by major racial/ethnic groups, disability status, ELL status, economic status, or other demographic variables)*
- grades, grade bands, and school types*
- content areas (e.g., multiple areas combined vs. areas separately)*
- time points (e.g., current year vs. trends over time across years (or within year)*
- combinations of some of the above (e.g., groups over time for schools of different sizes, MLLs with disability status in large districts over time)

Depending on internal capacity, the assessment team at the SEA could also review in more detail the technical information about the assessments, which is typically shared in technical reports from vendors. This could include the following:

- **Inspecting item and scale quality information such as:**

- reliability, measurement error, and information across the scale*
- equating and concordance information*
- difficulty, discrimination, and pseudo-guessing statistics
- differential item, bundle, or test functioning statistics
- item drift statistics
- other global or local model-data fit statistics
- prevalence of unusual responses (e.g., skipped or nonsense responses)
- prevalence of blank responses that receive 0s (e.g., true blank, insufficient response, critical response)

- **Reviewing assessment designs for changes in key design aspects such as:**

- skill and competency (“construct”) definitions*
- item and task types (e.g., innovations in technology-enhanced items)*
- scoring approaches (e.g., human and automated scoring)*
- performance level descriptors, cut scores, and reporting categories
- evidence about form comparability
- evidence about validity, accessibility, and fairness

We discuss the specific data sources listed in Figure 1 next.

Source 1 - Population Characteristics

Understanding key changes in the composition of the student population across years is important since these changes might affect interpretations about notable changes in assessment performance. Insofar as population changes also affect the relative size of particular student groups, they can impact the accuracy and trustworthiness of group comparisons and the resulting understanding of any single group's trends.

In addition to standard demographic student groups, changes in the following student groups should be tracked in particular:

- learners whose first/home language is not English (e.g., ELLs, MLLs)*
- learners with severe cognitive disabilities*
- learners in special education programs*
- learners who are experiencing unstable living situations (e.g., homelessness, migrancy)*
- learners with high mobility, either within the state or in-and-out of the state*
- learners in gifted or talented programs
- learners experiencing disciplinary actions (e.g., suspension, expulsion)
- learners in non-traditional learning contexts, including the juvenile justice system

In addition to doing this for each of these groups in the aggregate, it can also be meaningful to break them down further. For example, one of our partners noted that their SEA tracks not just ELLs but also various subcategories that include students who were ELLs at any point regardless of their current status, current ELLs, ELLs within 2 years of exiting status, and ELLs in their first year of being assessed on the state proficiency assessment.

This could also include [intersectionality analyses](#) in which multiple demographic variables are combined. For example, one could review performance information for ELLs with disability status or ELLs who are experiencing economic hardship, depending on what is known about the challenges these populations face in a state at the moment and the educational programs that are currently implemented statewide to support them. Moreover, the theoretical framings of the underlying constructs have important implications for the interpretations of the resulting model parameters and statistical effect size estimates (see, e.g., chapter 12 in [this book](#)).

As we noted in the first subsection, it is helpful to compute descriptive statistics and create visualizations of the trends in population composition at the core structural levels of the data on an annual basis. For example, analyses of academic achievement can be done globally, by student group, by district, for students with different rates of absenteeism, and for districts or schools with varying levels of test participation, among others.

We caution against a mechanistic approach in which all kinds of descriptives are computed without any particular attempt at more informed sense-making based on data interrogation work. This can have negative consequences, such as when certain student groups are inadvertently blamed for undesirable state assessment performance once their performance patterns are reviewed.

In essence, each analysis should be done with a thoughtful set of hypotheses around the analysis in mind. This idea is aligned with best practices in the social sciences and related disciplinary areas. Put simply, having well-formed hypotheses at the outset—or forming and refining these iteratively through data interrogation—is important for meaningful data interpretation.

Source 2 - Assessment Information

State Assessments

The main source for understanding the current and historical state of student performance is, obviously, state assessment data. We use the term “state assessment” to refer to all statewide administered assessments. Thus, we include the state assessments for core content areas, which are predominantly ELA, mathematics and science as well as, in certain states, social studies or civics. This also includes alternate assessments for students with the most severe cognitive disabilities and assessments of English language proficiency (ELP).

As the Center discussed in a recent [position paper](#), state assessments undergo highly rigorous technical quality evaluations on a variety of dimensions (e.g., validity, reliability, fairness, accessibility, comparability) and are thus a valuable source of information for understanding student performance across the state using a common instrument. More specifically, beyond their monitoring function, they support continuous improvement work, encourage transparency about the state of learning and public engagement around it, and signal rich learning expectations.

Interim Assessments

Statewide Interim Assessments.

If a single statewide interim or through-year assessment is being administered, the reports from this assessment will contain information about student performance throughout the year and at the end of the year. Note that the availability of information throughout the year does not automatically determine how it is used for summative determinations at the end of the school year for accountability purposes, which is a separate issue.

An important consideration in the (potential) analysis of interim data is their design. For example, interim systems designed to maximize instructional utility will have different designs than interim systems with a through-course structure in which the computation of a year-end summative score for accountability purposes is one explicit intention. The Center’s recent paper on [through-year assessments](#) illustrates several of these points.

District-selected Interim Assessments.

Common district-selected assessments in core content areas such as ELA and math include [Acadience](#), Curriculum Associates’ [iReady](#) suite, NWEA’s [MAP](#), and Renaissance’s [STAR](#). Moreover, specialized integrated learning and assessment platforms like Apprendis’ [Inq-ITS](#), Pearson’s [Navvy](#), or other curriculum-alignable assessment suites can also provide relevant data if used at scale in the state.

However, unlike statewide interim assessments, these systems are not necessarily aligned to the state standards, thus limiting direct comparability of information. Nevertheless, with a thoughtful review of the design of the local assessments vis-à-vis the state assessments, it is possible to identify valuable directional signals in their resulting data.

Information about the relationship between interim performance and state assessment performance may be available from vendors in specialized reports or may have to be developed internally at the SEA if capacity is available.

Districts can choose which of these systems to make available to their schools. As a result, there is typically notable variability in which systems are used for which content areas and grades across the state. Therefore, it can be helpful to have information about:

- which districts use which assessment suites
- which grades and content areas these are used in
- their predominant approach to using them (e.g., formatively, summatively)

We recognize that the collection of this information at scale is likely complicated and burdensome, but could perhaps be advisable for districts that receive special analytic attention (e.g., those with schools that are designated for improvement and support or those with schools that excel notably).

College Entrance Exams

College entrance exams are standardized assessments used by postsecondary institutions to evaluate applicants' academic readiness and compare students across different schools and regions; the two most common exams in the United States are the SAT and ACT. They generally carry important stakes for individual students as well as schools in accountability contexts, although some students take them for accountability purposes and do not use them further for their own postsecondary education.

By design, these exams are not directly aligned to state content standards, but are aligned to the vendors' internal college- and career-readiness standards. These, in turn, are influenced by other major standards in the field, such as the Common Core State Standards, to some degree, although they include unique components. As a result, reviewing the design information for these assessments is important for properly interpreting student performance.

As a result of these factors, direct comparisons to state assessment results are not warranted, but they can provide valuable additional insight about postsecondary readiness. They can be useful for directional trend comparisons in related content areas, as long as differences in assessment design, student populations, and test-taking conditions are acknowledged when interpretations are made.

Educational Surveys

The major national educational survey is NAEP, which provides important information about aggregate student performance over time, predominantly in ELA and math. Importantly, scores are available only for groups of students, not for individual students.

NAEP is not designed to measure each state's content standards, but is based on an assessment framework adopted by the [National Assessment Governing Board \(NAGB\)](#). Historically speaking, NAGB has provided very strong oversight for NAEP, ensuring that results from the test are scientifically trustworthy given the goals of the survey.

Moreover, the sampling design and associated statistical machinery underlying NAEP can help distinguish random year-to-year fluctuations in data or random differences in student group performance from systematic changes via statistical significance tests. Information about the technical properties of NAEP can be found in a variety of [technical reports and manuals](#).

NAEP can be a helpful high-level contextual indicator of student achievement, particularly when it comes to trends in student achievement over time across all states. As with interim assessments and college readiness exams, direct comparisons to state assessment results are not warranted but, directionally, NAEP can offer valuable additional insight. If a state includes a district that is participating in the [Trial Urban District Assessment \(TUDA\)](#), this information can be used to dig further into the performance relationships across assessments for that district.

Schools may also participate in other national or international surveys that allow states to contextualize state-level performance. Examples include:

- [Third International Math and Science Study \(TIMSS\)](#) [math, science; 4th and 8th grade]
- [Program for International Student Assessment \(PISA\)](#) [reading, math, science, 15-year olds]
- [Progress in International Reading Literacy Study \(PIRLS\)](#) [reading, 4th grade]
- [Programme for the International Assessment of Adult Competencies \(PIAAC\)](#) [adult competencies for workforce readiness, 16-65 year olds]
- [International Computer and Information Literacy Study \(ICILS\)](#) [computer and information literacy, 8th-grade]

For more information on a range of surveys that may be relevant, see the web portal of the [National Center for Education Statistics \(NCES\)](#). While the educational surveys are likely not the primary sources for annual data analyses and their data-collection cycles and participation rates differ, they can be valuable supplementary information sources to consider in moments.

Source 3 - Accountability Information

Statewide System

Statewide accountability systems fulfil critical roles in a state that include:

- monitoring educational outcomes across a state using common metrics
- engaging with school and district leaders in evidence-based school improvement conversations
- providing families and community members with actionable information about school quality
- signaling the most critical educational priorities

See this recent [Center position paper](#) for a more detailed discussion about the value of statewide accountability systems.

A large portion of state accountability systems is composed of state assessment information (e.g., achievement and growth in ELA, math, and science, including on alternative assessments; progress in ELP). In addition, the following information may be included in accountability systems:

- graduation rates*
- dropout rates*
- measures of student and teacher engagement and sense of belonging (often surveys)*
- measures of school climate and safety (often surveys)
- core measures of readiness (e.g., completions, credits, certifications, recognitions) for*:
 - advanced coursework (e.g., AP and IB courses)
 - career and technical education (CTE) courses and pathways
 - dual enrollment courses
 - visual and performing arts, physical education, world languages, and other enrichment courses
 - apprenticeship programs, internships, or other types of workplace-based learning
- measures of other aspects of postsecondary readiness such as students:
 - participating in job fairs
 - enrolling in colleges of choice
 - finding employment in desired professions
 - earning an Advanced International Certificate of Education (AICE) diploma
 - completing assessments of military readiness (e.g., the Armed Services Vocational Aptitude Battery (ASVAB))
 - completing assessments of postsecondary workforce readiness (e.g., WorkKeys, Worldwide Interactive Network National Career Readiness Certificate (Win NCRC))

Importantly, while the rules for how the information of these component indicators are combined to create aggregate indicators and school designation classifications can often be complex, the individual indicators are first and foremost valuable as independent sources of information for understanding student performance. The degree to which these are relevant for particular data interrogations depends again on the research questions / analytic goals that drive these analyses.

Local Systems

In local accountability systems, additional indicators may be collected that may be helpful when certain districts are analyzed more closely as use cases (i.e., “deep-dive analyses”). These may include:

- indicators of opportunities to engage in deeper or vibrant learning
- completion rates for cumulative student presentations / defenses of learning
- student involvement in extracurricular enrichment activities
- student involvement in community service or civic projects
- student representation on school committees
- student participation in peer mentoring programs
- student and parental/family attendance at school events
- assessments or evaluations of socio-emotional skill development (often surveys)
- assessments of digital and information technology literacy skills

As with the state system indicators, the relevance and potential utility of these indicators varies based on the nature of the research question/analytic goal.

Source 4 - Conditions of Student Success

Most of the information in the previous section was anchored in the perspective of student performance (e.g., how well students are performing on assessments, whether they complete courses, attend school, or graduate). In this section, we discuss how the fundamental school quality work by schools and districts creates conditions of learning that empower students for success in the first place.

State-guided Initiatives

SEAs are often engaged in specific initiatives that have indirect effects on state assessment outcomes and may thus be worth monitoring. These include:

- improving access to high-quality instructional materials*
- improving teacher pre-service and professional development training*
- increasing the focus on deeper, student-centered learning*
- increasing support for the development of social-emotional skills*
- increasing outreach to parents/caregivers to support student learning
- changing policies that affect classroom practice, including assessment practices

The degree of impact would depend on the spread of implementation, the length of implementation, and the alignment considerations listed above. Again, we underscore that such connections are not causal arguments without suitable experimental or quasi-experimental designs and analytic techniques.

SEAs may also provide general direction, guidance, and professional development for targeted programs that are intended to boost learning, either directly or indirectly, by affecting the conditions that make successful learning possible. For example, general instructional guidance and/or specific programs may be focused on areas such as early literacy and reading, STEM education and enrichment, social-emotional learning (SEL), college and career readiness, and CTE pathways.

Therefore, it is generally helpful to be able to understand the answers to a few key implementation questions:

- Which major innovative educational programs are implemented in the state?*
- What are their expected impacts on state assessment performance?*
- How long and with what intensity have they been implemented in different districts?*
- What is known about their effects, beyond potential effects on assessment results?

For example, if an SEA has made a statewide push to support early literacy and/or numeracy skill development, and many districts have implemented them successfully, then it might be reasonable to expect that the state will eventually see some gains in reading scores in earlier grades and for certain student groups in particular.

Alternatively, an SEA may have pushed for innovation in the area of deeper, more personalized learning. They may have created a portrait of a learner with performance outcomes and indicators and have supported districts in various initiatives for transforming learning practices in classrooms; see [this report](#) as an example. These practices might be designed to change familiarity and engagement with more complex, authentic tasks and would thus change opportunities to learn and transfer the learnt skills to assessment contexts.

Locally-guided Initiatives

While it is not the purview of the state to have access to information on all local initiatives, there are certain situations in which it pays off to systematically collect such information. As illustrated by the third use case in Section 4, this information is often heavily qualitative and developed through resource-intensive efforts that include:

- interviews with individual school leaders*
- focus groups with teachers, students, and/or families/caregivers*
- classroom observations and reviews of teacher materials*
- independent evaluations of student work
- systematic reviews of strategic plans for multi-tiered support

This information is often collected in more depth exclusively—or at least predominantly—for schools that were identified for support and improvement through the federal accountability system. Arguably, some of the same ideas for data collection could also be used for schools that are performing better to celebrate their success, share lessons learned, and help everyone better understand local drivers of differences in student performance.

In addition to considerations related to the needed investments of resources for this work, a better understanding of these local implementations typically requires collaboration across several departments or teams within an SEA as well as a close collaboration with district leads and concerned data-collection efforts throughout the year. This is most effective within a broader culture centered on mutual learning, trust between the state agency and local districts, and a desire to innovate.

In the next section we discuss three use cases that illustrate how the cross-cutting data analysis practices at the outset of this section and the four sources of information can be used to generate actionable insights for pressing problems of practice.

SECTION 4 - ILLUSTRATIVE USE CASES

In this section we present three use cases that illustrate different ways in which some of the data sources laid out in the previous sections come together to support deeper exploration and (at least initial) potential explanations of observed patterns and trends.

Use Case 1: Public State Assessment Results Reporting

This is a common annual task for SEA teams. Practically speaking, this can be internally framed as the task of creating a compact executive report for state leadership and communications teams that lays out the key take-aways from the assessment results.

Report Creation

This report likely takes form as a presentation/conversation and may be informed by any of the following:

- relevant data tables and graphic visualizations
- a summary memo or paper
- talking points for engaging with the press or stakeholders
- slide decks comprising key points and data

Steps to prepare data for reporting can be standardized so that, with suitable data governance, architecture, and engineering practices in place, reports can be semi-automated. This work can be done in modern computational environments such as GitHub with programming languages such as Python, R, or SAS.

Moreover, if slide decks are used, the repeating components/layout features of these slide decks can be captured in a master deck that gets updated every year with new data and resulting descriptions of interpretations. Importantly, this approach is intended to create efficiencies that allow for deeper and richer data discussions, not to facilitate a plug-and-play mindset without a renewed critical eye.

Internal Engagement

As with any complex story, simply creating such a report or slide deck and handing it over to SEA leaders or communication teams is not sufficient. Instead, in-depth conversations are necessary to untangle some possibly confusing and seemingly contradictory facts. This helps everyone to:

- understand and be able to describe the current outcomes and trends
- engage in informed and appropriate discussion about how the state will use the results and other data to inform decisions about supports or guidance it will provide
- develop materials that will help school and district leadership talk about their results (e.g., sample talking points, sample family letters, sample press releases)

SEA leadership may find it helpful to anticipate different potential framings and follow-up questions that can be asked around these data. They have a critical role in helping frame how improved performance can be a celebration and a call to action at the same time.

Beyond these internal conversations, it can also be helpful to onboard members of the media earlier on in the reporting process to build trust and allow them to better understand some of the trends and their possible explanations. This can help to reconcile possibly conflicting narratives in advance.

Implementation Challenges

In talking to colleagues about these processes, they repeatedly mentioned some challenges that affect these processes:

- tight timelines between having frozen/verified data from all districts, internal leadership conversations (e.g., less than a week), and public reporting events (e.g., less than a month)
- missing or non-existent data on the quality of implementation of important initiatives across districts, especially those that are state-provided but not mandated (e.g., early literacy, early numeracy, social-emotional learning support)
- competing priorities of internal staff who could conduct more in-depth analyses but are requested to do more mundane data management and associated communication tasks
- multiple communication processes with distinct audiences that require coherent messaging at a higher level but are overseen by different agencies within the SEA (e.g., SEA school improvement teams working with school improvement teams at districts, communications office teams working with the press, assessment and accountability teams working with various audiences)

Advanced planning and longer-term capacity building by SEA leadership are needed to foresee and address these challenges.

Use Case 2: Local Outreach on Assessment and Accountability Results

This use case focuses on the individual or small-team conversations between SEA staff and school leaders to review the state report card results and discuss possible contextual explanations and strategies for local school improvement.

These conversations commonly include at least the schools that were identified for improvement and support based on results of the federal accountability system. However, if SEA capacity allows, they could be extended to include all districts in the state or at least additional districts or schools based on prioritized state criteria.

These conversations are also informed by accreditation-style reports for these schools that are often created by vendors. These provide qualitative and quantitative snapshots of each school based on interviews, focus groups, classroom observations, surveys, and the like. This information helps paint a much more detailed picture of the realities on the ground. Similarly, local leaders will bring to these conversations their own dashboards, relevant reports, or staff that can help them fill in additional pieces of information as-needed in the moment.

Context

The work in this use case was led by an office of school and district improvement within an SEA. The work was situated within a broader strategic effort to better understand what data are available that the SEA can use in conversations with district and school leaders to improve their local school improvement work.

Conversations had historically been framed in terms of “how to use accountability results to improve schooling,” rather than anchoring them in a norm-referenced comparison with other districts or schools. This framing had set the tone for the districts to continually leverage data to be strategic in their planning and monitoring efforts.

These conversations were part of the SEA’s broader “roadshow” on annual data. Conversations with local leaders and teams were customized to the specific informational needs of individual districts or schools.

Data Platform

In this example, all districts maintained local accountability systems, including strategic improvement plans. Importantly, all strategic plans for both schools and districts were brought together in a robust, integrated strategic planning platform.

SEA staff had access to all of these plans at all times, supported by state legislation. The platform also allowed customization to include specific federal and state reporting requirements for schools with formal improvement designations as well as information about the implementation of state initiatives.

The strategic planning platform also pulled in all relevant longitudinal data for assessment and accountability indicators, including contextual indicators such as attendance and participation rates. On the backend of this system was an integrated data warehouse.

For example, if a district was setting a goal for increases in scores on the SAT, the system included relevant historical data on SAT performance, allowing users to interrogate how realistic those goals were and to see whether past goals had been met. This included various visualization capabilities to make working with these data easier.

Improvement Conversations

Officers at the SEA reviewed the state assessment and accountability results with local school improvement teams, principals, and other relevant leaders, with a focus on trends over time at different levels of the system. These trends were then interrogated and brought into connection with the patterns of the main contextual indicators such as attendance, participation, and so on.

Areas of particular concern were then identified and captured in the district's strategic plan. Districts were encouraged to monitor progress with indicators that were locally meaningful to them along with any relevant state indicators. The use of various data sources conjunctively to gain additional insight had proven powerful before in many instances.

These conversations were required for schools within the most common designations for support and improvement but the SEA had an open-door policy, capacity allowing, for other local districts or schools to reach out to them as well. So far, the SEA has never had to say no to these requests and it has made about 100 such visits during the last two years.

Importantly, the SEA team had received requests not just from schools or districts that are struggling more broadly, but also those with overall strong student performance and more isolated areas of weakness.

The SEA team typically brought specialists from the assessment and accountability teams, content teams, special education teams, and/or other teams with relevant expertise with them. Internally, there had also been other forms of professional development to support this work.

For example, the assessment team had done item-level walkthroughs of state assessment forms to help the school improvement team explain those aspects better to local leaders. While this was resource-intensive work, the participating SEA teams as well as the local leadership teams very much felt that it was worth it.

Problem of Practice 1

In this specific example, one of the schools noted that a particular group of students within a particular race/ethnicity category was struggling in one of the core content areas. One of the

teachers hypothesized that this was likely due to the fact that students in this group had much higher absenteeism rates than students in other groups. However, the data showed that this was not true. In other words, it appeared that factors unrelated to attendance—rather than not having access to instruction—should be examined.

The team then looked at whether there was an abnormally high proportion of students in that group who were accessing highly targeted, intensive tier-III interventions. This included considerations of whether students were over- or under-identified for such interventions as well as whether these interventions were working effectively.

However, not a lot of students from that group had been receiving tier-III interventions. Alternatively, it was also possible that tier-I or tier-II intervention efforts had been suboptimal. To disentangle those issues, the team looked at the beginning-of-the-year benchmarking assessments to see how the students were performing then. If those that needed support were appropriately identified for tier-III interventions, then another reasonable inquiry might have focused on the effectiveness of tier-III intervention strategies. This could have included insufficient accommodations, for instance.

An additional in-depth analysis of the intervention revealed that this was, in fact, not the issue either. Instead, one of the issues that this analysis brought to the fore was the lack of cultural relevance of the curricular materials that led to reduced engagement with the materials for these students in particular. Plans were then put in place to revise the curricular materials to address those limitations.

Importantly, in this example, the data that were inspected included both district and individual school data. At the district level there were too many students included in this group to be able to do an individual, case-by-case review. However, at the school level, the local teachers and principals would be familiar with the individual students as there were fewer of them overall. This allowed for a more targeted analysis of their situations and how to address them.

Problem of Practice 2

In this example, SEA leadership was trying to understand how to better serve their MLL population. To start, the SEA team pulled data from all districts that had identified this group of learners as a focal group of interest for support based on historical performance that had been significantly below expectations.

The SEA leaders worked with ELL specialists to develop a series of professional development modules for district leaders to improve their curriculum planning, instructional models, and local teacher professional development approaches to better serve this population. This included training resources targeted at professionals in very specific roles (e.g., school principals). The local districts then updated their improvement plans to include the state-provided and state-suggested services.

This work was supported through federal Elementary and Secondary School Emergency Relief (ESSER) funds at the time. Similar work had been done to support students with disabilities and address issues of attendance, for example.

In addition, the SEA also worked with a vendor to visit local schools most critically in need of support for more in-depth, mixed-methods analyses based on classroom observations, focus groups, interviews, and the like. This had uncovered a need for systematic improvement in the area of curriculum implementation, for example.

This second use case combines ideas about internal strategic planning at an SEA with resulting district engagement. The third use case in the following section fleshes out internal planning work in a bit more detail.

Use Case 3: Internal Strategic Planning to Support Districts

This use case focuses specifically on developing strategies at the SEA that can guide the engagement with LEAs around school improvement efforts.

Context

The release of an annual state report card made it clear to SEA leadership that math achievement and growth continued to lag behind both the developmental trends in other states and general expectations for performance in the state itself.

As a result, an interdisciplinary, cross-office team was put together to iteratively interrogate the data and better understand the variation across districts and possible variables/factors impacting statewide performance.

From this work, the team made strategic recommendations for program improvement and followed certain districts more closely to better understand their lived reality, how the suggested supports were taken up, and what intended and unintended effects of these efforts were.

Team Structure

The team that was assembled included about 10 people:

- state superintendent and deputy
- division administrators
- academic officer
- accountability director
- assessment director
- content and learning team director for mathematics
- research and data science lead
- math specialist/consultant
- administrative staff

A series of three to four monthly conversations took place. The data work was mostly descriptive with no particular formal modeling at that point (e.g., no regression-type, classification, or multivariate models).

Quantitative Data Interrogations

The initial data interrogations were driven by an exploratory slicing-and-dicing of the data using the commonly available structural variables (e.g., region ID, district ID, district location, district size, student demographic variables, grades, school type) to identify places where performance was particularly strong or troublesome.

This included excluding the largest district in the state, which was large enough to shift the statewide performance distribution. Importantly, data included both current-year data and as much historical data as was available for each indicator at each relevant structural level.

This part of the work helped to iteratively identify places where to shine the data spotlight next and to dig deeper. The nature of such follow-up explorations was driven by the kinds of patterns that emerged and the hypotheses that were formed (e.g., identifying certain issues as racial issues vs.

regional issues vs. global issues helped to sharpen questions and affected the nature of potential strategies for addressing the issue).

These data interrogations also served as an important internal level-setting exercise to ensure consensus that there were indeed systematic, noteworthy areas of concern in mathematics (e.g., that various observed differences were statistically significant and not merely within expected bounds of random year-to-year fluctuations).

The analyses also helped to uncover a few incorrect presumptions that staff made and to understand, in many cases, that issues were not necessarily attributable in simple ways to districts or schools with particular structural characteristics. This helped to empower all members of the interdisciplinary team in their messaging efforts.

Data were presented in multiple ways that included pre-developed slide decks and handouts with relevant breakdowns that were shared in advance of the meetings with all members. Moreover, the research and data scientist lead pulled up relevant breakdowns from their common database during the meetings as new questions were generated by the team through conversations.

Qualitative Data Enrichments

During these conversations, team members shared what they knew about key factors that were likely to affect mathematics instruction and student performance. These included:

- transition to new state standards in mathematics
- the increased prominence of AI-supported tools for problem-solving in classrooms
- new performance tasks in various high school math courses
- new instructional practices grounded in deeper learning through innovation initiatives
- information from statewide surveys and focus groups managed by education centers
- independent vendor reports about school quality for CSI/TSI/ATSI schools
- anecdotal information from select schools and districts
- anecdotal information from SEA colleagues

Results from a statewide survey on the use of instructional materials, informed by guidance by EdReports, were used as part of this work.

Care was taken not to over-interpret the anecdotal information and to be aware of gaps or biases in the data that were not collected for all districts (e.g., looking at which districts did not participate, looking at participation by demographics).

Strategic Planning

The discussion referenced current policies to facilitate understanding of the levers the state would be able to pull for support (e.g., waivers). This discussion extended beyond formal policies and included reflections on the kinds of initiatives—resources, training, data support—that the state could put into place given its primary role, how it could collaborate with regional education centers to bring these resources to districts, and how it could incentivize participation in such initiatives.

This led to the eventual creation of stronger initiatives on high-quality instructional materials. The SEA teams were also better able to understand, and support, proper interpretations and uses of various types of data locally. It also helped to identify additional research questions that the SEA teams wanted to pursue in the future with the support of their academic partners.

Finally, this work also led to new practices at the SEA, which speaks to the impact of this work on the human systems at play. For example, members of the assessment and accountability teams started

to accompany the school improvement teams on at least one annual site visit to get a better, richer sense of the realities on the ground and the nature of conversations that were being held. This included secure access to internal reports and professional development to better understand how different resources were used.

CONCLUDING THOUGHTS

As these three use cases have shown, engaging in the five overarching practices named in Section 2 and the data-analysis practices named at the outset of Section 3 with the data in the framework can empower SEA leadership in effective, meaningful internal planning as well as external outreach efforts. There is certainly no one-size-fits-all approach that works best under all circumstances, but the examples in Section 3 and the use cases in Section 4 demonstrate their value again and again based on our experience.

We sincerely hope that this document contained a few helpful ideas and inspirations to further refine your work. If you have additional best practices or use case examples that you would like to share with us that could further enrich this document in the future, please do not hesitate to reach out. We would love to consider them for incorporation into a future revision of this document.

EDITORIAL NOTES

Several SEA representatives provided invaluable formative feedback, use cases, and otherwise inspirational thoughts:

- Angela Foxhall, Nick Heckel, and Nikki Combs, IL Department of Education
- Phyllis Lynch and Kristen Danusis, RI Department of Education
- Zachary Warner, NY Department of Education

We cannot thank them enough for their support of this work. In addition, special thanks to Chris Domaleski for providing detailed feedback on an earlier version of the document and to Scott Marion for initiating the creation of this document in the first place.



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

www.nciea.org