IMPLEMENTING BALANCED ASSESSMENT SYSTEMS:

A Practical Guidebook for Districts & Schools

Center for Assessment

NATIONAL ACADEMY of EDUCATION

Implementing Balanced Assessment Systems: A Practical Guidebook for Districts and Schools

Table of Contents

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. For more on the Center for Assessment, visit www.nciea.org.

The **National Academy of Education (NAEd)** advances high-quality research to improve education policy and practice. Founded in 1965, the NAEd consists of U.S. members and international associates who are elected on the basis of scholarship related to education. The Academy undertakes research studies to address pressing educational issues and administers professional development fellowship programs to enhance the preparation of the next generation of education scholars. For more on the NAEd, visit www.naeducation.org.

Digital Object Identifier: 10.31094/2025/3

Suggested Citation: Evans, C. M., Wylie, E. C., Landl, E., & Marion, S. F. (2025). *Implementing Balanced Assessment Systems: A Practical Guidebook for Districts and Schools*. The National Center for the Improvement of Educational Assessment & The National Academy of Education. https://doi.org/10.31094/2025/3

Acknowledgements: We want to acknowledge the thoughtful feedback provided by those who attended the Center's annual convening, the Reidy Interactive Learning Series, in the fall of 2025. The topic was "What It Takes to Implement Balanced Assessment Systems" and participants read a draft version of this Practical Guidebook and provided feedback. We are grateful for the insights shared that made this Guidebook better. We also appreciate our Editorial Director, Catherine Gewertz, who helped improve the readability of this Guidebook.

SECTION 1: WHY & HOW TO READ THIS PRACTICAL GUIDEBOOK

We know that assessment in schools can have a significant impact on students. When done well, assessment can provide meaningful feedback for students and teachers—an important element of supporting ambitious instruction and deeper learning for students. Unfortunately, current practice does not always support such rich visions of assessment and learning.

Purpose of This Guidebook

This Guidebook is designed to help school and district leaders evaluate and improve their existing assessment system by addressing one or more of the following questions in Table 1-1.

TABLE 1-1 List of Questions Addressed in this Guidebook

Coherence	• Is our assessment strategy and approach aligned to and coherent with the research about how students learn and gain more complex and sophisticated disciplinary knowledge, skills, and understandings? • If an outsider were to look at our assessments, what might they surmise about our goals for education, how students best demonstrate learning, and the types of information needed to help schools, educators, and students understand and improve teaching and student learning? • Are our locally controlled assessments supporting our goals for learning, curriculum, and instruction, or hindering our goals?
Usefulness	• What types of information do our teachers, specialists, leaders, parents, students, and school board members need about student academic achievement and growth, when do they need that information, and for what purpose? • Are educators and other interest holders getting the information they need, when they need it, and is it of the type that is useful for the decisions they make?
Efficiency	• Why are we assessing students the way we are? Do we have a clear strategy for how, what, and when we assess student learning? • Do we need all of the assessments currently in place?

These questions probe how students are assessed over the course of a school year and the extent to which that collection of assessments works with curriculum and instruction to reflect what we know from the research about how students learn and develop more sophisticated ways of knowing over time. If that set of local assessments—from the classroom to the district level—are thoughtfully selected and designed, they should work together as a coherent *system*.

A *system* is a collection of interacting components or parts that work together to accomplish a particular goal or function. A key distinction between a system and a simple collection of assessments is the intentional linkages and interdependencies between and among the parts. Each assessment should have a clear function or purpose that contributes to the overall coherence, efficiency, and usefulness of the system as a whole.

A balanced system of assessments is one that is coherently designed to support curriculum and instruction through a common model of learning and to provide a comprehensive understanding of what students know and can do.[1] Balance, in this context, is meant to reflect a dynamic state where each user has the information they need from assessments to make better educational decisions that support student learning. Balance is also intended to signal that the primary function of assessment in the educational system should be to support teaching and learning, not to feed external accountability demands.

With this in mind, this Guidebook can help school and district leaders address three additional questions:

- How balanced is our local assessment system and how would we know?
- What might be out of balance with our local system of assessments, given our instructional vision?
- How could we begin to improve the balance of our local assessment system?

Our Goals

We wrote this Guidebook to serve as a companion to the National Academy of Education's 2024 volume, *Reimagining Balanced Assessment Systems*.[2] Our goal was to start translating this research into practical support and guidance for district and school leaders as they engage in this vital work. As we have engaged in balanced assessment design, implementation, and evaluation over the past 20 years, we keep returning to some fundamental questions. Most importantly: do the assessments intended to support student learning and ambitious teaching practices actually do so? If they are not used for these purposes, are they used for other critical purposes in the education system without getting in the way of student learning and high-quality teaching? Do we have evidence to support claims that the assessments in the system are serving defined uses as well as possible while avoiding negative consequences?

[1] National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. https://doi.org/10.17226/10019

[2] Marion, S. F., Pellegrino, J. W., & Berman, A. I. (Eds.). (2024). *Reimagining balanced assessment systems*. National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems-project/

Having a clear, research-based understanding of how students learn and develop makes it challenging to design assessments to provide a coherent picture of student learning from the statehouse to the classroom. In fact, many have argued that is too ambitious of a goal and we should be focusing our attention on school districts.

A major challenge to creating balanced assessment systems has been the lack of shared visions of student learning beyond merely pointing at the content standards. The standards are necessary, but insufficient for creating this vision because they define only the endpoints of what students are expected to learn. Thus, our goals in writing this Guidebook are to provide tools and tangible examples to help district, school, and state leaders design, implement, and support balanced systems of assessment by better understanding the most common threats to balance.

We recognize that there are many other aspects of schooling, not to mention policies and structures, that have an impact on the decisions made about assessment in schools. We are not able to discuss them all here, but our goal is to start translating research into practice, and we think we have accomplished that goal.

Suggestions for Where to Focus Your Reading

Section 2**: **Background on Balanced Assessment Systems provides a brief history of the research on and the criteria for balanced assessment systems, as well as barriers to design and implementation noted in the previous research literature. If you already have a strong background in balanced assessment systems or you just want to get to work using the Guidebook, we recommend you skip Section 2 and start with Section 3.

***Section 3*:** The central ideas of this Guidebook are in ***Section 3: Threats to Balance & Practical Next Steps*.** **Everyone should read this section.** In Section 3, we discuss nine common threats to developing balanced assessment systems. These threats extend previous research and writing on the barriers and challenges to implementing balanced assessment systems and provide explicit "look fors" and open-source resources that can be used to investigate and address the threats in local contexts. We end Section 3 with practical next steps for those who want to start investigating the balance in their local assessment system. These high-leverage actions can be taken by any district or school who wants to explore the balance and quality of their local assessment system.

Section 4: Those interested in connecting the threats to balance with real-life examples from districts and states should read ***Section 4: District & State Vignettes*.** We provide vignettes from several districts and states that have been tinkering towards balance—treating it as a continuum rather than a fixed end state. By highlighting their experiences we hope to help districts and schools envision how they might overcome these common barriers to balanced assessment systems and move toward assessment systems that foster ambitious teaching and richer student learning.

Section 5: Leaders will find ***Section 5: Managing the Change Process*** important to help them sustain and build upon this work in their local context. We want to help local leaders ensure the time and effort put into exploring their local assessment system will result in changes that sustain leadership transitions and educator attrition because they have been instantiated into the collective knowledge base and assessment culture of the local context.

Appendices: We addressed a few other critical topics related to balanced assessment systems in the Appendices:

- ***Appendix A: The Role of Assessment Culture***
- ***Appendix B: The Importance of an Instructional Vision***
- ***Appendix C: Examining Coherence***

Throughout the Guidebook we indicate when and how these topics are relevant to the current discussion, allowing readers to delve deeper, if they wish.

SECTION 2: BACKGROUND

Connection to the 2024 National Academy Volume

The National Academy of Education (NAEd) released *Reimagining Balanced Assessment Systems* in 2024.[3] This volume updates and extends the seminal 2001 publication, *Knowing What Students Know: The Science and Design of Educational Assessment*.[4] The new volume was intended to provide a roadmap for developing, implementing, and using balanced assessment systems to support ambitious and equitable teaching and learning. Despite the valuable contributions made by the new NAEd volume, it is too conceptual and abstract to guide district and school leaders as they wrestle with design and implementation decisions regarding balanced assessment systems.

When they started the project, the NAEd project leaders and steering committee (of which we were a part) envisioned a companion guidebook to accompany *Reimagining Balanced Assessment Systems*, in the same way that *Building Educational Equity Indicator Systems: A Guidebook for States and School Districts* was written to help operationalize *Monitoring Educational Equity*.[5] **We wrote this practical guide to help district and school education leaders implement the teachings from *Reimagining Balanced Assessment Systems*.** We also intend state leaders to use this guide to help them support the efforts of local education leaders to make their assessment systems more balanced.

We wrote this practical guide to help district and school education leaders implement the teachings from *Reimagining Balanced Assessment Systems*.

Reimagining Balanced Assessment Systems revisited prior struggles in implementing balanced assessment systems and updated its theoretical underpinnings related to human learning and development. The volume also situated the work of balanced assessment systems within classrooms, with the goal of supporting ambitious and equitable teaching and learning with robust assessment literacy and professional learning for educators.

[3] Marion, S. F., Pellegrino, J. W., & Berman, A .I. (Eds.). (2024). *Reimagining balanced assessment systems*. National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems-project/

[4] National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. https://doi.org/10.17226/10019

[5] National Academies of Sciences, Engineering. and Medicine. (2020). *Building educational equity indicator systems: A guidebook for states and school districts*. The National Academies Press. https://doi.org/10.17226/25833; National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. The National Academies Press. https://doi.org/10.17226/25389

The authors and steering committee members of the National Academy volume offered an updated definition of balanced assessment systems:

> "Balanced assessment systems and practices [i.e., assessments inside the classroom under teacher control] are intentionally designed to provide feedback to students and information for teachers to support rich instructional and learning opportunities… Assessments outside of the classroom, at the district and state level, provide aggregate data to policymakers and education leaders, allowing for the monitoring of educational opportunities and support for high-quality instruction indirectly through the provision of appropriate curricular resources and professional development opportunities."[6]

The 2024 volume also included chapters offering guidance for state and district leaders to help design and implement balanced assessment systems in their locales. Leading researchers authored the chapters, and the steering committee and reviewers worked to ensure the entire volume provided a roadmap to developing balanced assessment systems centered on ambitious and equitable teaching and learning. Despite the best intentions of the leaders of the *Reimagining* volume, it is not written in a way that school and district leaders can easily use in their daily work.

Background on Balanced Assessment Systems

The call for balanced assessment systems began more than 20 years ago to correct the distortions and adverse effects that occur when large-scale tests are prioritized and often linked to high-stakes decisions. The seminal publication *Knowing What Students Know: The Science and Design of Educational Assessment* included a recommendation that "[t]he balance of mandates and resources should be shifted from an emphasis on external forms of assessment to an increased emphasis on classroom formative assessment designed to assist learning."[7]

The push for balance signified much more, however, than merely increasing the amount of informal testing done in classrooms to equal the weight of state-level tests. The original intention was to fundamentally change the character of classroom assessment practices to make them a part of effective teaching and learning. Indeed, the *Knowing What Students Know* study committee was convened to consider how measurement models and assessment methods should be revised in light of current conceptions of learning and knowledge development. Advances in learning research present in 2001, and even more so today, demand fundamental shifts in the representation of authentic learning goals and processes.[8] At the classroom level, a balanced assessment system will support assessment practices that are thoroughly integrated with day-to-

[6] Marion et al. (2024), p. 2-3.

[7] National Research Council (2001), p. 310.

[8] Nasir, N. S., Lee, C. D., Pea, R., & McKinney de Royston, M.(Eds.). (2020). *Handbook of the cultural foundations of learning.* Routledge; National Academies of Sciences, Engineering, and Medicine, 2019; National Research Council (2001).

day instructional practices and support deep disciplinary learning. At the level of school districts and states, a balanced assessment system will provide broader aggregate evidence of student attainment to inform policy decisions, including resource allocation. Overall balance may be achieved when the right assessments are used by the right users at the right time to make key decisions that either directly move student learning forward or move system structures and programs forward so they better support student learning.

Criteria for Balanced Assessment Systems

Knowing What Students Know outlined three criteria—coherence, comprehensiveness, and continuity—to characterize and define balanced assessment systems.

> "… systems are balanced when the various assessments are *coherently* linked through the specification of the learning targets, *comprehensively* provide multiple sources of evidence to support educational decision-making, and *continuously* document student progress over time."[9]

Overall balance may be achieved when the right assessments are used by the right users at the right time to make key decisions that either directly move student learning forward or move system structures and programs forward so they better support student learning.

The 2001 study committee believed that these three properties were necessary for creating a high-quality system of assessments rooted in a common model of knowing and learning. Unfortunately, test developers and users have struggled to understand and implement these criteria. We offer brief definitions of each below, and we refer interested readers to *Knowing What Students Know* for an in-depth discussion of these definitions.

Coherence signifies the need to connect assessments with a shared, research-based model of human learning. A *coherent* assessment system must be compatible with how student learning is expected to progress within an instructional domain. In an ideally balanced system, assessments at different levels of the educational system (state level to classroom) are compatible because of their close ties to the learning model. However, they will differ in grain size or specificity. Also, a *coherent* assessment system means that assessment information is strongly linked to curriculum and instruction.

Comprehensiveness describes the range of assessments used to provide a rich view of student learning. It also refers to the breadth of assessments used to serve the multiple purposes in the assessment system. In some cases, a desire to meet the ***comprehensiveness*** criterion may have

[9] Marion et al. (2024), p. 4.

led state and local assessment leaders to overbuild collections of assessments, leading to confusion and incoherence.

Continuity refers to the capacity of the assessment system to evaluate students' progress over time, which may be across units, within a school year, or across multiple school years. State and district leaders often turn to quantitative measures of student growth derived from commercial interim or state assessments to address ***continuity***. Student longitudinal growth measures have value for making comparisons among jurisdictions and over time, and researchers have been working on content-referenced approaches to student growth that focus attention on qualitative distinctions in student learning progress inferred from changes in assessment performance.[10] However, continuity does not require quantitative growth measures. Student portfolios of work, for example, offer a valid way to evaluate student progress over time.

We discuss two additional criteria that the 2024 volume added to the original three: utility and efficiency.[11] ***Utility*** (more simply referred to as usefulness) is the degree to which the assessment system provides the information necessary to support its multiple and often diverse purposes, such as instructional usefulness and program evaluation.[12] ***Efficiency*** is maximizing the use of assessment resources and eliminating redundant, unused, and untimely assessments. Efficiency determinations identify and reduce assessments that are not serving the stated purposes or are redundant with other, more useful assessments. Unfortunately, many district personnel assume a set of assessments constitutes a system if it contains at least summative, interim, and formative components, but this menu approach to designing a system is not supported by the research literature. Efficiency could conflict with comprehensiveness if district leaders continue to add assessments to the system to serve multiple users and purposes. Thus, we contend that having assessment system designers or evaluators think about efficiency can be a helpful hedge against the proliferation of unnecessary or redundant assessments.

In Section 3, we draw on three of the five criteria that we have found resonate with school and district leaders—efficiency, usefulness, and coherence—to organize the specific list of threats to achieving more balanced assessment systems.

---

[10] E.g., Student, S.R. (2022). Vertical scales, deceleration, and empirical benchmarks for growth. *Educational Researcher, 51*(8), 536-543. https://doi.org/10.3102/0013189X221105873 ; Briggs, D.C., McClure, K., Student, S.R., Wellberg, S., Minchen, N., Cox, O., Whitfield, E., Buchbinder, N. & L. Davis (2025). Visualizing and reporting content-referenced growth on a learning progression. *Educational Assessment.* https://doi.org/10.1080/10627197.2025.2503288

[11] Chattergoon, R. & Marion, S.F. (2016). Not as easy as it sounds: Designing a balanced assessment system. *National Association of State Boards of Education, 16*(1), 6-9. https://www.nasbe.org/measuring-up-recalibrating-systems-of-assessment/

[12] Evans, C. M. & Marion, S. F. (2024). *Understanding instructionally useful assessment*. Routledge.

Barriers to Balanced Assessment Systems

Previous writing has explained the barriers and factors that impede the design and implementation of more balanced systems of assessment. For example, as seen in Table 2-1, Marion et al. (2019) wrote about four barriers to balanced assessment system design and implementation to which two more were added in the most recent volume (Marion et al., 2024), and Polikoff and Hutt (2024) wrote about two categories of factors that have hindered the growth and impact of balanced assessment systems.[13]

TABLE 2-1 Barriers and Factors that Impact Balanced Assessment System Design and Implementation from the Previous Research Literature

Barriers to Balanced Assessment System Design and Implementation Marion et al. (2019) & Marion et al. (2024)	**Factors that have hindered the growth and impact of balanced assessment systems** Polikoff & Hutt (2024)
1. Influence of politics, policy, and political boundaries 2. Influence of commercialization and proliferation of assessments 3. Lack of attention to curriculum and learning in the design of assessment systems 4. Lack of assessment literacy at multiple levels of the system 5. Limited understanding of human development and student learning 6. Misconceptions associated with the meaning of balance	1. Technical Challenges ○ Measuring multiple complex domains ○ Interpreting information across grade levels for multiple dimensions ○ Weighting multiple measures ○ Scoring student work ○ Technological challenges and opportunities 2. Political and Practical Challenges ○ Poorly designed assessment and curriculum policies ○ Shifting political barriers ○ The challenges of embedding assessments in the curriculum ○ Lack of capacity across levels of the system ○ Instructional reform in the context of loosely coupled systems

[13] For previous research and writing on barriers and challenges to balanced assessment systems, see: Marion, S. F., Thompson, J., Evans, C., Martineau, J., & Dadey, N. (2019). *The challenges and opportunities of balanced systems of assessment: A policy brief*. The National Center for the Improvement of Educational Assessment. https://files.eric.ed.gov/fulltext/ED598421.pdf; Marion, S. F., Pellegrino, J. W., & Berman, A. I. (2024). Reimagining balanced assessment systems: An introduction. In Marion, S. F., Pellegrino, J. W., & Berman, A. I. (Eds.). *Reimagining balanced assessment systems*. National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems/chapter1; Polikoff, M. S., & Hutt, E. L. (2024). The struggle to implement balanced assessment systems: Explanations and opportunities. In Marion, S. F., Pellegrino, J. W., & Berman, A. I. (Eds.). *Reimagining balanced assessment systems*. National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems/chapter2

These barriers and factors affect the educational ecosystem and set the systemic constraints within which an assessment system resides. Lack of assessment literacy, for instance, is noted by both publications as having a significant impact on assessment systems because all aspects—from design to interpretation, use, and evaluation— depend on that foundational condition being met.

The next section of this paper extends this previous work and thinking by identifying and explaining what becomes "out of balance" in observable ways as a result of these previously identified barriers and factors that hinder the design and implementation of balanced assessment systems. We describe these as threats to balance and explain what they are, how they can be observed or investigated in local contexts, and the practical next steps district and school leaders can take to start to address some of the underlying barriers and factors that are impeding balance.

SECTION 3: THREATS TO BALANCE & PRACTICAL NEXT STEPS

Before exploring processes for improving balance in assessment systems, it is necessary to first describe *how* those systems can be out of balance. We have developed a list of what we call *threats to balance*, drawing on ideas presented in the NAEd volume *Reimagining Balanced Assessment Systems*, previous research, and our own experiences working with school districts and state education agencies.[14] We grouped these nine threats (see Figure 3-1) under the most relevant criteria of balanced assessment systems for this discussion (efficiency, usefulness, and coherence).[15]

Common Threats to Efficiency:
Threat 1: Too much testing overall, particularly early or later in the year
Threat 2: Redundant assessments
Threat 3: Unused assessment results

Common Threats to Usefulness:
Threat 4: No clear match between the assessment purpose, design, and use
Threat 5: Assuming all tests can inform instruction

Common Threats to Coherence:
Threat 6: Inconsistency between assessments and instructional vision
Threat 7: Policies and politics that distort practice
Threat 8: Over-emphasizing the role of summative assessment
Threat 9: Under-emphasizing the role of formative assessment

FIGURE 3-1 Nine threats to balanced assessment systems.

This list of threats is not intended to be exhaustive, but we believe it includes common areas of imbalance in local assessment systems. Several of the threats are interrelated. For example, if there is too much testing happening at one or more grade levels, it is also possible that some of those assessment results are not actually used or are redundant. Rather than thinking of these as separate categories, think of them as different viewpoints or ways to look into one's local assessment system and investigate possible areas of imbalance.

[14] Marion et al. (2024).
[15] Chattergoon & Marion (2016); National Research Council (2001).

Some of the threats occur when an assessment system is not designed purposefully to consider information needs of various interest-holders in a systematic way. This can happen over time if the system is not critically examined periodically and when assessments are added without removing others. When new assessments are being considered, the information needs of teachers and leaders should be considered holistically. For example:

- What assessment information is needed by classroom teachers?
- When do they need that information?
- What should they use that information to do, or what decisions will be adjusted based on that new information?

Similarly, the same thought exercise can be extended to other interest-holders in the system such as school and district leaders, policymakers, and parents.

Other threats relate to how the system is implemented or the nature of policies or behaviors of key interest-holders in the system. For example, do various school, district, state, or federal policies support or undermine the goals of the local assessment system?

In the remaining parts of this section, we will describe each group of threats and suggest some ways to look for the threat in action within schools and classrooms. For each set of threats we provide a table with resources and suggestions to guide investigating threats to efficiency, usefulness, and coherence. Section 3 ends with practical next steps for schools or districts interested in knowing where to start this work, who should be involved, and what might come next.

Supports or Conditions Necessary to Begin This Work

We think there are a few supports or conditions necessary to begin this work. First, this work should be led by district and school leadership. Second, those leaders should then involve others in the district, especially classroom teachers, as they begin to investigate and explore the efficiency, usefulness, and coherence of their local assessment system. And, finally, there is a foundational level of knowledge needed about assessment and testing to identify threats to balance (sometimes referred to *assessment literacy*).

Assessment literacy can be defined as the knowledge, skills, and understandings necessary to design, implement, and use assessments to support educational decision-making that ultimately improves student achievement.[16] Assessment literacy is crucial and it is discussed at length in previous research and writing on barriers to the design and implementation of balanced assessment systems (see Table 2-1). However, we don't want any school or district leader to stop reading here and decide not to engage in this work because they are not sure what we mean by assessment literacy or how much is necessary to get started. We believe that open and frank discussions about the threats to balance, listening to a wide variety of interest holders, and using the open-access resources provided in this Guidebook can help deepen participants' knowledge, skills, beliefs, and understandings necessary to meaningfully engage in this work and develop more assessment literacy along the way.

[16] The term 'assessment literacy' was first coined by Stiggins, R. (1991). Assessment literacy. Phi Delta Kappan, 72, 534–539. Available online: https://www.jstor.org/stable/i20404445. Since then, assessment literacy as a concept has expanded and become more nuanced for different users of assessment information (teachers, school/district leaders, policymakers, parents, students). For a good review of the research on what is included in teacher assessment literacy, see: Brookhart, S. M. (2024). Educational Assessment Knowledge and Skills for Teachers Revisited. *Education Sciences*, *14*(7), 751. https://doi.org/10.3390/educsci14070751. Table 3-2 in this Guidebook provides links to open-access assessment literacy resources and standards for different users.

Common Threats to Efficiency

Threat 1
Too much testing overall, particularly early or later in the year

Explanation

No matter how useful assessment is, there is only so much data that leaders and teachers can use. While there is no hard and fast rule about how much testing time is too much, the U.S. Department of Education suggested in non-regulatory guidance that no more than 2% of instructional time should be spent on state testing.[17] It is hard to imagine that many leaders would want to intentionally layer on district- and school-required tests well beyond that threshold. This means that if there is 1 hour of math instruction per day over 180 days of school, then 2% of instructional time is no more than 3.6 hours of state- and district-administered math testing per year.

Beyond the overall amount of testing, when that testing occurs can also be problematic. For example, the amount of testing early or later in the year can become excessive if not carefully monitored. Decisions can be made in silos without considering the totality of assessments already in place. Early elementary students are often bombarded with universal screeners early in the year and prep for the state test later in the year. Elementary and middle school students may take multiple interim assessments early in the year. For multilingual learners or students with disabilities, there are likely additional tests which can result in less instructional time. Similarly, for students in Tier 2 or 3 interventions, the frequency of progress monitoring is an additional layer of testing.

Multiple sources of assessment information can be very useful to build a more complete picture of students' strengths and to identify students who may need additional support and interventions; however, it can also be a distraction if the information provides conflicting signals or is not used to make different decisions for students than would be made using existing assessment information from the previous year.

[17] Donnelly, K. (2015). *Obama Calls to Curb Standardized Testing in Schools*. NBC News. https://www.nbcnews.com/news/us-news/obama-calls-curb-standardized-testing-schools-n451741

Look-Fors

- Count up the number of instructional minutes the general education student spends taking teacher-, school- or district-required assessments across the four core content areas in each grade. Count everything except day-to-day formative assessment practices. There is no correct answer for the amount of time spent, but interest holders with various perspectives should discuss whether they believe it to be too much or about right.
    - Consider how much time is lost to instruction for assessment. In a school year of about 180 days and a 1 hour math block each day, 18 hours of math testing would mean that a student spent 10% of their possible math instructional time being assessed, and likely more considering time lost to assemblies, field trips, or other school events.
- Look specifically at the number of instructional minutes students in each grade spend taking tests in the first two months of the school year and the last few months across the core content areas. Do interest holders consider this a reasonable trade-off?
    - As part of this, consider the amount of testing from the student perspective. How many assessments are they taking per week, per month—does that seem reasonable? Remember that English language learners, students with disabilities, and students receiving services within multi-tiered systems of support are likely taking more tests in addition to the ones taken by the "general education" student.

Section 4 vignettes that address this threat:

- Cupertino Union SD recognized that it was time to reevaluate the need for some additional assessments it had adopted during COVID so it could understand what was needed and used.
- Katy ISD was concerned about the testing burden on students, so it used an assessment system audit to explore those concerns.
- Connecticut DOE was concerned about the amount of testing taking place in districts. It released a report with strategies for districts to explore.
- Oregon DOE recognized the burden of excessive testing in the first weeks of kindergarten, so it replaced the state's kindergarten readiness assessment with a qualitative tool focused on family-teacher dialogue.
- Pennsylvania DOE was concerned about the amount of testing taking place in districts, so it created an assessment mapping process to help districts examine their system.
- Wisconsin DPI was concerned about the amount of testing taking place in districts. It worked with state intermediary units to provide district support for assessment audits.

For more information related to this threat, see *Reimagining Balanced Assessment Systems*:

- Chapter 2 (p. 29, Efficiency)
- Chapter 7 (p. 208-209, Local Control; p. 228-230, Local Assessment System Auditing Tools)

Threat 2
Redundant assessments

Explanation

Assessment takes time away from instruction (excluding formative assessment practices, which should be inseparable from instruction). There is a cost-benefit analysis that must be considered when evaluating how much testing time is being spent in one content area, let alone multiple content areas. Not only does formal assessment take time away from instruction, it can also overburden students who are taking assessments across content areas.

Efficient assessment systems use high-quality information from state tests in lieu of adding other assessments that tell the same story. For example, we believe state assessment results from the end of one school year are often underused at the beginning of the next school year leading to redundant assessments. If a school or district wants to identify students who need additional support or intervention, it makes sense to use their end of year state test results and then use targeted tools to monitor student progress.

Additionally, redundancies creep into an assessment system when district-created or -purchased tests are administered around the same time as state-mandated tests that provide similar information. It is common for students to take the state test in one week and then a district-required interim soon thereafter. What is the purpose of this duplication? And what is the impact on students if this duplication is happening across multiple content areas?

Redundant assessments can also linger in a system when legacy assessments remain and new assessments are added over time without careful scrutiny and consideration of need and use. This may occur when different levels of the system (teacher level, school/district administration level, state level) all want their own assessment for specific uses without realizing the impact on students, loss of instructional time, and general overlap of information collected.

A specific area of caution on redundant assessments relates to states with reading laws that mandate frequent screening in grades K-2. Such laws can disproportionately impact multilingual learners. Particularly in kindergarten, entering multilingual learners may take an English language proficiency screener (to determine if they'll be classified as English learners or not), a kindergarten readiness assessment, and a literacy screener, all of which will likely ask them to name letters, identify objects in pictures, and write something. Each of these tests has a different purpose, but the questions are likely very similar.

Look-Fors

- Examine the list of assessments administered to students in each grade and see if there are any that measure the same content or require the same skills as another assessment administered around the same time. In other words, can any assessments be removed because they aren't providing new or useful insights not already available from other sources?
- Ask whether decisions about students, instruction, resources, or programs are likely to be different if one or more assessments were removed. If not, then which assessments can be removed because they are not being used to inform better educational decision-making?

Section 4 vignettes that address this threat:

- Katy ISD was concerned about duplicative assessments in its assessment system, so it used an assessment system audit to explore assessment overlaps.
- Connecticut DOE was concerned about redundant assessments in district systems. It released a report with strategies for districts to explore.
- Oregon DOE was concerned about redundant assessments in districts' assessment systems. It completed a mandated state audit that identified some areas for improvement.
- Pennsylvania DOE was concerned about redundant assessments in district systems, so it created an assessment mapping process to help districts examine their systems.
- Wisconsin DPI was concerned about redundant assessments in district systems and worked with state intermediary units to provide district support for assessment audits.

For more information related to this threat, see *Reimagining Balanced Assessment Systems*:

- Chapter 2 (p. 29, Efficiency)
- Chapter 7 (p. 228-230, Local Assessment System Auditing Tools)

Threat 3
Unused assessment results

Explanation

Education leaders and teachers are often overwhelmed with the amount of assessment data they get during the school year, and the timing of that data might not match up with when leaders or teachers need to make decisions. Sometimes assessments are administered, and results passed along to others in the district or school, but the data are not used in any substantive way to inform decision-making. In other words, little critical information or insight would be lost if the assessments were not administered. All assessments given to students should be used by someone within the local setting to inform an important decision that otherwise could not be made without the assessment information.

For example, many schools or districts purchase or create interim assessments defined as assessments that can be aggregated to the school or district level to help inform policy.[18] Interims are often administered three times during the school year, but the within-year information is not always used to make decisions during that year or even the next academic year. This raises a key question: are these types of tests actually needed, or have they just become a tradition or a compliance exercise?

Look-Fors

- Examine the extent to which all assessments administered are used by someone within the local setting to make better educational decisions for students or evaluate program effectiveness.
- Ask for specific examples of how the assessment information is used and how decision-making would be affected without the assessment information.
- Examine the timing of assessments and assessment results by month (or another period of time) over the course of the school year. Are leaders or teachers getting very similar information from different assessments in a relatively short period of time or not until it is too late to inform decision-making about programs, personnel, professional learning, school structures, instruction, etc.?
- What doesn't appear to be needed or used and can therefore be eliminated?

[18] Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, *28*(3), 5-13.

Section 4 vignettes that address this threat:

- Howard-Suamico SD realized that teachers were not making use of information from assessments used to measure students' skills in their Graduate Profile and so worked with teachers to incorporate the Profile skills into curricular-based assessments.
- Katy ISD was concerned about whether all the assessment data being collected was being used. It used an assessment system audit to explore how each assessment was being used.
- Connecticut DOE was concerned about unused assessment data and released a report with strategies for districts to explore.
- Pennsylvania DOE was concerned about unused assessment data, so it designed an assessment map to support school and district leaders in interrogating their assessment systems.
- Wisconsin DPI was concerned about unused assessment data, so it worked with state intermediary units to provide district support for assessment audits.

For more information related to this threat, see *Reimagining Balanced Assessment Systems*:

- Chapter 2 (p. 29, Efficiency)
- Chapter 7 (p. 228-230, Local Assessment System Auditing Tools)

How to investigate and evaluate threats to efficiency

The look-fors that we provide for each threat are meant to provide initial insight into whether a local assessment system is out of balance when it comes to efficiency. However, they don't describe systematic processes and procedures to investigate and evaluate the efficiency of local systems. Therefore, in Table 3-1, we suggest some systematic ways to explore the efficiency of local assessment systems.

We suggest two high-leverage and high-impact strategies for approaching this work: one is oriented towards understanding the problem more deeply; the other combines understanding the problem and identifying solutions.

1. **Information-gathering**. Gathering information via focus groups, surveys or less formal means may be a first step to understanding general perceptions about the amount, time or usefulness of the assessments currently being used. This step may not be necessary if you already are confident that these threats are present. However, hearing from a broad range of interest holders may uncover conflicting perceptions and help local school/district leaders better understand the scope and nature of the problem, assuming one exists. The focus group discussions may also serve to surface related aspects of practice that fall under the broader umbrella of "assessment culture," which is discussed in Appendix A.
2. **An assessment system review or audit**. A comprehensive assessment system review is a concrete way to document, analyze and evaluate assessments being used with the intention of making informed decisions to change. This process can provide information for many of the look-fors associated with the three threats to efficiency such as the number, timing, and use of assessments. There are a variety of tools available to support this work.

TABLE 3-1 Possible Strategies, Tools and Resources to Further Explore Common Threats to Efficiency

Threats & Indicators	Strategies	Selected Tools and Resources
Threat 1: Too much testing overall, particularly early or later in the year **Threat 2:** Redundant assessments **Threat 3:** Unused assessment results	● Information-gathering (e.g., surveys or focus groups with students, teachers or leaders) to explore perspectives on the amount or type of testing, and use or non-use of results ● Assessment system auditing tools (see the column to the right)	Open-access tools that support an assessment system auditing process: ● Assessment System Review Online Learning Path (2.0 Version) by the Center for Assessment & California Collaborative for Educational Excellence (online modules, videos, and tools) ● Fewer and Better Local Assessments: A Toolkit for Educators (Playbook) by Education First (online toolkit with rubrics and other tools; accompanies Achieve's inventory) ● Student Assessment Inventory for School Districts by Achieve (online inventory tool and guidance document) ● Assessment Inventory Project by the Georgia Partnership for Excellence in Education (online inventory tool, facilitation guide, and other resources) ● Assessment Inventory Resource by The Center on Standards & Assessment Implementation (online inventory tool and evaluation criteria with questions)

Common Threats to Usefulness

Threat 4
No clear match between the assessment purpose, design, and use

Explanation

Not having a clear match between the assessment purpose and design has arguably been the single most common issue with assessment in schools. Too often, claims are made as if every assessment given to students can meet all information needs and users. The truth is that every assessment is typically designed to fulfill a limited number of purposes really well, whether that is to adjust instruction, evaluate or monitor programs, or diagnose specific learning needs. Sometimes the name of a test implies one purpose, but the actual design supports a different purpose. In a balanced system, not all tests are supposed to directly inform classroom instructional decisions, but we should be clear about which is which; otherwise, the usefulness (or utility) of the assessment for a particular purpose is challenged.

To limit over-inflated claims about what assessments can do, curriculum providers, test vendors, and local educators who create common assessments should not only be clear about what their test is designed to do and what claims the test is intended to support, but should also provide validity evidence in support of such claims and intended uses. This supports best practice because it attempts to prevent vendors and designers from overselling what their products or materials can be used for and conflating summative and formative uses, which ultimately prevents misuse.

Look-Fors

- Examine the list of assessments required by the district, state, or teachers (whether generated as part of an assessment audit or other process) to determine the alignment between the assessment design and the intended use and purpose of the assessment.
    - Is there agreement between teachers and leaders on the purpose of each assessment?
    - Does the information provided allow users to make the kinds of decisions they intend?

- Evaluate commercially purchased tests and locally created common assessments by reviewing vendor or internal documentation against quality criteria.
    - Are the purpose and intended uses of the assessment clearly articulated and reflected in the assessment design?
        - Vendor technical documentation should articulate the purpose of the assessment, describe the design of the assessment, the rationale for that design, and how it supports the intended purpose and use of the results.
        - For locally created common assessments, there likely will not be formal technical documentation, but there still should be a clear statement of intended purpose and use. Does the design of the assessment support that intended purpose and use?

Section 4 vignettes that address this threat:

- Chicago SD recognized that interim assessments were being used to support a number of high-stakes decisions which resulted in unintended, negative impacts on instruction.
- Katy ISD recognized a mismatch between mostly multiple-choice assessments and their vision for student learning. The district leadership team is working with principals to deepen collective learning about the purpose and use of assessments.
- Oregon DOE recognized that districts were struggling to know how to use the range of assessments they had access to, so it created a guidance document to support districts.
- Pennsylvania DOE recognized that lack of clarity between assessment purpose and design can undermine other assessment efforts, so it designed an assessment map to support school and district leaders in interrogating their assessment systems.
- Wisconsin DPI recognized the potential for mismatches between assessment purpose and design, so it collaborated with its regional service agencies (CESAs) to co-design training on a range of assessment literacy topics.

For more information related to this threat, see *Reimagining Balanced Assessment Systems*:

- Chapter 2 (p. 21-23, What are Balanced Assessment Systems?; p. 28, Utility)
- Chapter 7 (p. 221-222, State Action 4: Mitigate Misuse of the State Test Through Clear Reporting and Guidance)

Threat 5
Assuming all tests can inform instruction

Explanation

This threat is fundamentally about a mismatch between assessment purpose and use (Threat 4), but we highlight it separately because so many educators assume all tests can directly inform instructional decisions. In our experience, this assumption has contributed to over-testing in schools and to the general frustration with the lack of useful information derived from testing experiences.

Many assessments are described as being able to "inform instruction," but the nature of the reports and grain size of information make that unlikely. This is most often the case with school- or district-required assessments that are actually designed for leaders to monitor and evaluate programs, curricula, instructional quality, and school improvement. Those external-to-the-classroom assessments do not provide substantive insights into student thinking that can be used by teachers to monitor or adjust daily or weekly instruction. Information that can inform instruction is more likely to come from 1) open-ended tasks or activities where student thinking is made visible or 2) class discussions where educators have the opportunity to probe student thinking and analyze resulting student work to better understand what students have learned and how well they have learned it.

Look-Fors

- If you aren't sure if an assessment is designed to support direct instructional uses:
    - Conduct think-alouds with teachers where you ask them to explain how they would use the assessment results to inform their instruction the next day or within a very short amount of time, for whom, and why.
    - Ask for specific details about what teachers would change in their upcoming instruction.
- Consider key features of assessments likely to most directly inform next instructional steps.[19] Discuss whether the assessments that are intended to inform teachers' daily instructional decision-making reflect these features. For example:
    - Do they focus on a relatively narrow set of standards?
    - Are they embedded within the curriculum?
    - Do they make student thinking visible?
    - Do they provide timely information (moment by moment, day by day)?

[19] Evans, C. M. & Marion, S. F. (2024). *Understanding instructionally useful assessment*. Routledge.

Section 4 vignettes that address this threat:

- Katy ISD leaders recognized that they were assuming some district-mandated assessments were useful to inform instruction. They are now working to improve assessment literacy as a mechanism to enable principals to be stronger instructional leaders.
- Oregon DOE mandated a state assessment audit that identified some misconceptions about which assessments can inform day-to-day instruction, and is considering how to address them.

For more information related to this threat, see *Reimagining Balanced Assessment Systems*:

- Chapter 1 (p. 2-3, Balanced Assessment Systems, Redefined)
- Chapter 4 (p. 94-95, Conceptualizing Ambitious Teaching)
- Chapter 6 (p. 172-173, Assessments to Support Ambitious Teaching and Learning)

How to investigate and evaluate threats to usefulness

In Table 3-2, we suggest some systematic ways to explore the usefulness of local assessment systems. We suggest two high-leverage and high-impact strategies for approaching this work:

1. **Build off an assessment system review process to closely examine purpose, design, and use of selected assessments.** Ensuring that there is an appropriate alignment between purpose and use of assessments is an important step toward improving the local assessment system. While an audit can be completed in a relatively short period of time, reviewing and, if necessary, revising local assessments can take more time. Participants may need to engage in professional learning to support this work.
2. **Deepen assessment literacy knowledge and skills specifically related to the concept of instructional usefulness in order to review local assessments.** We noted that *Threat 5: Assuming all tests can inform instruction* is a specific instance of the more general *Threat 4: No clear match between the assessment purpose, design, and use.* That said, given the common use—and misuse—of the phrase "this test can be used to inform instruction," a book study on this topic could be very valuable to help teachers reconsider what kinds of classroom assessment will be most useful and how that depends on assessment design.

TABLE 3-2 Possible Strategies, Tools and Resources to Further Explore Common Threats to Usefulness

Threats & Indicators	Strategies	Selected Tools and Resources
Threat 4: No clear match between the assessment purpose, design, and use **Threat 5:** Assuming all tests can inform instruction	● Following the initial accounting of assessments from an audit, exam selected assessments with teachers and leaders to explore the match between design and purpose. The assessment literacy resources can be used to support deeper discussions. ● For Threat 5, teachers and leaders could use *Understanding Instructionally Useful Assessment* as a book study, followed by an application of learning to examine local assessments from this perspective.	**Tools or resources that support assessment literacy:** ● Educational Assessment 101 for K-12 Teachers & Leaders by the Center for Assessment (online modules, videos, and tools) ● Assessment Resources & Learning Opportunities by the Michigan Assessment Consortium (online modules and resources) ● The District Assessment Procurement Protocol by EdReports, Center for Assessment, & Lyons Assessment Consulting (tool for selecting interim assessments) ● *Understanding Instructionally Useful Assessment* by Carla M. Evans & Scott F. Marion (book)

Threats & Indicators	Strategies	Selected Tools and Resources
		Assessment literacy standards: ● Assessment Literacy Standards by the Michigan Assessment Consortium (separate standards for teachers, building and district administrators, local and state policymakers, students and their families)

Common Threats to Coherence

Threat 6
Inconsistency between assessments and instructional vision

Explanation

One aspect of system coherence is that curriculum, instruction, and assessment are all working toward a common set of learning goals. This occurs when assessment is seamlessly integrated with curriculum and instruction so that teachers get information back that they can interpret and use within the normal scope and sequence of teaching and students get feedback that they can use to improve while in the midst of an instructional unit. In contrast, when curriculum, instruction, and assessment are not aligned, teaching and learning are interrupted to prepare students for an external-to-the-classroom assessment that covers some proportion (or all) of the grade-level standards, including some standards that may not yet have been covered in the curriculum or instructional sequence.

Another aspect of systems that can lead to incoherence is the common practice of assessing students mainly with selected-response items that mimic the format of the state test. State tests are unlikely to change dramatically in format for many reasons. However, district, school, and classroom assessments need not mimic large-scale testing, which has notable limitations in being able to assess higher-order, integrated skills embedded in real-world contexts and problems.

For example, if the district's instructional vision is to support critical thinking, deep conceptual understanding, real-world application, and other higher-order thinking skills, as well as reflect research on how students develop disciplinary knowledge, skills, and practices, then the assessments should complement that vision rather than present conflicting goals for learning. This doesn't mean that every assessment needs to be a performance-based assessment; however, each assessment should align with the standards being assessed and the collection of assessments should support the district's instructional vision rather than hinder it.

Look-Fors

- Re-read the district's mission and vision statements and the district's instructional vision (if it exists). Gather a purposeful sample of end-of-unit classroom assessments and any school- or district-required assessments. Examine the assessments to see if they are consistent with the mission and vision statements (and instructional vision):
    - What item types are used and how does that reflect the system's goals? (i.e., mainly selected-response such as multiple-choice, constructed-response such as open-ended questions, and/or performance tasks).
    - How are the disciplinary practices embedded in the standards (e.g., argumentation and justification; writing testable questions) represented in the assessments?
    - Do students have opportunities to demonstrate their durable skills (e.g., collaboration, critical thinking, communication, creative thinking)?
    - For a more extended discussion on how to develop an instructional vision, see Appendix B. For more on how to examine the coherence between contemporary views on how students learn and the assessments being used, see Appendix C.

Section 4 vignettes that address this threat:

- Cañon City Schools wanted coherence among assessment practices, their instructional vision and Portrait of a Graduate. It incorporated critical assessment practices into its annual school review process.
- Chicago SD recognized that any potential changes to its assessment system had to be grounded in a clear, district-wide instructional vision.
- Howard-Suamico SD realized that Portrait of a Graduate skills were not integrated into teaching and assessment, so it worked with teachers to incorporate those skills into curricular-based assessments.
- Oregon DOE recognized that the state's kindergarten readiness assessment added to the burden of excessive testing, which was not in line with the department's vision of instruction. It replaced the test with a qualitative tool focused on family-teacher dialogue.

For more information related to this threat, see *Reimagining Balanced Assessment Systems*:

- Chapter 2 (p. 25-26, Coherence)
- Chapter 6 (p. 168-170, Introduction)

Threat 7
Policies and politics that distort practice

Explanation

State, district, and school policies are important to support fair and consistent practices, processes, and procedures, but almost all policies have the potential for unintended negative consequences. State or local politics can also impact assessment systems through state-level assessment mandates or local politics that can impact teacher professional learning time (e.g., collective bargaining agreements) or other aspects of practice. District and school leaders play a key role, along with policymakers, in reviewing current policies that can distort practice and preventing new policies from being enacted that lead to problems of practice.

For example, local grading policies can create situations in which teachers administer assessments because they are required to enter a minimum number of grades in their gradebook each week, quarter, or semester. Some policies require teachers to grade every piece of student work, which distorts formative assessment practice. Local leaders have direct control over grading policies and practices.

Similarly, state or district policies on progress monitoring or universal screeners can lead to over-testing and redundancy when they are applied to all students regardless of other demonstrations of learning that show a student no longer needs to be progress-monitored or screened for dyslexia, for example.

Districts may also have policies on the assessment of English learners (ELs) or students with disabilities (SWDs) which can have unintended consequences such as additional testing time, which in turn means less time on instruction for these already vulnerable student populations. Clearly, there are instances in which additional assessments are needed for ELs and SWDs. But in other instances, it seems possible to use pre-existing assessment information rather than give another test.

Teacher evaluation provides another example of policies that can distort practice. Evaluation systems that require teachers to demonstrate “effectiveness” by showing how they helped students “learn” and “grow” during the year can also have an impact on assessments. These policies often require that comparable tests are given at two or more time points, which often adds to the amount of testing and the type of testing occurring in classrooms. If these tests are used only to set student learning objectives, for example, and don’t inform or support best practices, then the policy distorts practice. Even as some states and districts have moved away from growth measures as part of teacher evaluation, the assessments and practices may remain.

Look-Fors

- Review common practices across classrooms in the district. For example, survey a purposeful sample of teachers (across grade levels and content areas), school leaders, and district leaders about policies that hinder impactful teaching, learning, and assessment.
- Review existing policies within the district and school that relate to assessment. Identify policies that may be leading to less-than-ideal practices. For example:
    - Are there local grading policies that require a certain number of grades to be entered every week or grading period? Is that number the same across grades and content areas? How are grading policies related to curriculum materials and the number of assessments expected to be graded from such high-quality instructional materials?
    - What are the assessment-related policies for ELs or SWDs?
    - Are there state or local teacher evaluation policies that are impacting assessment practices?
    - Are there other assessment-related policies that emerge from discussion as having unintended, negative consequences?

Section 4 vignettes that address this threat:

- Chicago SD recognized that policies on the use of interim data related to grade promotion and other decisions had a negative impact on teaching and learning.
- Connecticut DOE identified ways in which policies in some areas of the educational system have unintended consequences for student assessment, and is working on how to counter them.
- Oregon DOE identified concerns about districts' capacity to select commercial assessments that meet their needs. It advocates using resources to specifically address this issue and build district leaders' assessment literacy.

For more information related to this threat, see *Reimagining Balanced Assessment Systems*:

- Chapter 6 (p. 188-191, Grading)
- Chapter 9 (p. 280-284, Other Policy Influences and Actors; p. 284-285, How Policies Are Designed to Shape Instruction, Learning, and Assessment)

Threat 8
Over-emphasizing the role of summative assessment

Explanation

The original call for balanced assessment systems arose from a concern that large-scale, standardized assessments and other types of summative assessments were becoming far more important than formative classroom assessment practices and processes that support student learning. Teachers and students need moment-by-moment and day-by-day information about student progress toward learning goals. On-demand, point-in-time measures of what students have achieved are needed less frequently.

However, both direct and indirect messages about the importance of the end-of-year assessment can be miscommunicated or misunderstood at the district, school or teacher level. For example, teachers may believe that they should only use item types that are on the state assessment for their classroom assessment, or teachers may feel pressure to engage in extensive test preparation activities. Whether real or perceived, the pressure to boost school or district performance on state tests can lead to classroom practices that prioritize test preparation over genuine learning. This, in turn, increases the risk of narrowing the curriculum to include only content and item types that appear on state-mandated tests.

Look-Fors

- Ask teachers, principals, other school and district leaders about explicit or implicit messages about summative assessments. For example:
    - What messages are they getting (or inferring) about the role of state assessments, test-preparation activities, etc.?
    - What messages do they think other interest-holders in the system receive about the importance of state, interim, or classroom summative assessments?
    - How much emphasis are teachers expected to place on classroom summative assessments in comparison to formative assessment practices?
    - Do different groups have similar perceptions? Why might that be the case?
- Ask teachers to report how much time they are spending on test preparation activities at different times of the year, and the nature of those activities.

- Examine the types of professional learning and instructional coaching offered to teachers over the last five years, including via teacher professional learning communities or classroom walk-throughs and observations.
    - What has been communicated to teachers about the purpose and intended use of state and interim tests results, either explicitly or implicitly?
    - How much professional learning has focused on data-based decision-making or making sense of interim or state assessment results in comparison to an emphasis on formative assessment processes and student work analysis?
    - What messages have been sent about test prep, and specifically, what is appropriate and necessary test prep—and what is not—and why?
- Use focus groups or other means to analyze the assessment culture in the school or district. What are the collective expectations, values, and beliefs about the role, value and purpose of assessment? (see Appendix A for more on assessment culture)

Section 4 vignettes that address this threat:

- Cañon City Schools recognized that to promote appropriate balance of summative assessment with other forms of assessment it needed consistent messaging. One way the district signals this is through the evaluation categories in its Instructional Program Reviews.
- Cupertino Union SD realized that students wanted to engage in deeper learning that mattered to them. It developed a Portrait of a Learner and is working now on ways to assess the important skills described in the Portrait.
- Oregon DOE provides professional development resources to support teachers implement and interpret summative assessments appropriately.
- Wisconsin DPI and its CESAs provide targeted professional development for districts that includes improving end-of-unit summative assessments.

For more information related to this threat, see *Reimagining Balanced Assessment Systems*:

- Chapter 6 (p. 171-172, A Culture of Testing; p. 176-177, Distinguishing the Role of Distal Assessments)
- Chapter 7 (p. 207-208, Failure to Envision State Assessments Within the Broader System of Assessments)

Threat 9
Under-emphasizing the role of formative assessment

Explanation

Formative assessment—often described as a process or set of practices—refers to the kind of assessment that is interwoven with classroom teaching and learning.[20] There is an extensive literature base spanning more than two decades that provides empirical support for the positive impact of formative assessment on student learning.[21]

One common description of formative assessment is that it enables teachers and students to answer three questions: Where is the learning going? Where are learners right now in relation to the learning goals? How can the gap be closed between current and intended status?[22] Different aspects of the formative assessment process address these three questions.

- Having clear learning goals and success criteria that describe intended learning and what success looks like provide explicit roadmaps to answer the question about where learning is going and how the learning goals relate to and build upon one another.
- Evidence of student learning comes from what students say, write, do and make as they are learning. Opportunities to elicit evidence of student thinking aligned with learning goals can occur throughout a lesson and should be intentionally designed as part of lesson planning. Student self-assessment and peer assessment can also provide evidence of student learning.
- Formative feedback, information intended to help a student get insight into what they can do and how to improve their work, can come from a teacher, peers or the individual student. Reflecting on and using feedback is one way to close the gap between intended learning goals and current status. A teacher may also make instructional adjustments for individual students, small groups or the whole class to support learning. A student may adjust their learning strategies based on self-reflection or feedback from the teacher or peers.

[20] Council of Chief State School Officers. (2018). *Revising the Definition of Formative Assessment.* Council of Chief State School Officers. https://www.michigan.gov/-/media/Project/Websites/mde/2018/07/09/Revising_the_Definition_of_Formative_Assessment.pdf?rev=109be7413a2f49ba9e4c5514e9d71ff6

[21] See, for example: Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(*1*). https://doi.org/10.1080/0969595980050102; Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

[22] Wiliam, D. (2018). *Embedded formative assessment* (2nd ed.). Solution Tree Press.

While formative assessment is regularly talked about in schools and districts, there are frequent misunderstandings about what it is. Sometimes it is reduced to frequently administered, short assessments such as a weekly quiz. While quizzes can provide evidence to address the question of where are learners right now, this limited view of formative assessment ignores other important aspects. When that is the case, the promise of improved student learning indicated from the research base is unlikely to be realized because the full range of formative assessment practices are not engaged.

Look-Fors

- Examine what is called formative assessment and how closely aligned the information is with moment-by-moment, day-to-day teaching and learning.[23] For example:
    - Are end-of-unit assessments or interims/benchmarks called formative?
    - Are score reports considered formative feedback?
    - As part of the umbrella of formative assessment, is attention paid to whether and how learning goals are communicated to students?
    - Are students provided with developmentally appropriate, scaffolded learning opportunities so that they learn how to engage with self or peer assessment and how to respond to feedback in meaningful ways?
    - Do teachers have opportunities to look at student work with their colleagues to talk about the evidence of what students currently understand and what they do not yet understand for planning next instructional moves?
    - When leaders talk about assessment with teachers, how much time is spent talking about standardized assessment procedures and score reports versus talking about formative assessment practices and learning from/responding to student work?

Section 4 vignettes that address this threat:

- Cañon City Schools recognized that to promote formative assessment practices, it needed consistent messaging about its importance. One way the district signals this is through the evaluation categories in its Instructional Program Reviews.
- Katy ISD realized that teachers should spend less time reviewing summative data reports and more time on formative assessment evidence to inform timely, targeted instructional moves.
- Connecticut DOE recognized the importance of assessment literacy in ensuring that formative assessment is appropriately valued.
- Oregon DOE provides professional development resources to support teachers in developing rich understandings of formative assessment.

[23] Heritage, M., & Wylie, E.C. (2020). *Formative Assessment in the Disciplines Framing a Continuum of Professional Learning*. Harvard Education Press.

- Wisconsin DPI and its CESAs provide targeted professional development for districts that includes formative assessment as an instructional tool.

For more information related to this threat, see *Reimagining Balanced Assessment Systems*:

- Chapter 4 (p. 113-115, Formative Assessment)
- Chapter 5 (p. 134, Classroom Assessment)

How to investigate and evaluate threats to coherence

In Table 3-3, we suggest some systematic ways to explore the coherence of local assessment systems. We suggest two high-leverage and high-impact strategies for approaching threats to coherence: one strategy is to develop an instructional vision that can then be used as a lens through which to examine the degree to which assessments support that vision. Since Threats 7-9 represent more subtle threats that can impact assessment practices, the second strategy focuses on explicit policies and implicit messages that exist or are perceived in a school or district.

1. **Create an Instructional vision.** It is difficult to identify inconsistency between assessments and an instructional vision without having gone through a process to develop the vision. In order to create an instructional vision that reflects current understanding of how students develop proficiency in a content domain, participants in the process may need to engage in some collaborative learning on the topic first.
2. **Examine Assessment Culture.** Explicit policies and implicit messages about summative assessment impact the assessment culture of a school or district. The Understanding Assessment Culture tool provides a way to examine these and their impact on assessment practices. More broadly, we suggest a couple of chapters from the 2024 NAEd volume that could be particularly beneficial for a book study group.

TABLE 3-3 Possible Strategies, Tools and Resources to Further Explore Common Threats to Coherence

Threats & Indicators	Strategies	Selected Tools and Resources
Threat 6: Inconsistency between assessments and instructional vision	● Create an instructional vision	● See Appendix B in this Guidebook ● How to Articulate an Instructional Vision During a Materials Adoption by EdReports ● Sample Instructional Vision Statements by EdReports ● The Improving Instructional System Coherence Toolkit by RAND

Threats & Indicators	Strategies	Selected Tools and Resources

Threat 7: Policies and politics that distort practice
 	
Threat 8: Over-emphasizing the role of summative assessment
 	
Threat 9: Under-emphasizing the role of formative assessment
	• Focus groups with students, teachers or leaders on explicit policies and implicit messaging related to assessment and assessment culture
	• Chapter or Book Study
		• Understanding Assessment Culture (tool)
		• *Reimagining Balanced Assessment Systems* by the National Academy of Education (open access book)
		  ○ Chapter 4: Classroom Activity Systems to Support Ambitious Teaching and Assessment
		  ○ Chapter 5: Assessment Literacy and Professional Learning
		  ○ Chapter 6: District and School Practices and Assessments to Support a Learning-Centered Vision
		  ○ Chapter 7: State Practices and Balanced Assessment Systems
		  ○ Chapter 9: Policy Influences on Ambitious Classroom Instruction, Assessment, and Learning

Practical Next Steps

What Should We Do First?

We started this Guidebook by explaining our rationale for the creation of this practical companion document to the 2024 National Academy of Education volume, *Reimagining Balanced Assessment Systems.* We wanted to empower district and school leaders to investigate their local assessment systems for common threats to imbalance such as inefficiencies, incoherence, and lack of usefulness. We also wanted to showcase some examples of districts and states who are working in areas connected to balanced assessment systems—whether directly or indirectly—and use those examples to illustrate some potential next steps for local or state education agencies. These examples can be found in Section 4.

But, at the end of the day, local education agency leaders may simply ask:

- What should we do first?
- How should we start or continue on this journey?

To answer these questions, we looked back to the *Practical Next Steps* we identified under each Threat to Balance category and pulled out a few high-leverage actions in Table 3-4.

TABLE 3-4 High-Leverage Actions to Explore Threats to Balance

IF you are concerned about the **efficiency** of your local assessment system because of over-testing or unused or redundant assessment data ...	**… *THEN*** we recommend you start with a local assessment system review or audit (see Table 3-1).
IF you are concerned about the **usefulness** of your local assessment system because of a mismatch between assessment design, purpose, and use, and confusion about why all tests don't directly inform instruction ...	**… *THEN*** we recommend you start with building the knowledge and skills of your leadership team so they better understand how assessments are designed and how that impacts the intended versus actual use of different types of assessments (see Table 3-2). Leaders are then able to work with teams of teachers and others to critique specific locally used assessments and their design in light of intended purpose and use.
IF you are concerned about the **coherence** of your local assessment system because of inconsistency between the assessments given to students and your district's mission or instructional vision …	**… *THEN*** we recommend you start with either creating or reviewing your district's instructional vision and evaluating the alignment among the curriculum materials, instructional practices, and assessments administered (see Table 3-3 as well as Appendix B and Appendix C).

IF you are concerned about the **coherence** of your local assessment system because policies seem to be shaping practices in ways you don't think are right, or there is an outsized focus on summative assessments …	**… *THEN*** we recommend you learn more about assessment culture and evaluate whether systemic shifts are needed to support the implementation of best practices (see Table 3-3 and Appendix A).

These actions are starting points, not destinations. But we believe that these high-leverage actions will help local education leaders, and by extension state education leaders, know how to take what has been a more conceptual and theoretical conversation about balanced assessment systems and turn it into practical next steps. Start with your most pressing concerns. If you are concerned about all of them equally, we recommend starting with an assessment system review or audit and embed in conversations about assessment design, purpose, and use, as well as coherence with your district's instructional vision to make recommendations about what assessments remain, need to be revised, or need to be removed.

Who Should Be Involved?

Up to this point we've specified that the audience for this Guidebook is primarily district and school leaders. But leaders cannot enact change on their own. They need to gather a group who can provide different perspectives and insights based on their role in the local educational system. We strongly recommend including classroom teachers and educators who represent the interests of students with disabilities and English language learners along with building and district administrators. Gathering a diverse group will ensure that those who are tasked with using student assessment results have a voice.

What Should We Do Next?

Using the high-leverage actions to explore threats to balance (see Table 3-4), are *reactive* actions district or school leaders can take to start looking for where imbalances might be in their assessment systems—whether it be from inefficiencies, inadequate assessments, or incoherence between assessments and other parts of the system. The actions will help leaders understand what is happening, why, and if it is what is desired or identified as important.

Similarly, so is knowing how to *proactively* create systems and structures that support rich classroom learning environments and how to right-size the role of assessment within those environments. When districts and schools intentionally ground assessment decisions and practices in an instructional vision (see Appendix B) and examine their assessment culture (see Appendix A), they are promoting coherence among curriculum, instruction, and assessment (see Appendix C). Further, an advantage of making progress toward a more balanced assessment system—especially when that involves removing redundant or unnecessary assessments from a

local system—is that it can free up instructional time, as well as teacher and student energy, to focus on instructional resources and assessment practices that advance student achievement.

Research suggests that formative assessment practices (when implemented well) can have the single largest effect on student achievement of any other school-based intervention.[24]

> "Formative assessment is a planned, ongoing process used by all students and teachers during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended disciplinary learning outcomes and support students to become self-directed learning."[25]

Districts and schools cannot go wrong in focusing professional learning, coaching, PLCs, and school improvement efforts on strengthening and supporting educators' formative assessment practices. These practices are the moment-by-moment, day-by-day observations, informal probes, sharing of success criteria and learning goals, evaluating responses to oral and written questions during instruction, self-reflections, peer and teacher feedback, and so on, where assessment information is immediately used to gauge a student's level of understanding and adjust teaching and learning strategies. There are many ways this can be accomplished; however, we provide an annotated list of a couple open-source professional learning resources that can help move practice forward in Table 3-5.

TABLE 3-5 Open-Source Professional Learning Resources on Formative Assessment Practices

Name of Resource on Formative Assessment Practices	Description of Contents
Michigan Assessment Consortium https://www.michiganassessmentconsortium.org/assessment-resources/formative-assessment/	• A collection of videos and short articles developed and collected by the MAC over multiple years. There are also learning modules that require a fee to access.
Oregon Formative Assessment for Students and Teachers https://ode.instructure.com/courses/122	• Five modules on formative assessment practices with narrated presentations, video and reflection guide, activities, etc.
Center for Assessment, Formative Assessment Resources and Math Learning Progressions https://sites.google.com/view/cfa-formative/home	• Compendium of videos, slides, readings, and resources on formative assessment practices. • Walk-through protocol for observing formative assessment practices and giving group feedback. • Connections to math learning progressions with slide decks and examples.

[24] See, for example: Black, P., & Wiliam, D. (1998) and Hattie, J., & Timperley, H. (2007).

[25] CCSSO. (2018). *Revising the Definition of Formative Assessment.* Council of Chief State School Officers. Washington, DC. https://ccsso.org/resource-library/revising-definition-formative-assessment

What is Not in This Guidebook, But is Also Important?

This Guidebook is not a "How to create a balanced assessment system from scratch" guide, nor does it cover every important topic. We chose to focus on observable threats to balanced assessment systems, which could serve to show district and school leaders the areas where they may need to investigate the efficiency, usefulness, and coherence of their local assessment system; that is, where their local system may be less balanced. We provided practical next steps to support those investigations and to lead, hopefully, to new design and implementation decisions.

Our chosen focus means that important topics such as increasing the **assessment literacy** of our policymakers, building and district administrators, teachers, parents and students are discussed only at a cursory level in this Guidebook. We know that lack of assessment literacy is one of the key barriers and factors that impede the design and implementation of balanced assessment systems (see Section 2: Barriers to Balanced Assessment Systems). We do provide open-source materials in Table 3-2 that can support professional learning related to assessment design, selection, interpretation, and use—all foundational topics under the umbrella of assessment literacy.

An equally important topic that does not receive deep attention in this Guidebook is the importance of **assessment quality**. Assessment quality affects the inferences we make about students and our confidence that assessment results are an accurate and adequate reflection of what each and every student knows and can do. Key considerations related to assessment quality, from state-level assessments to classroom-level assessments, include:

- Designing the assessment to match the intended purpose and use of the assessment
- Aligning the assessment to the depth and breadth of the identified content standards
- Designing the assessment to be fair and accessible to all learners
- Scoring the assessment using clear criteria or rubrics

SECTION 4: DISTRICT & STATE VIGNETTES

Background

After creating our list of the most common threats to balance, we wanted to check our understanding against others who we know are working in this area. We selected a purposive sample of district and state education agency leaders and interviewed them for about one hour to get their perspectives on two key questions:

1. Do you agree with these common threats to balance? What do you think is missing, erroneous, or needs to be revised?
2. What have you done as a district or state leader (or worked with others to do) to address one or more of these threats to balance? How did you do that, and what was the result?

We do not claim that this purposive sample is representative or generalizable; however, we believe it provided us with valuable insights that helped refine our original list of threats and offered examples of how districts and states are approaching balanced assessment systems.

We don't expect any example to show a perfectly balanced system—balance exists along a continuum, after all; it is not an on/off switch … but we hope with a wide range of examples, districts or schools can identify areas they could tackle to improve balance in their assessment system.

The goal of these interviews was to capture real-life experiences and stories, infusing what is often a conceptual or theoretical conversation about balanced assessment systems with practical examples written as vignettes. We do not expect any example to show a perfectly balanced system—balance exists along a continuum, after all; it is not an on/off switch. We also do not expect any one example to be comprehensive and cover all threats, but we hope with a wide range of examples, districts or schools can identify areas they could tackle to improve balance in their assessment system. We also hope the state vignettes will illustrate different approaches to supporting local actions, including the tradeoffs and tensions across various approaches.

We chose to include both district- and state-level examples because state actions related to assessment and accountability can have an outsized influence on local assessment systems. For example, states control the design of and messaging about state assessments, as well as the types of guidance, tools, and/or resources offered to local education agencies related to all types of assessment (formative, interim, summative, diagnostic, universal screeners, etc.). We include district vignettes because districts control decisions about many areas that directly affect the balance of an assessment system. For example, districts are the locus of control for

- local curriculum, instruction, and assessment decisions, including school- or district-required assessments and classroom assessments;
- selection and implementation of professional learning opportunities for teachers;
- local assessment policies and practices (e.g., grading policies); and
- structures that support assessment interpretation and use (e.g., PLC structures; instructional coaching models; student work analysis or other types of data analysis protocols; leadership support).

Below, we try to capture the stories of our district and state examples and connect their stories to the common threats to balanced assessment systems. Many of the threats are interconnected and not entirely independent, but we selected certain threats to highlight because they were most apparent in our interviews. The vignettes are organized by starting with the district leaders and then moving to the state leaders, incorporating selected quotes and key ideas from the interviews.

We recommend two approaches to reading the vignettes: read them all the way through from district to state, or use Table 4-1 to select specific threats that interest you the most. We provide a summary of key themes after each set of vignettes.

TABLE 4-1 Crosswalk Between Vignettes and Common Threats to Balance

	CCS	CPS	CUSD	HSSD	Katy ISD	CT	OR	PA	WI
Efficiency									
Threat 1: Too much testing overall, particularly early or later in the year			✓		✓	✓	✓	✓	✓
Threat 2: Redundant assessments					✓	✓	✓	✓	✓
Threat 3: Unused assessment results				✓	✓	✓		✓	✓
Usefulness									
Threat 4: No clear match between the assessment purpose, design, and use		✓			✓		✓	✓	✓
Threat 5: Assuming all tests can inform instruction					✓		✓		
Coherence									
Threat 6: Inconsistency between assessments and instructional vision	✓	✓		✓			✓		
Threat 7: Policies and politics that distort practice		✓				✓	✓		
Threat 8: Over-emphasizing the role of summative assessment	✓		✓				✓		✓
Threat 9: Under-emphasizing the role of formative assessment	✓				✓		✓		✓

Notes: CCS=Cañon City Schools; CPS=Chicago Public Schools; CUSD=Cupertino Union School District; HSSD=Howard-Suamico School District; Katy ISD=Katy Independent School District; CT=Connecticut State Department of Education; OR=Oregon Department of Education; PA=Pennsylvania Department of Education; WI=Wisconsin Department of Public Instruction.

Cañon City Schools

- Adam Hartman, Superintendent

Introduction

Cañon City Schools, in Colorado, has about 3,500 students. Like several of the other districts featured in these vignettes, Cañon City Schools has adopted a Portrait of a Graduate which guides their work. However, in this vignette we focus on how the district has developed a shared understanding of the purpose of school and how the adults in the system should function. These core beliefs then influenced what happens with teaching and assessment across all the schools. In this vignette, we focus on the way the district has proactively documented and evaluated instructional quality through the annual Instructional Program Review, as a means of communicating expectations about how those beliefs should show up in classroom instructional and assessment experiences. This example showcases one approach to helping schools consistently and coherently link their instructional vision to their assessment system and employ district policies in ways that support best practices in assessment.

Threat 6: Inconsistency between assessment and instructional vision

The district has a simple but powerful vision: Learning for Life! District Superintendent Adam Hartman provided one example that motivated his understanding of why the district needed a clear vision for the purpose of school. In a previous role, he observed a middle school math teacher who was struggling to connect the lesson to real-world applications. He noted,

> *"She had a beautiful window in her classroom, and she used to go over there and tell the kids, 'look, the real world is right out there.' And it was really bothering me that we created this distinction between what was happening in the classroom and the idea that someday, somewhere, you would be applying these things."*

From this concise vision statement, the district developed a set of four core beliefs:

- Core Belief #1: We meet the social-emotional needs of all students, putting Maslow's Hierarchy before Bloom's Taxonomy.
- Core Belief #2: We believe growth is what matters most. Learning requires risk-taking, and the work we do has the greatest impact on student outcomes.
- Core Belief #3: We are future-focused, committed to developing the traits and skills that prepare students for ever-changing careers.
- Core Belief #4: We emphasize what is good for kids over the needs and comfort of staff.

The second core belief—in a student's growth as what matters most—influences teachers' attitudes towards assessment. Since the goal is to help every student grow from wherever they are, teachers and students need to engage in meaningful formative assessment practices to understand what students can do so that next learning steps can build from there.

Instructional Program Reviews

The Instructional Program Reviews take place annually for each school and are the district mechanism for evaluating and monitoring quality across schools. It intentionally models Core Belief #2 as a continuous improvement process. The process itself communicates what the district values. In preparation, school leadership engages in self-reflection using a set of common rubrics and collates evidence and artifacts to share with the visiting review committee. The committee visits each school for a full day and meets with staff and students to review the information they have compiled and directly observe classroom instruction. Results are reported back to the school by the end of the school year, and each school's report for the past four years is publicly available on the district website.

Each school is rated on a four-point scale: ineffective, somewhat effective, effective, and highly effective on nine categories:

I. Focus on Climate, Culture, Vision, and Purpose
II. Focus on Student Health and Wellness
III. Focus on Innovative Instruction
IV. Focus on Equity of Opportunity
V. Focus on Opportunities for Learning Experiences
VI. **Focus on Effective Assessment Procedures**
VII. Focus on Improvement of Instruction and Learning
VIII. Focus on Support for Positive Student Behavior
IX. Focus on Resource Acquisition and Maintaining a Safe Learning Environment

Each of these categories has at least five indicators. For example, the nine indicators associated with Category VI: Effective Assessment Procedures are:

- Sound practices for student assessment and measuring the development of key skills & traits are evident.
- Teachers create and provide rubrics and exemplars to students.
- Teachers analyze assessment results to revise and improve instruction, with a primary focus on growing student achievement.
- Teachers use assessment data to engage students in the creation of student-centered growth targets.
- Assessment is conducted in a variety of ways, including formative, summative, project-based, and Capstone (applies only at the HS level).
- When available and applicable, teachers use common assessments for given subject areas.

- Students have opportunities to show what they have learned through collaborative participation, and in non-traditional ways.
- Time, resources and support are provided for teachers to analyze student learning progress together.
- Prompt, formative feedback is continually offered to students, aimed at growth, and occasionally offered by non-educators.

Each indicator has a rubric and suggested questions that the panel can ask students and staff to elicit evidence for that indicator. For example, Table 4-2 below shows the rubric and questions for the final indicator of Category VI.

TABLE 4-2 Rubric and Questions for One Indicator in Category VI

Indicator	**Ineffective**	**Somewhat Ineffective**	**Effective**	**Highly Effective**
Prompt, formative feedback is continually offered to students, aimed at growth, and occasionally offered by non-educators.	No evidence of formative feedback being given to students.	Some evidence of formative feedback is given to students. It is occasionally offered and is not impactful.	Clear evidence of formative assessment is happening daily in classrooms. Teachers are familiar with multiple formative assessment practices.	Teachers are familiar with, and use on a regular basis, formative assessment practices that drive instructional decision- making processes. Feedback is offered daily or when necessary more often during class.
Questions you might ask staff and students:	● How do your teachers measure what you are learning? ● How do you quickly check for a student's understanding of a particular subject? ● When students are unclear of the subject matter, what is your next step?			

The final report from the review committee for each school is about 40 pages long. For each of the nine categories, it includes the school reflection, reviewer comments, the indicator scores, and specific suggestions or directions for the school to consider implementing related to that category.

> Threat 8: Over-emphasizing the role of summative assessment
> Threat 9: Under-emphasizing the role of formative assessment

We highlighted the Assessment section of their Instructional Program Reviews given the focus of this Guide. The collection of nine indicators related to effective assessment procedures together serve to ameliorate Threats 8 and 9: an over-emphasis on the role of summative assessment and an under-emphasis on the role of formative assessment. For example, formative assessment practices are valued in the district as an important part of an assessment system and this is communicated from their Core Beliefs through to explicit inclusion in the Instructional Program Reviews. This throughline supports consistency of understanding and allows for these practices to withstand leadership change because they are integral to how the district operates.

Clearly, the district leadership has made a significant time commitment to develop the categories for the review, together with the indicators and rubrics, to conduct the day-long observations in school, and to compile the final report for each school. However, reading across several reports focusing primarily on Category VI (Effective Assessment Procedures), it is clear that there are many practices that show up across multiple school reports, such as the role of rubrics and exemplars in assessment, or the value that teachers place on student self-reflection. Hartman commented on this process:

> *"It is the biggest piece of work that I do every year. It is the most valuable work that we do every year. Every year I think, oh, gosh, we've got to do all the IPRs. And every year I put it to the leaders, and they said, do not take that away, because it helps keep us focused on all these things."*

Discussion

This vignette from Cañon City Schools illustrates quite a different approach to assessment system improvement, in part because the assessment system is explicitly one part of a much larger educational system. The district has a clear vision and set of core beliefs around the work of a district in supporting and improving school quality, and has further articulated those beliefs into a set of categories that are seen as sufficiently important to warrant annual reflection and review. The Instructional Program Review rubrics provides a common framework, language and expectations that are shared across school- and district-leadership and educators.

The development of this process took nine years, is adjusted and improved continuously. It requires an annual commitment to the process, including training and calibrating the members of the review panels who conduct the observations. Reading across reports from four years ago to the current year reveals incremental improvements.

While the review process does not guarantee that consistency across all schools in terms of assessment and instructional vision in the Cañon City Schools, it does mean that school and district leaders can identify where there are inconsistencies, which in turn allows them to address them, whether through professional learning opportunities, policy changes or other adjustments.

Chicago Public Schools

- Peter Leonard, Executive Director of Student Assessment and MTSS
- Dr. Sasha Klyachkina, Skyline Assessment Manager

Introduction

Chicago Public Schools (CPS) is the largest school district in Illinois, and one of the largest in the U.S., serving approximately 325,000 students. In this vignette we focus on the motivation and circumstances that led to the development of free, high-quality curriculum materials for PK-12 (called Skyline) that includes curriculum-embedded assessments. The district has invested significant time, money, and resources into Skyline and continuously improves the materials and associated professional supports over time. For Peter Leonard and Sasha Klyachkina, starting with an instructional vision has been key to Chicago's approach. This vignette illustrates how high-quality instructional materials with embedded assessments can promote coherence, efficiency, and usefulness.

Converging Factors

Starting Point:
Threat 4: No clear match between the assessment purpose, design, and use

Threat 6: Inconsistency between assessments and instructional vision

Threat 7: Policies and politics that distort practice

Klyachkina and Leonard described a number of converging factors that led to the creation of the district-created curriculum. First, before the work began, CPS used a commercial interim that was administered multiple times per year, and the data were used in many high-stakes ways such as a component of teacher evaluation, part of school quality metrics, and at certain grades levels the data were used as part of grade promotion or selective high school admission decisions. CPS district leadership was aware of the significant efforts being made by school leaders and teachers to align instruction with this assessment because of the consequences associated for students, teachers, and schools. Because of the way in which these data were being used, the district decided to discontinue use of this interim assessment.

Second, at the same time, there were national conversations about the need for high-quality instructional materials to provide all students with an equitable opportunity to learn grade-level standards. For example, reports, such as *The Opportunity Myth* from The New Teacher Project, highlighted the paucity of grade-appropriate student assignments experienced by students, particularly for students of color, those from low-income families, those with mild to moderate learning disabilities, and English language learners.

Third, district staff were aware that student experiences between schools and even within schools were not the same when it came to coverage of standards and content. To more fully understand these disparities and uneven student experiences, 500 CPS teachers were asked if their school provided curriculum in the area that they teach. In 2019, almost 50% responded that their schools did not provide these resources. Survey results also indicated that more than 35% of teachers were spending five or more hours per week searching for instructional resources.
It was within this convergence of local and national factors, that CPS leadership decided to take action. As Leonard explained: *"The development of Skyline was very much a recognition that the district needed to play a more active role in ensuring a focus on the instructional materials, the learning experiences connected to that, and assessment experiences that were supportive of that vision."* As such, in 2019, CPS decided they first needed to develop an instructional vision and then create free curriculum materials that would be aligned to that vision, with the goal of raising the floor on the educational experience of all students across the city. Currently, curricula have been developed PK-12 for English language arts, artes del lenguaje del español, mathematics, science, social science and world language (Spanish, French).

The Skyline Curriculum

The Skyline curriculum involves curriculum-embedded assessment materials at the lesson, unit, and interim levels, as well as formative assessment opportunities within lessons. These assessments are embedded within the scope and sequence of learning, which was critically important for CPS since they had already experienced what can happen when assessments are standards-aligned, but not instructionally aligned. Leonard put it this way, *"A curriculum is more than just a collection of instructional materials, and assessment is a part of our definition of high quality curriculum."*

Klyachkina also explained the rationale for the integration, *"If we are providing high-quality instructional materials but the assessments don't match that vision … [and] if teacher teams are engaged in the practices we want them engaged in and they're backwards planning from their end of unit assessment, coming together in teacher teams to analyze those assessments to inform instruction. If the assessments don't actually support that instructional vision, one of the things is going to go, right?"* She elaborated by noting that either teachers will compromise the instructional vision and reduce the rigor to the level of the student assessments, or they will go back to trying to create their own assessments which is not a good use of their time, nor always results in high-quality assessments.

The curriculum materials themselves play an educative role for teachers by having assessment opportunities embedded in instruction and including formative opportunities such as journaling, teacher observations, discussion questions, exit tickets, all of which are clearly intended to support within and between lesson adjustments. Unit assessments support backward planning

and help teachers and students gauge progress towards mastering grade-level standards. Finally there are interim assessments, aligned to the curriculum's scope and sequence, which help teachers evaluate transfer to learning across units.

Supporting Coherent Implementation

In a district as large as CPS, coherent implementation of any new policy or resource is dependent on the culture and leadership within school. The district leadership has worked on two fronts to address assessment system coherence. *From a design perspective*, they have codified their intentions and requirements for all assessments, as well as individual assessment types, into an Assessment Design Guidance document that is used by their curriculum design teams who create the materials and content teams who review and provide feedback and professional learning. This supports enactment because any designer creating curriculum (instructional activities and embedded assessments) for any content area or grade level is working with the same set of criteria, definitions, expectations, design tools, and examples for what is expected and ideally using them in the same design process anchored in Understanding by Design. There is also a five-part Assessment Design Professional Learning Series to support designers in internalizing and applying the guidance to their work.

From the implementation perspective, CPS has a series of school leader, coach, and teacher professional learning opportunities. Teacher learning is content and grade-level specific to provide space for teachers using the same materials to collaboratively plan for how they will use the different assessments before, during, and after instruction. They also have interpretation and use protocols for lesson, unit, and interim assessments that are intended to be used within teacher or leader professional learning communities. There is also an implementation framework for high-quality curriculum that school leadership teams can use to support the school-level practices and resources for strong implementation that will lead to engaging, rigorous and relevant instruction for students. This resource is intended to support teacher teams engaging in continuous improvement cycles within their buildings.

Discussion

This vignette from CPS illustrates several things that will be true regardless of district size. First, assessment reform should not occur without conversations about instructional vision. Otherwise, decisions about assessment become untethered from discussions about disciplinary learning and curriculum-instruction-assessment coherence. This can result in assessments that lead to pernicious effects on teaching and learning such as teaching to the test rather than teaching to the standards. If assessments do not reflect and support the instructional vision, one of them is likely to be compromised. Second, assessment reform cannot be a one-year endeavor but requires leadership to maintain a focus over time. Instantiating that focus in documents like the instructional vision and assessment design guidance can help ensure that it can outlast leadership

changes. And, finally, Chicago's story highlights the continued struggle with the decentralized nature of education in the United States and the impact of local control on high-quality instructional material implementation. CPS cannot mandate the use of Skyline—it is an optional curriculum, albeit free, but at the time of writing, 470 out of 519 CPS schools (91%) have adopted the Skyline curriculum in at least one content area and grade band. State education agencies often face similar challenges: they can tie state funds to the purchasing of only high-quality instructional materials, but they have little other control over local actions in this space.

Cupertino Union School District

- Allison Liner, Associate Superintendent

A District Experiencing Change

Like many districts in the United States, the TK-8 Cupertino Union School District (CUSD) in California has experienced a significant decline in overall student enrollment, while also experiencing an increase in the percentage of low-income families since COVID. Situated in the heart of Silicon Valley, with many parents working in the tech industry, CUSD has a diverse student population that speaks more than 45 different home languages. The district's history of academic excellence is a draw for many families.

Influencing Change

District leadership has been paying attention to calls for supporting the development of future-ready learning, along with results from the California Healthy Kids Survey and other evidence that shows that students want to engage with things that matter in school.[26] They also are aware of the requirements and expectations related to the state accountability assessments. On one hand, state results are a publicly reported metric used to judge student progress in English language arts, mathematics and science, but as Deputy Superintendent Allison Liner described it, *"We know kids will be asked to do more to show their understanding, and more importantly, be asked to solve unpredictable problems in the future. While state assessments capture knowledge from a singular moment in time, implementing future-ready skills is three-dimensional: infusing not only academics, but key life and learner competencies."*

Threat 8: Over-emphasizing the role of summative assessment

As part of CUSD's most recent strategic planning work, leadership solicited input from hundreds of parents, teachers, staff, students, and community members, which led them to articulate a Portrait of a Learner, which identifies seven competencies that CUSD will cultivate in their students.[27] The Portrait of a Learner has helped district leadership articulate the important durable skills or competencies that they want students to develop in conjunction with knowledge and skills articulated by their state standards.

[26] California Healthy Kids Survey: https://www.cde.ca.gov/ls/he/at/chks.asp

[27] CUSD's Portrait of a Learner: https://www.cusdk8.org/departments/educational-services/portrait-of-a-cusd-graduate

Implementing Change With Bottom-Up Piloting Efforts

The approach CUSD took to integrating the competencies in the Portrait of a Learner into the day-to-day work of teachers has been focused on supporting a group of fifty "accelerator" teachers and principals. This group of educators agreed to work together to flesh out descriptions of each of the seven Portrait of a Learner competencies and then to create one-page documents that describe the learner characteristics of each of the competencies in more detail. The accelerator teachers also provided input on progressions of learning that describe the continuum of how each characteristic develops over time.

CUSD also provided professional learning opportunities for the accelerator teachers, with a focus on project-based learning (PBL), using two complementary, but distinct approaches: PBL Works at the elementary level, and Defined Learning Tasks at the middle school level.[28]

Interestingly, Liner noted that while she thinks *"we're a couple years out from any district wide assessment changes,"* there is an organic growing interest in PBL in the district as a result of colleagues watching the success of the accelerator teachers. Liner noted that they had a PBL summer professional learning session, and due to the interest level had to open up a second session.

Working on Other Aspects of the Assessment System

Liner said that while they are supporting the work of the accelerator teachers, they are also cognizant of the assessment system as a whole and the impact that it has on teachers and students. CUSD has been evaluating their assessments based on teacher concerns about insufficient instructional time, and looking for opportunities to eliminate assessments, or in some cases better understand the purpose and use of some assessments. Some additional assessments were added when schools were remote due to COVID. While those assessments served them well during that period, district leaders are reexamining whether they are still necessary.

Threat 1: Too much testing overall

In reading and mathematics, CUSD uses a benchmark assessment that also provides personalized learning playlists for students based on their results. District leaders discussed with teachers whether the personalized playlists were necessary and were considering removing them. However, some teachers found them useful, so CUSD decided to make their use optional as a site-level decision. Sites that choose not to use this tool are expected to utilize other methods to ensure personalized learning is occurring for students.

[28] PBL Works: https://www.pblworks.org/; Defined Learning Tasks: https://definedlearning.com/defined-learning

CUSD is currently examining the suite of assessments that they are using in English language arts. Given the instructional shift to structured literacy, they want to make sure that teachers have the assessment information that they need. This fall, a new state-required reading screener will be implemented in grades K-2 to ensure that students who need extra support receive early intervention.

Last school year, a group of teachers and principals piloted a student-driven assessment designed to measure state academic standards as well as Portrait of a Learner competencies. Teachers tested the assessment over a period of days in their classrooms as a beginning step towards including a "cornerstone" assessment in the district's assessment suite. This work will continue with an expanded pilot this school year, with the aim of adding a new assessment to the district's assessment system in future years.

Conclusion

The story of the Cupertino Union School District is similar to that of Howard-Suamico School District. Both have adopted a portrait of a learner/graduate to help them focus on the kinds of durable skills that reflect their local community's beliefs about the purposes of schooling and the types of skills they want students to develop. Cupertino Union is taking this work slowly, with an initial investment in the learning and work of a cohort of accelerator teachers who are innovating around project-based learning and assessment as a means to developing and measuring these skills. This group is well-placed to help district leadership identify challenges such as too many or redundant assessments, so that strategic decisions can be made that will smooth the way for a district-wide adoption.

Howard-Suamico School District

- Amanda Waldo, Director of Teaching and Learning

Introduction

Howard-Suamico School District (HSSD), is a small district in Wisconsin just outside Green Bay. It enrolls nearly 6,000 students. By Wisconsin standards, the district is in the top 25%. District leaders know that students need to be both adaptable and innovative in order to thrive in a changing world, particularly in the context of a digital transformation. That recognition led the district into a journey that began in 2012 with work on the 4Cs of 21st-century skills (critical thinking, creativity, collaboration, and communication). In 2017, that work evolved to be more student-centered and the district developed its north star called the Graduate Profile with seven skills that all students now work towards (being a self-starter, a critical thinker, a collaborator, a communicator, adaptable, responsible, and a solutionist). In this vignette we pick up the story of HSSD and its Graduate Profile, particularly from the perspective of assessment.

Measuring the Right Things: Making Informational Meaningful to All

With the shift to the Graduate Profile, the district recognized the importance of not just *describing* the seven skills that they wanted to see students develop, but also of *measuring* those skills. The school board was taking an active interest in how students were developing the skills outlined in the Graduate Profile. Initially, the district worked with an external organization and used its assessments to provide information about student progress on the seven skills. The assessment data for students in grades 4, 6, 8, and 10 was reported to the school board and district-level leadership, but teachers and school leaders were not deeply involved in analyzing results at the classroom level. In addition, the initial assessments measured skills in a way that did not connect to the curriculum frameworks which further exacerbated the disconnect between what teachers were doing day-to-day in the classroom and these summative assessments of skills of the Graduate Profile. As Amanda Waldo, Director of Teaching and Learning, reflected on this period of time in the district, the two threats that best reflected their situation were Threat 6: Inconsistency between assessment and instructional vision and Threat 3: Unused Assessments (although in this case, perhaps "under-used" assessments since some interest-holders, but not the teachers, were examining the data).

> Threat 3: Unused assessment results
> Threat 6: Inconsistency between assessment and instructional vision

To address these threats, or challenges, district leaders realized that the assessments of the Graduate Profile skills had to be directly connected to the curriculum frameworks. Rather than using assessments developed by an external agency, the district brought teams of teachers from

grades 4, 6, 8, and 10 in order to examine their existing common summative assessments. The goal of these meetings was to identify opportunities for adjustments that would allow for purposeful inclusion of one or more of the Graduate Profile skills while still maintaining the assessment of an already-identified aspect of the curriculum framework.

Waldo gave an example of how the grade 6 math team modified a task from a unit within its adopted math resource on statistics and probability. The original task provided data from a radar-speed machine and showed speeds for 20 drivers. Students were asked to calculate the mean, median and mode of the data set, explain their work, to consider the impact of outliers, and to draw a box plot. The task required students to show their work and explain their thinking, but the math was decontextualized and it did not align with the vision of a HSSD assessment that emphasizes authentic learning and a demonstration of the Graduate Profile skills.

As the team members reviewed this task, they wondered why they would not use drivers' speeding information from the road directly in front of the school and connect the mathematics to students' everyday experiences. The revised task now involves partnering with the local police department to collect real data, and having students use statistical reasoning skills to decide whether there are safety issues, and, if there are, to design community recommendations. In addition to the math standards being assessed, the task now explicitly identifies three Graduate Profile skills (critical thinking, problem-solving, and responsibility within the community) with "I Can" statements that describe the student success criteria. The team also developed a rubric that addressed both the mathematical and the Graduate Profile skills.

Using a series of performance tasks over the course of a year, students and teachers get a stream of information about progress towards the seven skills and learning in terms of curriculum frameworks. By integrating the assessment of the Graduate Profile skills into curriculum framework assessments, the vision of learning (Graduate Profile), the vision of teaching (content frameworks), and assessment are much more aligned, and the assessment information is supporting improvements both in how students learn content and how they develop the skills. The revised assessments are no longer isolated from the everyday teaching, learning, and assessment in classrooms. The teacher guidance for the performance assessment includes the following note:

> *Prior to the assessment and throughout the year, provide students with real-world challenges in which they actively participate in the problem solving process, within and across disciplines. Provide a variety of information sources/documents that students read and evaluate in order to draw conclusions and provide evidence to support an option or claim.*

In other words, there is an explicit expectation that teachers are using tasks and activities in their instruction that enable students to connect their learning to real-world problems and to support the development of the Graduate Profile skills.

Extending the Work

The district is now looking to extend this assessment approach to the adjacent grades to encourage all teachers to attend to the Graduate Profile skills and to provide them with assessment information that will help them do so in a meaningful way.

Waldo acknowledged that the upper end of high school brings unique challenges, as students feel the pressures of ACT scores, GPA, and college applications. Although these pressures do not come from district leadership, they remain part of the reality that students face. To better prepare students for this, the district explored various ways to support students' demonstrations of growth. The district has now decided on a junior-year capstone experience. In this capstone, students showcase their growth through the Graduate Profile while practicing real-world skills such as interviewing, writing college essays, and/or presenting to authentic audiences, equipping them with confidence and readiness for their unique pathway after high school.

Conclusion

The work in HSSD illustrates one district's approach to promoting more consistency or coherence between its instructional vision and its assessment system. The district uses performance tasks and embeds Graduate Profile skills within content-based assessments. This assessment approach signals what is valued in classroom teaching and learning, which allows teachers to shift their instructional emphasis and practices. Engaging teachers in the redesign was an important aspect of this work.

Katy Independent School District

- Dr. Sanee Bell, Assistant Superintendent of Teaching and Learning
- Dr. Christine Caskey, Chief Academic Officer
- Natalie Martinez, Executive Director of School Improvement

Introduction

Katy Independent School District (ISD), is a large and diverse district in Texas just outside the city of Houston. Their story is a work in progress, but provides an example of how working on one set of issues (or threats) related to their assessment system led to deeper understanding of the challenges and opportunities with other threats. Katy ISD's example highlights the importance of deep engagement with a set of questions about what was being assessed, for whom and for what purpose.

Beginning the Journey: Threats to Efficiency

Leaders in Katy ISD reached out to the Center for Assessment for support to improve their assessment system. The goals that they identified at the first meeting included:

- Reducing the testing burden on students
- Ensuring that assessment data is actionable
- Establishing consistent approaches to assessment practices within and across grade levels, schools and content areas

Starting Point
Threat 1: Too much testing overall
Threat 2: Redundant assessments
Threat 3: Unused assessment results

Center staff led four grade-level teams through an assessment system review process: one grade per grade band (K-2, 3-5, 6-8, and 9-12). The assessment system review was conducted over 2.5 days, using a combination of whole-group training and grade-specific work. The goal of the whole-group training sessions was to ensure a common understanding of assessment terminology as applied to the Katy ISD system and provide a basic level of assessment literacy to support the evaluative activities embedded in the assessment system review. The goal of grade-specific work was to provide time for central office staff, school leaders, and classroom teachers to work together to inventory and evaluate the collection of assessments experienced by a general education student over the course of the year in the four core content areas (and foreign languages for high school). The inventory prompted the teams to capture the primary intended user, use(s), and time to administer each assessment in the inventory before evaluating the extent to which those identified users, use(s), and time was realistic, feasible, or appropriate.

During the Process: Threats to Efficiency and Coherence

Each team came to similar realizations as they reviewed their inventory and responded to a series of reflection questions: They were surprised by the total number of testing minutes per year, and found the quantity of assessments in the first six weeks of school particularly noteworthy. Reflecting on this initial learning, Executive Director of School Improvement Natalie Martinez said, *"Our inventory showed that we had more tests than there were days in the month."* In addition, the leaders began to see that the volume of assessment data could not possibly all be used. As Assistant Superintendent of Teaching and Learning Sanee Bell noted, *"We were going to give a beginning of the year [assessment], then another one two weeks later ... for what?"*

Another insight that arose from this process was the way in which policies have unintended distorting impacts. In Katy ISD, we heard from teachers about the requirements in the grading and reporting handbook that dictate a certain number of major, minor, and other grades to be added to the learning management system/gradebook per six-week cycle (secondary schools) and nine-week cycle (elementary schools). This policy constrained participants thinking about what assessments could be removed from a grade-level assessment inventory. Bell summed it up this way: *"... [I]t's hard to get people to think beyond what they've done because that's been the **rhythm** of assessment, as opposed to what is the **reason** of assessment."* This aspect of the Katy ISD journey is an ongoing one, as the leadership team who brought us in to conduct the assessment system review is not responsible for the grading policies. However, they are continuing to collaborate across departments and are working for revisions to these policies.

After the Process: Threat to Usefulness

During the assessment system review process, each grade-level team was asked to consider the collection of assessments in their inventory in light of the district vision and mission statements which describe the purpose of Katy ISD. The teams recognized that some of the purpose of schooling was not supported by a reliance on primarily multiple-choice assessments. This learning was reinforced after the assessment system review when the leadership team read Evans' and Marion's book, *Understanding Instructionally Useful Assessment.*[29] Bell captured their learning by saying, *"I felt like we were trying to solve a what and a how problem without really deeply understanding and being able to articulate the why and so we needed some common understanding, some common language, so we could speak to why a balanced assessment system."* The leadership team is now developing shared and common understandings about assessment purpose and design and what is

Where the Work is Now
Threat 4: No clear match between the assessment purpose, design, and use
Threat 5: Assuming all tests can inform instruction
Threat 9: Under-emphasizing the role of formative assessment

[29] Evans, C., & Marion, S. (2024). *Understanding instructionally useful assessment*. Routledge.

required of assessments to inform instruction in order to differentiate and make decisions about which assessments are needed, by whom, and for what purpose. This work is beginning with principals at the elementary level, and using assessment literacy as a mechanism to enable them to be stronger instructional leaders.

Bell described conversations she has had with leaders in which the focus has been shifting from summative data to the classroom formative assessment process information:

> *"I don't want us spending time writing lengthy assessment reports or analyzing endless district-level data points. Our focus should be on working with teachers, building their skills so they can use formative assessment to make timely, targeted instructional moves. From the C&I perspective, at both elementary and secondary levels, we'll help teachers see they don't need ten different tests to know if a student has learned the material. The goal is for them to know, based on quality formative assessment, what the student needs and how to respond."*

Looking to the Future

The work in Katy ISD is continuing. The assessment system review made visible some of the issues that teachers and leaders already implicitly and explicitly knew about their system: there was a great deal of testing happening. It also brought to the foreground the way in which district policies were distorting assessment practices. Moving forward, the leadership team is focusing on strengthening teachers' formative assessment practices, recognizing that change is incremental rather than happening in just a month or two.

Conclusion

The example of Katy ISD illustrates the way in which the threats to a balanced assessment system are inter-related and how working on one may result in a better understanding of additional threats. It also points to the importance of strong leaders who are willing to ask hard questions, even when they are not sure where the answers will take them.

Summary of District Vignettes

The district vignettes in this Guidebook illustrate examples of how districts have approached efficiency, usefulness, and coherence in their assessment systems. District examples varied in their starting points and timeframes, but a consistent theme across these examples is the role that stable leadership plays in supporting change. The importance of leadership will be discussed by Kristen Lewald and Cindy Goldsworthy in the Pennsylvania vignette, but worth reflecting here across these district examples. Many of the district leaders that we interviewed had considerable experience in their districts, which is not to suggest that it is necessary, but to recognize that stability of leadership provides some affordances when it comes to maintaining balance in an assessment system over time.

For some districts, work on their Portrait of a Graduate or Learner influenced how they conceptualize student learning, which then impacted how they think about instruction and assessment. But as we saw in the Katy ISD example, questioning whether the most appropriate combination of assessments is being used to serve the needs of both teachers and leaders does not necessitate the adoption of a Portrait approach. The work in Chicago Public Schools began with their instructional vision which drove curriculum redesign with assessment as an integral part of it.

For some districts, this focus on their assessment system is a recent initiative; for others it has been a longer journey. The work in Katy ISD started in the fall of 2024 with an audit of their assessment system, while leaders in Cañon City Schools, Chicago Public Schools, and Howard-Suamico School District discussed work that has been taking place over a decade. There is no right or wrong timeframe for the work, and enacting one change can frequently lead to identifying other parts of the system that also need to be modified. Leadership in a district that we worked with on an assessment system review explicitly noted that they were evaluating the impact of decisions made as a result of the review to determine if "it's going to make an impact instructionally for students and teachers or … if it would guide us into another trajectory of other things we need to do."[30]

[30] Evans, C. M. (2025). Eliminating unnecessary district-required Assessments. https://www.nciea.org/blog/eliminating-unnecessary-tests/

State Vignettes

Connecticut State Department of Education

- Ajit Gopalakrishnan, Chief Performance Officer
- Dr. Briana Hennessy, Research Analyst
- Abe Krisst, Bureau Chief
- Dr. Renee Savoie, State NAEP Coordinator

Introduction

The Connecticut State Department of Education (CSDE) has been actively working to promote more thoughtful and efficient assessment practices across the state. Motivated by both legislative mandates and concerns about over-testing locally, CSDE has undertaken a range of initiatives—from issuing statewide guidance through its *Sensible Assessment Practices* document to launching a legislatively required audit of district-required assessments. Through these efforts, the agency aims to help districts eliminate redundancy, reduce unnecessary testing, and improve the instructional value of assessment data.

At the same time, CSDE leaders are grappling with the realities of local autonomy and longstanding systemic pressures that shape district behavior. Despite state-level efforts to promote balance and purpose in assessment systems, local practices are often driven by political demands, evaluation policies, and vendor marketing. CSDE's experience underscores how a state can serve as a convener, guide, and thought partner, but also reveals the difficulty of aligning policy intentions with what happens in classrooms and boardrooms across hundreds of districts.

Providing Guidance on Sensible Assessment Practices

Connecticut has released multiple tools and resources to help local leaders and educators think more intentionally about assessment time, redundancy, and purpose. In particular, CSDE's *Sensible Assessment Practices* report outlines strategies for reducing over-testing and eliminating unnecessary assessments. The guidance urges districts to make better use of existing data, such as prior-year summative results, before administering early-year diagnostic or interim assessments. Gopalakrishnan explained the thinking behind this recommendation:

Starting Point
Threat 1: Too much testing overall
Threat 2: Redundant assessments
Threat 3: Unused assessment results

"A student actually has a bunch of historical information … Are you leveraging that? You can eliminate [fall testing] altogether or at least think more strategically about where and when you assess."

CSDE also maintains a robust student assessment webpage that communicates the state's evolving vision for a more sensible and purposeful system of assessment. Yet, state leaders acknowledge that guidance alone is insufficient to change practice. As Savoie noted, *"We can't penetrate that district firewall … What teachers are hearing is not necessarily what we're saying."*

This gap in communication between state intent and classroom reality underscores the need for ongoing professional learning and stronger connections between state policy, guidance, and district-level implementation.

Implementing a Legislatively Mandated District Assessment Audit

In 2024, the Connecticut General Assembly passed Public Act 24-93, requiring CSDE to conduct a comprehensive audit of district-required assessments. One unique aspect of this law was that CSDE helped draft its audit language, helping lawmakers pass something that both sides felt would help them get to the core of their questions. The stated goals were to "eliminate redundant assessments, discourage test preparation, reduce testing time, and maximize assessments for classroom teachers." CSDE launched the audit in 2025, collecting submissions from nearly 200 districts. The department has until January 31, 2026 to submit a report to the General Assembly on the audit.

CSDE emphasized that the audit is intended to provoke local reflection and conversation, not to penalize districts. But state staff also recognized the political complexity of publishing audit results, especially given concerns about how the findings may be received by families, school boards, and the media.

CSDE is currently considering how to report audit findings in ways that balance transparency with constructive district engagement. Options include summary reports, scatter plots, and side-by-side comparisons of assessment purposes and time—all designed to spark local review and refinement.

How Policy Incentives Distort Local Assessment Practice

CSDE staff reflected on how state and local policy incentives—particularly around teacher evaluation and school accountability—have distorted local assessment practices, often in ways that undermine coherence. For example, teachers were using interim assessment results to set baseline targets and monitor their "teacher effectiveness" as part of their student learning objectives. Even after Connecticut moved away from requiring test-based teacher evaluations, many

Threat 7: Policies and politics that distort practice

districts continued using commercial interim assessments to meet perceived evaluation or reporting needs.

Districts also face pressure to tell a continuous K–12 story, despite gaps in state-required testing for non-tested grades and subjects. Without a through line, some districts turn to commercial assessments as proxies, administered frequently across grades to track growth or prepare reports. Gopalakrishnan offered an alternative strategy: *"Instead of testing three times a year, maybe test at the end of grade K-2 … You can still reduce a lot of testing while preserving your through line."*

Yet, these pressures are reinforced by external forces. School boards want data, families demand accountability, and vendors market products as comprehensive solutions. As Hennessy put it: *"Districts want to show growth … They're being sold a package, and they need something to bring back to the board that says, '60% of our kids improved.'"*

Threat 9: Under-emphasizing the role of formative assessment

CSDE leaders are candid about the challenges of shifting entrenched practices, especially when interim assessments become "mini-summatives" that crowd out formative assessment practices and produce results with limited instructional value.

Conclusion

Connecticut's story illustrates how state education agencies can use tools like guidance documents and audit requirements to surface and respond to persistent threats to balanced assessment systems. But the CSDE experience also highlights the limits of state influence. Meaningful change will require not only policies and reports, but deep local reflection, clearer communication with stakeholders, and long-term investments in assessment literacy. As Connecticut leaders emphasized, the state cannot do it alone, and that makes relationships with districts—not just regulations—the most critical levers for change.

Oregon Department of Education

- Dr. Dan Farley, Assistant Superintendent of Research, Assessment, Data, Accountability & Reporting (RADAR)
- Dr. Andrea Lockard, Director of Assessment & Student Reporting

Introduction

The Oregon Department of Education (ODE) offers an example of how a state education agency can support balance in local assessment systems while navigating complex political pressures, local variability, and limitations on state authority. In recent years, ODE has provided guidance and tools/resources, developed student and staff surveys, removed the state's kindergarten readiness assessment, and implemented a legislatively mandated local assessment system audit. Their experience highlights both the promise and limitations of state influence in addressing threats to efficiency, usefulness, and coherence in local assessment systems.

Providing Guidance and Tools/Resources

The Oregon Department of Education (ODE) has a long history of providing guidance and resources to support school districts in transitioning to more balanced assessment systems. The state's theory of action states: "If Oregon educators increase appropriate uses and decrease inappropriate uses of tests and assessment data by increasing assessment literacy, then Oregon educators will make better instructional decisions that increase student learning."

We begin telling the story of Oregon's journey in the 2014-15 school year, when the state legislature funded interim assessments, but did not allow ODE to procure one for districts to use. To guide district decision-making, ODE provided districts with guidance about what to look for and how to evaluate the quality of commercial interim assessments prior to purchase. In the 2016-17 school year, ODE funded the Oregon Formative Assessment for Students and Teachers (OFAST) project, which provided a comprehensive course containing six training modules on formative assessment practices. That same year, ODE also worked with districts on performance assessment development.

Starting Point
Threat 4: No clear match between the assessment purpose, design, and use

ODE staff recognized that they were promoting and funding projects and initiatives that built the capacity of local educators to know how to select, design, interpret, and use different types of assessment (formative, interim, summative) in ways that support teaching, learning, and program evaluation. However, there was no overarching vision or theory of action describing how all of

those assessments should be used together. To address this issue, they developed a consolidated guidance document published in 2019: *The Right Assessment for the Right Purpose*.

The purpose of this more comprehensive guidance was to support "ODE's initial effort to develop assessment literacy, providing districts with clear communication about assessment and appropriate responses to assessment data as we build a more balanced approach to assessment in Oregon" (p. iii). The guidance provides background on balanced assessment systems, the importance of assessment literacy, and the components of a balanced assessment system. It explains how such an approach should address common problems of practice and support best practices, such as the appropriate use of time spent testing, appropriate uses of assessment results, and how to mitigate the misuse of assessment results.

ODE has continued over the years to provide additional resources, tools, and guidance that school and district leaders can use to "support local assessment-related professional development efforts." These professional development materials are curated on their state website. They are organized to support the implementation of balanced assessment systems by focusing on a range of topics including formative assessment practices, equitable grading practices, implementing and interpreting interim assessments, and implementing and interpreting summative assessments. At the same time, ODE is quick to note that simply publishing guidance or providing open-access professional development materials to build educator assessment literacy doesn't necessarily lead to better practice. Although ODE published *The Right Assessment for the Right Purpose* to help clarify how different assessment types should be used, uptake has been limited. Many educators still rely too heavily on assessments that produce a number—often at the expense of teacher-led formative assessment. As Lockard explained, *"There's a prioritization put on an assessment that spits out a number … that will get prioritized over teacher formative assessment practices."* ODE has intentionally incorporated student and staff surveys that provide qualitative information for its districts, providing necessary context needed to interpret the quantitative data that are available.

Threat 8: Over-emphasizing the role of summative assessment
Threat 9: Under-emphasizing the role of formative assessment

Dropping the State's Kindergarten Readiness Assessment

One action that is completely under state control is the design of state tests, as long as they meet federal peer review guidelines. Oregon used its authority to remove and replace the state's kindergarten readiness assessment to better support its vision of balanced assessment systems.

Threat 1: Too much testing, particularly early in the year
Threat 6: Inconsistency between assessments and instructional vision

ODE noted a longstanding concern about administering high-stakes assessments too early, particularly in kindergarten, when student-teacher relationships have not yet formed. This practice undermines the validity of the results and can be demoralizing for students. According to Dan Farley, *"You can't test students before they have a rapport developed with their educators ... Kindergarten is where we have the most concern ... their first experience of school is taking a test with someone they don't know."*

ODE eliminated the state's kindergarten readiness assessment and replaced it with the *Early Learning Transition Check-In*, a qualitative tool focused on family-teacher dialogue rather than student testing. ODE took this action to directly address over-testing (especially early in the year) and concerns about an inconsistency between assessments and instructional vision.

Implementing a Legislatively Mandated District Assessment Inventory

Threat 2: Redundant assessments
Threat 5: Assuming all tests can inform instruction

A statewide audit mandated by HB 4124 in the 2022 legislative session aimed to identify redundant district-mandated assessments. All districts were required to complete a survey about the kinds of academic tests they require students to take (and why) in grades pre-K through 12, excluding statewide summative tests. A full report was presented to the state legislature in May 2024. Key findings from the analysis of district-required assessments showed:

- Most required tests were reported to have multiple uses and purposes;
- Classroom teachers are reported as the primary users of tests results; and
- Students spend an estimated average of 14.8 hours per year on district-required tests in elementary school, 17.3 hours in middle school, and 18.4 hours in high school (page 3).

While the legislatively mandated district assessment audit or inventory provided valuable insight, ODE lacked authority to validate responses or require follow-up. Despite the audit revealing redundancy, there was little evidence that districts used the findings to change practice. *"The audit just led to more questions, more concerns, and more desire to do a better job of it the next time,"* Farley noted.

The audit also revealed a lack of assessment literacy, especially concerning the false belief that all tests can directly inform instruction. For example, classroom teachers were identified as the primary users of all assessments listed about 90% of the time. ODE expressed concern that districts often try to use whatever assessment data is available to guide instruction, regardless of whether the tool is designed for that purpose.

Additionally, ODE emphasized that the issue revealed by the audit isn't merely how much time testing takes, but whether the instructional time lost yields meaningful data that is actually used. Meaningful use of test data is dependent upon there being a match between the assessment purpose, design, and intended user. ODE is considering follow-up resources such as protocols or reflection questions to encourage local action in response to their audit information.

Previewing the Horizon

ODE sees off-the-shelf assessments as the single greatest threat to coherence in Oregon's assessment system. In particular, ODE flagged the influence of commercial vendors as a major source of noise in the system. Off-the-shelf assessments are often misaligned with state standards and lack transparency about their technical quality or accessibility features. According to Farley, *"Vendors are selling a product … they'll tell you it can do anything … that really is creating a lot of noise."*

Threat 7: Policies and politics that distort practice

Lockard added that many districts sometimes lack the capacity and/or necessary supports to critically evaluate vendor claims. She advocates using the District Assessment Procurement Protocol (DAPP) to help districts push back. ODE also noted their concerns about the implementation of progress-monitoring tools and universal screeners, two topics that they plan to address in the near future through guidance and professional learning supports. However, even with thoughtful guidance, Oregon and other state leaders face clear limitations in shifting local practice.

Conclusion

Oregon's story highlights how state education agencies can use district assessment audits and guidance to identify and address persistent threats to balanced assessment systems, as well as the importance of adding qualitative information to guide decision-makers. Yet, as ODE's leaders candidly admit, the power of the state is often constrained. Achieving efficiency, usefulness, and coherence at scale will depend not only on state policies and guidance but on sustained local partnerships, deeper communication, and a shared understanding of what high-quality assessment practice looks like.

Pennsylvania Department of Education

- Dr. Kristen Lewald, Statewide Projects Team for PDE, Lancaster-Lebanon IU13
- Dr. Cindy Goldsworthy, PVAAS Statewide Team for PDE, Consultant

Introduction

In Pennsylvania, years of work supporting school and district leaders have revealed a recurring pattern: assessment systems are often bloated, misaligned, and incoherent—not because of bad intent, but because of limited foundational knowledge, fractured leadership, and a lack of shared purpose. To address these challenges, Kristen Lewald (IU13) and Cindy Goldsworthy co-developed statewide professional learning resources to support the data and assessment literacy of Pennsylvania educators under the leadership of the Pennsylvania Department of Education (PDE).

Our interview with them focused on the guidance and practical resources (i.e., assessment maps) they developed to support the design, implementation, and evaluation of balanced assessment systems. Their work demonstrates how states can support balanced assessment systems through partnerships with intermediary organizations, focusing on leadership development, practical tools, and attention to system-level structures.

Mapping to Improve Efficiency, Coherence and Use

Lewald facilitates continuing education data courses for PDE's statewide program for school leaders, Pennsylvania Inspired Leadership (PIL) program. She and Goldsworthy co-developed one of the learning modules in the PDE PIL *School Level: Data to Action* course with the goal of supporting high-quality, local assessment systems.

Lewald and Goldsworthy emphasized that building a balanced assessment system begins with clarity of purpose and consistency in practice. Goldsworthy noted, *"One of the biggest threats is the lack of knowledge about assessment design and intent. Not all assessments are meant to do the same thing."* This misunderstanding leads to redundancy, wasted instructional time, and data that goes unused. To combat this, they introduced school and district leaders to the concept of an *assessment map*—a document that clearly outlines what assessments are administered, at what grade levels, for what purposes, and when. *"It's not an assessment calendar. It's much more intentional. Mapping helps identify gaps, redundancies, and mismatches between purpose and practice,"* said Lewald.

Starting Point
Threat 4: No clear match between the assessment purpose, design, and use

The origins of this work trace back to Goldsworthy's time as a district curriculum director. When she asked teachers to list the assessments they were using, *"We found over 100 different assessments being given across classrooms—many of them duplicative or low value."* To organize this chaos, she engaged teachers and school psychologists to create a *feature matrix* comparing assessments based on time, cost, purpose, and data usefulness. This process helped educators identify which assessments to keep and which to eliminate.

Threat 1: Too much testing overall
Threat 2: Redundant assessments
Threat 3: Unused assessment results

Supporting Implementation Using Structures and Routines

Assessment maps alone, however, are not enough. Both Lewald and Goldworthy stressed the importance of structures and routines that bring assessment data into instructional decision-making. *"If there's no data meeting structure, no routines, then even the best-designed map won't matter,"* said Lewald. In districts where Goldsworthy led change, they created tight alignment between assessment maps, data protocols, and professional learning communities (PLCs). Teachers were expected to come to data meetings prepared to analyze only the assessments listed on the map.

Over time, this structure helped shift practice. *"We had teachers who rolled in carts full of binders with their own assessments, but no one looked at those. They eventually stopped bringing them. The map changed the conversation,"* Goldsworthy recalled. The map focused the conversation on common, curriculum-embedded assessments and other assessments shared across classrooms and teachers, rather than idiosyncratic, teacher-created assessments used only in one classroom.

Addressing Other Common Threats to Balance

Goldsworthy and Lewald also underscored that leadership matters. Effective change required courageous district administrators who understood assessments well enough to lead hard conversations, unify general and special education practices, and push back against vendor marketing. *"We still see districts administering redundant early literacy assessments because no one is willing to drop something they already paid for,"* said Lewald. *"Or they adopt assessments that take 45 minutes when a 5-minute screener would suffice."*

The pair also highlighted a phenomenon they see often: "assessment whiplash." This occurs when new leaders adopt entirely new assessment tools, discarding what existed—regardless of quality or buy-in. *"In one district, leaders and teachers received training for three years on one tool, and then a new administrator came in and said, 'I don't know this. Let's use what I used in*

my last district.' Everything built was lost," Lewald said. This turnover not only disrupts coherence but also frustrates educators who must then live in a constant state of change.

Concluding Reflections

Lewald and Goldsworthy are the first to admit that one module on assessment maps isn't nearly enough. Without ongoing support, many district and school leaders continue to struggle with misaligned tools, weak protocols, and incoherent systems. Still, their experience shows that change is possible when assessment tools, practices, and leadership structures are aligned. In their words: *"A map provides direction, but it's the system around it—leadership, routines, accountability for implementation, and follow-through—that makes it useful,"* Lewald said. Through a mix of state-supported professional learning, practical tools like assessment maps, and intentional leadership development, Pennsylvania's model offers a compelling example of how to build assessment systems that are coherent, useful, and manageable—one LEA/district at a time.

At the same time, the limitations of this approach should be apparent. Effectuating change one district at a time reflects the problem of scale affecting the move towards more balanced assessment systems. There are approximately 12,000 to 13,000 school districts in the United States. While working deeply with a handful of districts is a start, it is not a complete solution.

Wisconsin Department of Public Instruction

- Lauren Zellmer, Standards-Based Assessment Consultant, DPI Office of Educational Accountability
- Kim Pencil, Assessment and Data Specialist, CESA 7
- Nick Joseph, Director of Secondary Curriculum & Instruction, Unified School District of De Pere

Introduction

The Wisconsin Department of Public Instruction (DPI) launched a statewide initiative aimed at improving assessment and data literacy by developing strategic assessment systems that enhance classroom practice. The initiative began in 2015, with the state providing guidance, tools, and resources to strengthen and improve the assessment and data literacy of its educators. Professional learning modules were created in 2017. However, we will focus on telling Wisconsin DPI's story starting in the 2023-2024 school year. This is when DPI partnered with the twelve Cooperative Educational Service Agencies (CESAs)—regional educational organizations—to deepen district-level engagement in designing and using meaningful classroom assessments. Our vignette focuses on the work of CESA 7, in Green Bay, which provided individualized coaching sessions to district leaders on a range of topics related to assessment literacy. The initiative demonstrates how states can promote more balanced assessment systems by empowering regional educational service agencies and districts to enhance their focus on instructional coherence and relevance.

Evaluating System Quality Using Assessment System Inventories

Starting Point
Threat 1: Too much testing overall
Threat 2: Redundant assessments
Threat 3: Unused assessment results

In the first year of the initiative, the Wisconsin DPI worked with CESAs to lead districts through an audit of their local assessment system in at least one or two of the coaching sessions. Both the state education agency and regional intermediary organizations felt it was important to use the audit to support LEAs in implementing a strategic assessment system in which a variety of aligned, high-quality assessments provide evidence of student learning with the least amount of testing. Districts were led through a process of reviewing the quality of their assessment system, using an inventory tool and set of evaluation questions.

Lauren Zellmer noted, *"I think the assessment audit was an important part of the assessment professional development series because it can provide a big picture overview of assessment within each school or grade level, identify redundancies or gaps within the system, and uncover*

how the results are or are not being used." However, she went on to explain that they decided not to continue with the assessment system reviews as a required component in year 2 of the initiative because of the amount of time required to lead districts through the process in a meaningful way and the specific coaching needs of their participants didn't always match. *"We found that to do this work well, a great amount of coaching time with districts was needed, so we decided to make the assessment audit an optional coaching tool for year 2,"* she said, which can serve as an important cautionary tale for others.

Building Strategic Assessment Systems

According to Zellmer, DPI intentionally *"emphasized that a balanced assessment system inside a classroom has several different ways of assessing students."* This framing is critical. The agency collaborated with CESAs to co-design the project structure and training topics—such as formative assessment, performance tasks, project-based assessments, and writing assessments—centered on improving local assessment practices and data use.

Threat 4: No clear match between the assessment purpose, design, and use

Zellmer explained, *"We really wanted to emphasize what we call Strategic Assessment Systems—not just naming all the parts of the assessment system, but making clear which parts matter most for learning."* DPI made these professional learning opportunities available for schools and districts at no cost, reflecting their commitment to improving student learning and equity.

Providing Tailored Coaching and Professional Learning

Kim Pencil, an Assessment and Data Specialist from CESA 7, described her role as part facilitator, part coach, and part translator between policy and classroom practice. Her work began by recruiting districts and then delivering a series of four-to-five-day trainings focused on different types of classroom assessments. Following the training, Pencil provided coaching that was fully customizable: *"A lot of my coaching has been full staff professional development, where participants said, 'We need this chunk of that four-day training brought back to the whole staff.'"*

Coaching included everything from revising rubrics with grade-level teams to developing common assessments and crafting assessment re-take policies. Sessions were flexible—in-person or virtual, during the school day or after school. *"We learned the one-size-fits-all approach for coaching doesn't work. We needed to meet districts where they were to provide them with the support they need,"* Zellmer noted.

Leaning Into Balance

In the Unified School District of De Pere, near Green Bay, Nick Joseph, the Director of Secondary Curriculum and Instruction, described how his already high-performing district benefited from this initiative, *"This reassured us on some of the things we've done and helped identify areas to improve."* With coaching support from CESA 7, the district focused on improving end-of-unit summative assessments through blueprinting and alignment checks and began more intentionally using formative assessment information to adjust instruction.

Joseph described a core principle guiding their work: *"We're not teaching to the test; we're testing to teach."* The district embedded a daily 30-minute "flex" period for small-group reteaching based on assessment results. Teachers are expected to align their unit pacing so students with similar learning gaps can be grouped and supported.

Still, challenges remain. Both Pencil and Joseph pointed to a persistent issue: assessment results being collected but not acted on. *"Teachers often just enter a grade and move on,"* Pencil said. Joseph added, *"If we're doing formative assessments and kids just see a grade, they won't value them unless we follow through with targeted support."* Pencil also raised a less commonly cited but growing concern: student disengagement from assessments. *"Too often, assessment is something that happens to students, not with them,"* she explained. DPI and CESAs have begun surfacing this issue in coaching sessions and reframing assessment as an instructional tool that actively involves students.

Threat 8: Over-emphasizing the role of summative assessment
Threat 9: Under-emphasizing the role of formative assessment

Looking Ahead

The Wisconsin DPI initiative offers a replicable model of how a state agency can broaden its theory of action to prioritize instructional improvement. Zellmer, Pencil, and Joseph all emphasized the importance of flexibility, local responsiveness, and relationship-building in making this work possible.

As Joseph reflected, *"The fact that this came from the state but was delivered through local CESAs meant that now when we hire someone from a neighboring district, they already speak the same language of balance."* This systemic coherence—across classrooms, districts, regions, and the state—is exactly the kind of alignment needed to bring balanced assessment systems from theory into practice. And yet, despite the early successes of this work, the amount of coaching time and funding commitments may not be replicable across state contexts. States face a tradeoff as they try to get more directly involved in local improvement efforts, especially efforts aimed at influencing what happens in the daily interactions between students, teachers, and the content.

Summary of State Vignettes

The state vignettes in this Guidebook illustrate examples of how states have supported the efficiency, usefulness, and coherence of local assessment systems. The state education agency leaders we interviewed focused on the tools, guidance, and/or supports they provided to schools and districts with the hope of promoting more balanced assessment systems at the local level.

In particular, state leaders described the support they offered around local assessment practices, the tools they provided around local assessment system reviews or audits, and the guidance they produced to build the assessment literacy capacity of educators (teachers and leaders) in their states. The focus on these types of resources was likely a by-product of the questions we asked state leaders in our interviews, as well as what is most relevant to the state role in supporting more balanced assessment systems beyond the design of and messaging around the purpose and use of state assessments.[31] Below, we synthesize successes and challenges associated with various approaches, as discussed in our interviews.

Local Assessment Practices Support and Assessment Literacy Resources

Local assessment practices support includes any resource offered by a state education agency to districts and schools related to high-quality local assessment practices. We showcased the efforts of Connecticut, Oregon, Pennsylvania, and Wisconsin to support improved local assessment practices by providing guidance, tools, and resources. Other states also provide guidance to districts to support the implementation of balanced assessment systems, although they may replace the term "balanced" with terms such as "sensible," "comprehensive," and "strategic," among others. For example:

- Michigan's Balanced Assessment System
- Rhode Island's Comprehensive Assessment System

Assessment literacy topics include designing/selecting, interpreting, and using classroom formative and summative assessment processes; evaluating, interpreting, and using commercial interim and locally created common assessments; and interpreting and using state test results. The distinction between local assessment practices support and assessment literacy resources appears to be a distinction without a difference in practice. In other words, during our interviews, state education agency leaders discussed the professional learning resources, guidance documents, tools, and other materials they provided as sets of assessment literacy materials intended to support and drive better assessment practices and uses in schools and districts.

[31] In Chapter 7 of the 2024 NAEd volume, *Reimagining Balanced Assessment Systems*, Evans and Landl discuss six state actions that support more balanced assessment systems at the local level. Providing these types of resources was included in State Action 6. We refer interested readers to that chapter for more details on a state's role in supporting balanced assessment systems.

Each state we interviewed agreed that providing these types of assessment literacy resources and supports is important and necessary, but does little to change local practice without direct support. While state education agencies often do not have the capacity to do this type of direct support to districts, intermediary organizations like CESAs and IUs serve to bridge that gap and offer critical implementation support. Yet a key tension remains about scale and spread. While intermediary organizations can support a handful of districts, they also lack the capacity and funding to support all districts or schools in a state. This is a persistent problem of practice related to scaling and spreading assessment literacy efforts or initiatives.

Local Assessment System Auditing Tools

Local assessment system auditing tools often entail a collection, analysis, and evaluation of all of the state, district, school, and classroom summative assessments administered to a student over the course of a year.[32] Overall, local assessment system audits should allow district and school leaders and teachers to ask and answer key questions such as:

- How can the collection of assessments administered to students over the year (both within the classroom and outside the classroom) best support our vision of high-quality instruction?
- How can we design or implement a system of assessments to ensure that different decision-makers (e.g., district/school leaders and teachers) have the information they need to make better educational decisions that ultimately support student learning? What assessments need to stay, go, be redesigned, or added?

In our interviews, we identified several ways in which state education agencies and state-level organizations are directly supporting local assessment system reviews or audits. These approaches include state-legislatively mandated assessment audits (e.g., Connecticut and Oregon) and state-supported assessment audits conducted through intermediary organizations (e.g., Pennsylvania and Wisconsin). These are not the only ways states are supporting local assessment system reviews. For example, some state education agencies provide tools and resources for conducting assessment system audits on their websites (e.g., Michigan Department of Education) or through state-funded organizations (e.g., California Collaborative for Educational Excellence).

While local assessment system audits are a valuable component of moving towards more balanced assessment systems, simply inventorying the system is not enough. State leaders discussed the constraints they experienced because of local control and ways in which audits must be paired with structures and routines in schools, leadership support, and continued growth in assessment literacy to support implementation.

[32] See Table 3-1 for a listing of free, open-access assessment system auditing and inventory tools.

SECTION 5: MANAGING THE CHANGE PROCESS

Designing a balanced assessment system is not like planning a balanced diet—mixing the occasional summative, with a little interim, and a lot of formative. Balance is not about having a specific type or number of assessments: it is about identifying and collecting the assessment information needed to complement curriculum and instruction in ways that promote rich classroom learning environments. That is why balance exists along a continuum rather than as an on/off switch. The goal of working toward balance is to provide teachers with the information they need to adjust their instruction and leaders with the information they need to monitor school and program quality and close achievement gaps.

In Section 3 of this Guidebook we identified some of the major threats to balanced assessment systems, provided look-fors to help with the determination of how prevalent those threats might be in your district or school, and suggested some resources to support that work. Although this document is not intended as a "how to guide," it would be remiss to not say something about the process of both threat identification and remediation from the perspective of organizational change management.[33]

For school and district leaders there are two aspects to consider: (1) the change process itself such as how potential changes to the assessment system will be identified, explored, prioritized and selected for action, and (2) the conditions that need to be in place for the desired changes to be successfully implemented and sustained. We will briefly address both topics in this final section.

The Change Process

Volumes have been written on the topic of managing organizational change, but in this conclusion we are going to focus on strategies supporting effective *implementation* (e.g., developing and implementing a new instructional vision for the district) and *de-implementation* (e.g., stopping, reducing or doing something quite differently such as removing unused assessments, or shifting how professional learning time is spent). Both kinds of change can be difficult, and understanding some of the human tendencies involved can help leaders avoid pitfalls and sustain and scale change over time.[34] Two of the most prominent tendencies are provided below.

---

[33] For an extended discussion of change management, see Peurach, D.J., & Russel, J.L. (2024). Developing, implementing, and institutionalizing complex educational innovations: Considerations for balanced assessment systems. In Marion, S.F., Pellegrino, J.W., & Berman, A.I. (Eds.). *Reimagining balanced assessment systems*. National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems/chapter8

[34] Hamilton, A., Hattie, J., & Wiliam, D. (2023). *Making room for impact: A de-implementation guide for educators*. Corwin Press.

1. Human nature is more inclined to add than remove. Research has shown that when people try to solve all sorts of different problems across disciplines, they are more likely to add than remove something.[35]
2. We are more inclined to stick to what we are currently doing than to change. Change can feel daunting so we keep doing what we are doing even if we know it isn't working because we have invested so much time and effort already in trying to make it work. This is sometimes referred to as the sunk-cost fallacy.

Table 5-1 suggests a four-stage process for identifying and applying implementation or de-implementation changes to address the threats identified in Section 3. These four stages should not be considered a linear implementation plan; but an iterative process similar to a Plan-Do-Study-Reflect cycle.

TABLE 5-1 Four Stages for Implementation and De-Implementation in the Context of Assessment System Threat Analysis and Improvement

Stages	Specific Application to Assessment Systems
Discover	The **purpose of the discovery phase** is to help districts build a shared understanding of how well their current assessment system functions and where targeted improvements can enhance coherence, efficiency, and usefulness. This phase requires the engagement of senior leadership, a cross-functional team with adequate time to review materials and share ideas in a safe space, and a commitment to follow through to be successful. • Part of setting the stage for the work is having a senior leader at the district provide the mandate and leadership for exploring the efficiency, usefulness, and coherence of the local assessment system, and who can communicate why these changes are worthwhile exploring. • This work should be a collaborative process with a core cross-functional team who can draw on additional perspectives and insights as the work evolves. The combination of insights from district- and school-leaders with classroom teachers is critical. • The initial phase of identifying candidate threats to explore might be done with this core team reviewing the threat descriptions in Section 3 to see which ones resonate with the local context. • Additional time should then be spent on a more thorough exploration of one or two candidate threats which could be done using the look-fors associated with each threat. For example, this exploration work could

[35] Klotz, L. (2021). *Subtract: The untapped science of less*. Flatiron Books.

	entail: ○ Developing or refining an instructional vision in order to use it as a lens through which to examine coherence in the system. ○ Conducting an assessment system audit at one or more grade levels. ○ Engaging in a series of listening tours to better understand intended and perceived messaging about assessment to identify potentially problematic policies.
Decide	The **purpose of the decide phase** is to help districts move from an investigation of threats to a specific action plan to be implemented. • At some point one or more recommendations for change will be made by the team and then the focus will shift from problem-discovery to solution-planning. Plans will vary in terms of whether solutions are focused on implementation (e.g., new professional learning to deepen knowledge and practice of formative assessment) or de-implemention (e.g., redesigning some unit assessments to include performance tasks), or a combination of both. ○ If the plan involves implementing some new things, the team may also want to consider whether they also need to take some things off teachers' work load to make time or space for the new activities (i.e., de-implement some other duties). ■ If teachers or leaders need opportunities to improve their assessment literacy, consider what structures will support meaningful learning that can be sustained over time. ■ Other aspects of assessment change, such as teachers reviewing student work together, will also require attention to the structures in school days to make this possible. ○ It is perhaps easier to decide to implement than de-implement so it is important to consider four ways to de-implement:[36] ■ Remove: stop doing it entirely (e.g., remove unused assessments from the system). ■ Reduce: doing it less often or by/to fewer people (e.g., changing how frequently people meet to review certain types of assessment results). ■ Re-engineer: do it more efficiently (e.g., reorganize schedules so that grade-level teams have time to develop or revise curriculum-aligned common assessments together).

[36] Hamilton, A., Hattie, J., & Wiliam, D. (2023). *Making room for impact: A de-implementation guide for educators*. Corwin Press.

	■ Replace: substitute it with a more efficient or effective alternative (e.g., change whole-staff professional learning to grade-level team time that focuses on aligning curriculum-embedded assessments with the instructional vision). • For any identified change, a plan needs to be developed to ensure potential issues are identified and mitigated in advance: is the change to be piloted with a small number of teachers or schools, or is it a change for an entire content area, grade level, etc.? What will teachers and students do who are not involved in the pilot? What if educators are not willing to make the desired change (e.g., discontinue a particular assessment practice). • What will success look like? What information will need to be collected and when to inform the evaluation? Over how long of a time period?
Implement or De-Implement	The **purpose of the implement or de-implement phase** is to put specific plans into action, along with an ongoing evaluation of how those plans are working. • During this stage the plan for change and evaluation are put into action. • Communication is important as decisions roll out to a broader audience. • Check in on how the plan is going, and making minor adjustments or providing additional communications can help make this a smooth process. • What structures are needed to ensure effective communication across the district about insights the core planning team has? What structures will enable this team to systematically get feedback from other teachers and leaders in the district?
Re-Decide	The **purpose of the re-decide phase** is to begin the process over again with a new area of focus. • When the evaluation information has been collected the core planning team should meet again to review it and make an informed decision whether to discontinue the pilot, modify the pilot, or move to full implementation. • What needs to be done differently if this change is to become permanent? • What is next? What is the next threat to be considered? Return to the Decide stage.

While the four stages seem quite linear as laid out in Table 5-1, in reality the process is iterative. For example, deciding what to change may uncover some areas that were not fully understood which could result in a return to a Discover stage.

The core team might have some hypotheses in the early stages about both threats and potential solutions and may fast-track one area of exploration looking for a quick turnaround and early win, recognizing that another threat may take longer to understand and problem-solve. Alternatively, the team may decide that they need to take a slow-and-steady approach to work on a single large threat. Communication both within the implementation team and with the larger system will be vital to build trust and support progress.

Sustaining Change

While the change process laid out in Table 5-1 contains some suggestions that are related to sustaining change (e.g., engaging a team to guide the process, communication) the primary focus is change initiation. Schools experience many initiatives, sometimes competing with each other during the same school year, and so it is worth considering a framework for managing and sustaining complex change. One common framework comes from Lippitt (1987), later modified by Knoster (1991).[37] This model identifies five components for achieving sustainable change: vision, skills, incentives, resources, and action plan. Table 5-2 describes each component and what can happen if that component is not in place.

TABLE 5-2 Five Components for Sustaining Change and Impact if Missing

Component	What it is	What happens without it
Vision	A clear reason or motivation for the change	Without a guiding vision, educators will be **confused** about the purpose of change
Skills	New knowledge and skills may be needed to work in new ways	Without opportunities to develop the requisite knowledge and skills, educators will be **anxious** about expectations being placed on them
Incentives	The benefits (short-term and longer-term) that are expected as a result of the change	Without incentives to change, educators will be **resistant**
Resources	Time, materials or personnel	Without resources to support change, educators will be **frustrated**
Action Plan	Plans for how to integrate changes in existing ways of working	Without an action plan to guide next and subsequent steps, educators will have **false starts**

[37] Lippitt, M. (1987). *The managing complex change model*. Enterprise Management Ltd.; Knoster, T. (1991). *Factors in managing complex change*. Material presentation at TASH conference, Washington D.C.

This framework can be applied to proposed plans about changes to the assessment system within a school or district:

- Has the leader or team provided a **vision** for what needs to change and what the ultimate benefit will be for students? Has this vision been communicated to teachers, families, and/or other interest-holders?
- Have learning opportunities been created/identified for teachers or other educators to develop any new knowledge and **skills** that they might need to implement the change? What structures are needed to sustain these learning opportunities over time?
- Have the **incentives** or benefits of the change been communicated to teachers and other interest-holders?
- Are the appropriate **resources** in place? If the changes will add to or change teacher workloads, do they have time? Do they have the materials they might need? Do they need additional support in their classrooms while they develop new ways of working? How can coaches or other school leadership support them in this change process?
- Is everyone clear on the **action plan**: not only the immediate change but any snowball effects?

For some, the thought of more change can bring out the cynics who will mutter about "reform du jour" or "just wait out this school year and it'll go away again." If it has been a habit within the district to bounce among initiatives, then leadership will need to acknowledge this and also plan for how this effort will be different. Examining previous initiatives in light of the components in Table 5-2 can help identify previous challenges and start the process of remediating them with this new effort. In addition, leaders may want to consider school culture broadly, or assessment culture specifically, to better understand whether there are additional barriers to implementing change (see Appendix A for more on assessment culture).

Concluding Thoughts

Implementing change within a school or district—whether that involves implementation or de-implementation, learning new practices or un-learning old ones—is complex. Using the discovery–decide–implement–re-decide cycle is one approach to support a longer-term process of continuous improvement. Sustaining those change processes will require a clear vision, new knowledge and skills, known incentives/benefits, resources, and action plans.

One question that we anticipate district or school leaders asking is this: If we address all nine threats identified in this Guide, will that result in a balanced assessment system. Our response is to remind readers that at the beginning we noted that we believe the threats identified are major threats to balance, and that others exist that could be identified locally. However, if a district made progress on all nine threats, or even a subset of the ones that were relevant to their district, we are confident that they will have made significant progress to a *more balanced* assessment system.

We conclude this Guidebook with an encouragement to all who are working towards designing or implementing more balanced assessment systems to start and end every discussion about this topic with the question: What impact will this assessment or assessment practice have on teaching and learning in classrooms, and is that the type of impact we want?

APPENDIX A: THE ROLE OF ASSESSMENT CULTURE

Creating assessment systems that demonstrate coherence, efficiency, utility, and the other characteristics discussed in this document is no small feat. Addressing deficits in assessment literacy, policy dynamics, the impact of commercialization and other factors that drive assessment decisions requires more than professional development. It demands a thoughtful strategy for assessment system change—one that identifies and addresses the cultural factors that sustain the current state.

The notion of an assessment culture originated from Lorrie Shepard's seminal article, "The Role of Assessment in a Learning Culture."[38] We reference assessment culture as a distinct area of focus within this Guidebook because we believe it can have a powerful influence on efforts to modify assessment systems. Specifically, understanding a school's assessment culture can illuminate why certain barriers exist—and how they might be overcome.

Relation to School Culture

A large body of research has examined ***school culture*** and its impact on policies, practices, interpersonal relationships, and educational outcomes.[39] School culture is typically defined as the expectations, beliefs, values, and assumptions that direct the activities of the personnel and students in a school.[40] Though not often explicitly stated, these cultural norms are evident in how classrooms are organized to support learning, how educators and students interact, how the school communicates with families, how success is defined and celebrated, and even in the posters and awards displayed throughout the building.[41]

Viewed through a different lens, these artifacts, interactions and structures can reveal something about a school's assessment culture. We define ***assessment culture*** as the collective expectations, values, and beliefs a school has about the role, value and purpose of assessment.[42]

---

[38] Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14.

[39] See. e.g.: Harrison, C. J., Könings, K. D., Schuwirth, L. W., Wass, V., & Van der Vleuten, C. P. (2017). Changing the culture of assessment: The dominance of the summative assessment paradigm. *BMC Medical Education*, 17(1), 1-14. http://dx.doi.org/10.1186/s12909-017-0912-5; Hinde, E. R. (2005). School culture and change: An examination of the effects of school culture on the process of change. *Essays in Education*, 12(1), 1-12. http://openriver.winona.edu/eie/vol12/iss1/5; Jerald, C. D. (2006). School culture. Center for Comprehensive School Reform and Improvement. https://files.eric.ed.gov/fulltext/ED495013.pdf

[40] Hinde (2005).

[41] Jerald (2006).

[42] Wylie, E.C., & Landl, E. (2024). *Understanding assessment culture.* The National Center for the Improvement of Educational Assessment and ISTE. https://www.nciea.org/wp-content/uploads/2024/08/Understanding_Assessment_Culture.pdf

*We define **assessment culture** as the collective expectations, values, and beliefs a school has about the role, value and purpose of assessment.*

Like school culture, a school's assessment culture is not written down; rather, it is revealed in the way educators select or design assessments, how they discuss the purpose of different types of assessment, the amount of time students spend either preparing for or taking tests, how parents and students talk about assessment results and what they believe the results mean, how results are communicated and reported and the structures in place to support educators' understanding and use of assessment processes and results. Assessment culture is shaped not only by the broader school culture but also by the individual experiences and predispositions of its educators, leaders, students, and community members.

Impact of Assessment Culture

Taking the time to understand a school or district's assessment culture is important because it can impact assessment decisions in ways that inadvertently support—or constrain—practices that advance more balanced systems of assessment and, ultimately, teaching and learning.[43]

Specifically, it influences decisions made about which assessments are administered to students, when, and why. For example:

- A school that *believes trend data must be maintained at all costs* (e.g., giving a specific assessment to all students three times per year no matter what other assessment information is available) may prioritize assessment continuity over efficiency, quality, or alignment.
- A school that *values assessment as a tool for learning* is more likely to invest in professional development focused on formative assessment practices, rather than strategies to boost interim or summative results.
- A school that *uses assessment to control behavior* (e.g., student data walls, teacher evaluation) is more likely to equate assessments with grades or events to prepare for rather than opportunities for reflection and understanding.

Furthermore, research has shown that school reforms, whether related to assessment or other areas, are unlikely to succeed if they conflict with one or more of the underlying assumptions that shape a school's culture.[44] In many cases, these assumptions and their influence remain hidden until they are intentionally surfaced. This suggests that in order for districts and schools

[43] Shepard (2000).

[44] Finnan, C., & Swanson, J. D. (2000). *Accelerating the learning of all students: Cultivating culture change in schools, classrooms, and individuals.* Routledge.

to make meaningful and lasting changes to their assessment systems, they must first identify and address the assumptions and predispositions that may hinder progress.

Relationship With Balanced Assessment Systems

Figure A-1 uses the previously provided examples to highlight aspects of assessment culture that can help or hinder efforts to develop or implement balanced assessment systems. The figure is divided into four cells. The left column provides aspects of assessment culture while the right column provides characteristics of balanced assessment systems. The row on top is considered "desired" while row on the bottom is considered "not desired" in light of their potential impact on each other and, consequently, teaching and learning.

The relationship between assessment culture and balanced assessment systems (purple arrow) is bidirectional because the tools and activities necessary to move a system toward balance should reinforce a supportive assessment culture in the same way improving assessment culture should provide for greater balance. For example, conducting a thoughtful assessment system audit requires educators to carefully interrogate the components of their existing assessment system for efficiency, utility, and coherence with the system's instructional vision. This activity can improve assessment culture by highlighting inaccurate assumptions, building trust, and facilitating ongoing communication about the purpose and value of different types of assessment.

The top left cell includes assessment-related beliefs and practices that are likely to support teaching and learning. Examples include valuing assessment *for* learning, school structures that support collaborative analysis, alignment with instructional goals, and so on. These aspects facilitate the likelihood that desired characteristics of a balanced assessment system, as represented in the top right cell, will be observed. The bottom left cell includes aspects of assessment culture that are likely to interfere with efforts to establish more balanced assessment systems. For example, a hyper-focus on data trends, use of assessment mainly for accountability, rigid structures focused on test prep. These undesirable aspects of assessment culture make it more likely that a local assessment system will be less balanced as illustrated in the bottom right cell.

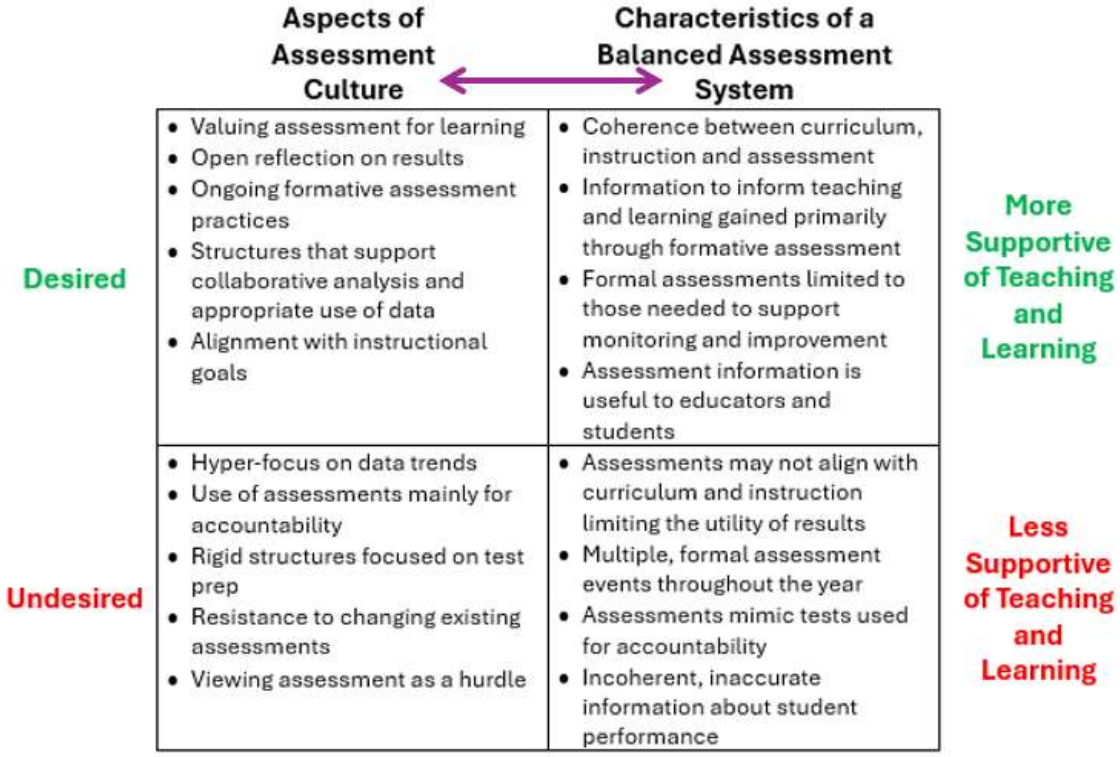

FIGURE A-1 Aspects of assessment culture that can help or hinder balanced assessment systems.

Changing an assessment system is hard, but changing assessment culture is likely harder.

Changing an assessment system is hard, but changing assessment culture is likely harder. As suggested throughout this Guidebook, focusing on small, incremental areas for improvement can go a long way in helping a school or district achieve its goals. A good place to start is by conducting a thoughtful analysis of the assessment practices and beliefs that currently exist within your system. The assessment culture continuum, developed by the Center and ISTE, was designed to support this process. The continuum includes pairs of statements that illustrate how aspects of assessment culture might manifest in practice ranging from less to more supportive of teaching and learning. The authors also sketch out a process that can be used to inform the discussion and identify priority areas for intervention.

Engaging in this type of process with an understanding of the specific challenges you need to address to improve your assessment system can help to streamline efforts even further. If, for example, a school knows that educators put too much emphasis on the state summative assessment, as seen with hours of instructional time spent teaching the test and in test preparation activities, focusing on cultural aspects related to that challenge (e.g., beliefs around the purpose of the assessment; how summative assessment results are reported and discussed within the system), can help ensure that these are accounted for when establishing a strategy for change.

APPENDIX B: THE IMPORTANCE OF AN INSTRUCTIONAL VISION

Having a clear instructional vision is a way of developing an understanding of and communicating how current research on human learning can be translated into classroom teaching and learning practices. This foundation then supports the examination of coherence or alignment among curriculum materials, teaching practices, and assessments.

What is an Instructional Vision?

An instructional vision is the image of ideal classroom practice.[45] As the name suggests, it represents a future-oriented aspiration, rather than a statement of what is actually happening in every classroom. Instructional visions often describe *what instruction should look like* and reflect what the school or district *hopes teaching accomplishes in the long run*.[46] Instructional visions can start as a general statement about instruction, but should move into the discipline-specific (e.g., science, mathematics, English language arts), because there are distinct disciplinary concerns that a district or school will want to make sure are clearly articulated and understood by all educators.[47]

> *"An instructional vision articulates what teaching and learning should look like in a particular content area. While standards name the specifics of what students should know and be able to do and the instructional shifts within the standards describe the types of instruction the standards call for, an instructional vision is a district's articulation of what students should experience daily and the overall goals for student learning."*
> (EdReports, 2020)

Research has found that instructional visions are key levers of school reform because they support coherence or consistency across and among different elements in the system, such as *curriculum* enactment, *instructional* practices and routines, *assessment* decisions, *professional development* opportunities, and *supervision/evaluation* criteria.[48] Ideally, instructional visions will be developed at the district level, with input from teachers and based on research in the relevant discipline.

---

[45] Hammerness, K. (2001). Teachers' visions: The role of personal ideals in school reform. *Journal of Educational Change*, 2(2), 143–163. https://doi.org/10.1023/A:1017961615264 ; Munter, C. (2014). Developing visions of high-quality mathematics instruction. *Journal for Research in Mathematics Education,* 45(5), 584–635. https://doi.org/10.5951/jresematheduc.45.5.0584

[46] Vaughn, M., Wall, A., Scales, R. Q., Parsons, S. A., & Sotirovska, V. (2021). Teacher visioning: A systematic review of the literature. *Teaching & Teacher Education,* 108, 103502. https://doi.org/10.1016/j.tate.2021.103502

[47] See section, "Discipline-Specific Learning," in the National Academies of Sciences, Engineering, and Medicine. (2018). *How People Learn II: Learners, Contexts, and Cultures*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24783

[48] Cobb, P., Jackson, K., Henrick, E., & Smith, T. M. (2018). *Systems for instructional improvement: Creating coherence from the classroom to the district office*. Harvard Education Press.

Instructional visions complement district mission and vision statements because they give specificity to overall system goals for students. This doesn't mean that teachers all need to teach the same way, but every teacher should have a clear picture of the overall goals for their instruction, and then they can make decisions that lead students in that direction.[49] Table B-1 provides three example instructional vision statements from EdReports—English language arts, math, and science.[50] Figure B-1 is an excerpt taken from Chicago Public Schools' *A Call to Action for Social Science in CPS*, which provides a comprehensive instructional vision for social science curriculum, instruction, assessment, and professional learning in the district.

TABLE B-1 Example Instructional Vision Statements from EdReports

English Language Arts	[District] students will develop the necessary skills in reading, writing, speaking, and listening that are the foundations for creative and purposeful expression in language. Students will engage with high-quality, complex text and participate in writing and discussions anchored in evidence from the text to promote comprehension. Students apply these skills while reading independently and see themselves as readers and writers who persevere in the face of challenging work.
Math	In [district], we strive to build mathematically proficient students prepared to meet the demands of college and career and apply mathematics in the real world. In our classrooms: ● Students do the thinking. They engage in productive struggle and work to persevere through real-world problems. Students see themselves as problem-finders and problem-solvers and have multiple opportunities to grow as mathematicians. ● Students do the bulk of the work (both thinking and talking) and actively participate in mathematics conversations. As a result, students know how to think and communicate like a mathematician, taking risks and challenging misconceptions. ● Students learn the hows and whys of math and are engaged in deep learning that builds on what they already know, and how that applies to the world around them.
Science	In [district] science classrooms, our teachers intentionally design experiences that allow students to facilitate their own learning. This leads to students who are actively engaged in explaining phenomena through discourse, collaboration, and asking questions in order to build critical-thinking and problem-solving skills.

[49] Kaufman, J. H., Wang, E. L., Kennedy, K. E., Schweig, J., & Giglio, K. (2023). *The improving instructional system coherence toolkit: A resource for K-12 districts and schools.* RAND. https://www.rand.org/pubs/tools/TLA2168-1.html

[50] EdReports (2020). *Sample instructional vision statements.* https://edreports.org/resources/article/sample-instructional-vision-statements

Center for Assessment

NATIONAL ACADEMY of EDUCATION

Social Science

An Imperative for College, Career, and Civic Life

Interrogate, Evaluate, Assess, Collaborate, Decide: Actions for a Complex World

Engaging students in powerful social science instruction is critical to our democracy and essential to the success of every child in whatever their chosen college and career pathways. To meet the demands of college, career and civic life, our students must be engaged in deep and meaningful exploration of history, geography, economics and civics—the core disciplinary domains of the Illinois Social Science Standards.

In social science classrooms students not only build critical literacy and numeracy skills, they also develop the skills, habits and dispositions to **critically interrogate the world** around them, to **evaluate and assess information** from a variety of sources, to **work collaboratively with others,** and to **make informed decisions** about who they want to be and plan how they will **take action** in the world.

In launching the new state standards in 2017, the Illinois State Board of Education described the social sciences in this way:

> ***"At its heart, social science explores the relationship between individuals and society, from friends and family to local communities and global networks. In a school setting, the disciplines of civics, economics, geography, and history are central to our students' preparation for college, career, and civic life. Through social science, students develop skills critical to success in college and careers, including creativity, critical thinking, working in diverse groups to solve complex problems, global awareness, and financial literacy."***

At CPS, students have opportunities to build, practice, and deepen inquiry skills, content knowledge, and social, cultural, and political competencies from pre-kindergarten through high school, **at every grade level, every day, and throughout the year.** The powerful learning we describe is available in some schools. Our challenge is to expand these opportunities to **all students.**

To realize the **district's vision** for rigorous, culturally sustaining classrooms that prepare all students for college, career, and civic life, high quality, stand-alone social science instruction must be made available to all students, in every school in our district.

To ensure that students are positioned for success, **CPS is redoubling its efforts** to ensure that all students have access to powerful social science instruction. Our vision for powerful instruction is informed by the new Illinois state standards, current research in learning science, culturally sustaining pedagogy, ethnic studies pedagogy, and aligned to the Common Core and the **College, Career, and Civic Life (C3) Framework.**

The social science strategic plan and implementation plan, as well as the soon-to-be-released K-12 curriculum, were developed to support stakeholders at all levels in our shared responsibility to make this vision a reality.

Call to Action

CURRENT STATE

- Eurocentric, dominant narrative.
- Fact-focused (people, places, events) with a fixed, chronological march through time.
- Teacher-centered (teacher is the holder of knowledge)
- History is the main discipline of focus.
- Skills are solely aligned to Common Core.

FUTURE STATE

- Narratives are varied and inclusive of multiple perspectives and framed in asset-based characterizations
- Inquiry learning with a thematic approach to content: Students curate, investigate, analyze, and synthesize
- Centralize and sustain student and community identity, agency and meaning-making (student-driven)
- Core disciplines integrated (History, Civics, Geography, Economics)
- Skills rooted in Inquiry: criticality, collaboration, communication, problem-solving, informed action for effective navigation in an increasingly digital and media dominant world

FIGURE B-1 An Excerpt from *A Call to Action for Social Science in CPS* (Chicago Public Schools)

Importantly, instructional visions should be based on the most updated and widely accepted research on human learning and cognition (see Appendix C for a summary), as well as disciplinary-specific instructional practices. The example instructional visions shared above do just that. Instructional visions often share common elements because they ask and answer similar questions, such as:[51]

- What academic competencies and lifelong transferable skills are prioritized and taught within and across content areas?
- What would you hope to see and hear instructionally as you walk around from classroom to classroom?
- What are teachers' primary roles in supporting student learning?
- What should students be doing to engage in learning?

Creating an instructional vision is a starting point, not a destination; other supports and structures should align with the instructional vision to support coherent implementation. For example, curricular resources, professional learning opportunities, supervision and evaluation frameworks and rubrics, and other school supports and structures should all align with and support the implementation of the instructional vision. Additionally, an instructional vision is only useful if it is communicated and there is shared meaning across a district or school. We provide a list of resources that can inform the creation of an instructional vision at the end of this Appendix.

Why is an Instructional Vision Important?

We often talk about assessment as one vertex of an instructional triangle. Assessment in schools is supposed to be aligned to and coherent with what is taught (curriculum and standards) and how it is taught to students (instruction). However, alignment and coherence among curriculum, instruction, and assessment can only be evaluated if there is a clear and shared understanding of what is supposed to be taught and how it is supposed to be taught.

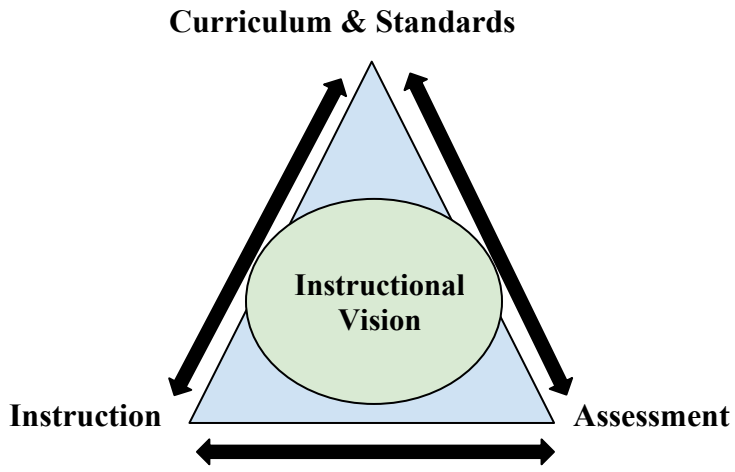

FIGURE B-2 Illustration of how an instructional vision unifies curriculum, standards, instruction, and assessment.

[51] EdReports (2021). *How to articulate an instructional vision during a materials adoption.* https://edreports.org/resources/article/how-to-articulate-an-instructional-vision-during-a-materials-adoption

As shown in Figure B-2, an instructional vision serves to unify the instructional triangle, ensuring consistency and coherence across and among all elements.[52]

> *… alignment and coherence among curriculum, instruction, and assessment can only be evaluated if there is a clear and shared understanding of what is supposed to be taught and how it is supposed to be taught.*

In the indicators presented in this Guidebook, Threat 6 focuses on a common problem of practice that can signal an imbalance in a local assessment system—namely, inconsistency between assessments and instructional vision. This poses a threat to balance because when assessments do not align with the district or school's instructional vision, they often exert a significant negative influence on curriculum and instruction. We see this most poignantly when considering the effects of the test-based accountability movement on teaching and learning.[53] Schools and districts changed their assessment practices to align with the standardized tests used for accountability, and curriculum and instruction shifted as a result.[54] Teachers not only narrowed their instruction to focus on those standards and subjects tested, but they also changed how they taught the material and how they organized the classroom learning environment.[55]

Attending to an instructional vision is an important first step in evaluating whether the instructional triangle is functioning as desired or in identifying which assessments are actively working against the larger system's goals and aspirations for teaching and learning. Rebalancing assessment systems is ultimately about making decisions about which assessments to retain, which to remove, and which to redesign, so that the entire set or system of assessments better supports a school or district's vision for curriculum, instruction, and assessment. Looking for consistency or inconsistency among the assessments implemented in practice right now with your district or school's instructional vision is a useful tool when considering the balance of your assessment system.

---

[52] Readers familiar with the features of balanced assessment systems will rightly associate this explanation with horizontal coherence, as discussed in *Knowing What Students Know* (2001). In that National Academy report, there is also an assessment triangle, which illustrates how every assessment in the system should be based on a theory of cognition, elicit observable evidence of what students know and can do, and lead to correct interpretations of students' knowledge, skills, and/or abilities.

[53] Shepard, L.A. (2008). A brief history of accountability testing, 1965-2007. In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 25–46). Routledge.

[54] Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, *36*(5), 258–267. https://doi.org/10.3102/0013189X07306523

[55] Hamilton, L. S., Stecher, B., Marsh, J. A., McCombs, J. S., Robyn, A., Russel, J. L., Naftel, S., & Barney, H. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. RAND. https://www.rand.org/pubs/monographs/MG589.html

Resources That Can Inform the Creation of an Instructional Vision

The resources listed in Table B-2 should be used in conjunction with a deliberate and inclusive process to create an instructional vision. EdReports provides an overview of a five-step process, along with content-specific resources and questions that can drive the development of an instructional vision; we refer interested readers to that resource for more information:[56]

1. Gather documents
2. Review documents
3. Consider the characteristics of effective vision statements
4. Brainstorm ideas and identify what to include in the vision statement
5. Articulate the vision statement

TABLE B-2 Resources That can be Used to Shape an Instructional Vision

Resource	How the resource can be used to shape an instructional vision
State content standards	Most state content standards begin with an opening introduction or section that outlines key design considerations or desired instructional shifts reflected in the standards. For example, the Common Core English/language arts standards open with key design considerations, such as a focus on results rather than means, an integrated model of literacy, research and media skills blended into the standards, and shared responsibility for students' literacy development (reading and writing across the curriculum). Additionally, state content standards signal instructional priorities based on how they are organized, clustered, or outlined. For example, the Next Generation Science Standards outline three dimensions: disciplinary core ideas, science and engineering practices, and cross-cutting concepts. Instruction that is based on these standards should reflect these dimensions. In other words, students should have many opportunities to learn and practice the interplay of core ideas, practices, and cross-cutting concepts in daily instruction.
State curriculum frameworks	Some states go beyond just providing content standards and also provide curriculum frameworks, either embedded within (e.g., Massachusetts) or separate from the content standards (e.g., California). Curriculum frameworks are typically intended to provide guidance for implementing the content standards adopted by the state. The curriculum frameworks explain the desired instructional shifts in more detail than what is typically found in state content standards documents. For example, curriculum frameworks usually

[56] EdReports (2021).

Resource	How the resource can be used to shape an instructional vision
	include a longer and more detailed explanation of the rationale, guiding principles, vision, organization, and scope & sequence of the standards and desired instructional shifts by grade band or grade level. Some curriculum frameworks also provide links to supplementary resources and materials that classroom educators can use to implement the standards according to the state's instructional vision.
Portrait of a Graduate	Some states and districts have worked with their community to create a Portrait of a Graduate (PoG). These portraits outline a vision of what it means to be a successful graduate from a K-12 education system. In other words, what do parents, educators, and the community believe schools should accomplish in the long term, which is an aspect of an instructional vision. PoGs typically list the knowledge, skills, experiences, competencies, and/or attributes students need to succeed in life and graduate "future-ready." See examples from Battelle for Kids, North Carolina Department of Public Instruction, and Fairfax County Public Schools.
District mission and vision statements	District and/or school mission and vision statements are typically one to two sentences and provide a high-level description of what the system hopes to accomplish in the long term for students. These statements are typically too generic to be used as instructional visions, but they do provide a good starting point to ensure that the more specific instructional vision will lead to the accomplishment of the district or school's mission and vision.
Key publications from national professional organizations	The standards-based reform movement was led, in part, by professional organizations like the National Council of Teachers of Mathematics (NCTM), which produced standards and described instructional shifts necessary to improve the quality of mathematics teaching and learning across the nation. Other professional organizations followed suit, which makes perusing these national organization's websites for standards, curriculum frameworks, or other publications a key resource (e.g., ELA/NCTE, Social Studies/NCSS, Health & Physical Education/SHAPE, Arts/NAEA, World Languages/ACTFL). These documents can be particularly helpful if you plan on creating content-specific instructional visions.
National Academy publications	National Academy publications typically review and synthesize the best and most updated thinking about disciplinary teaching, learning, and assessment. A few really well-known documents that shaped instructional shifts reflected in state and national standards and curriculum frameworks include:

Resource	How the resource can be used to shape an instructional vision
	• A Framework for K-12 Science Education • Educating for Civic Reasoning and Discourse • How People Learn & How People Learn II (Appendix C also has a brief summary of important ideas about human development and learning)

APPENDIX C: EXAMINING COHERENCE

As we noted in Section 2 of this Guidebook, balanced assessment systems are defined by five features, one of which is coherence. Assessment systems are more balanced when the various assessments—from the statehouse to the classroom—are coherently built on a common conception of student learning.[57] In other words, the way learning manifests through the assessments in a system is critical for ensuring that we are measuring what we think is essential.

Unfortunately, finding coherence in an assessment system has proved elusive, mainly because we (the larger "we" representing assessment professionals) have not done a good job defining coherence beyond "we'll know it when we see it." Also, the likelihood of having the same model of learning guiding the design of assessments from the state level to the classroom level is not feasible in current practice. That said, establishing coherence from the district to the classroom is possible because those assessments are under local control.

How Will We Know It If We See It?

Assessment should reflect what we know about how people learn and how that learning is represented in instruction and curriculum. Unfortunately, many people involved in assessment work are not steeped in learning theory. Additionally, most large-scale (including interim assessments) tests are designed to be curriculum agnostic and neutral. As Polikoff (2021) noted, standards-based reform has achieved some positive outcomes. However, it is limited in its ability to provide the fine-grained learning goals and activities found in high-quality instructional materials. We have experienced this firsthand. When we have asked colleagues in districts and states to describe their vision of learning or how they envision students to move from fragile to more complex understandings within a content domain, they often point to the state or district content standards. Content standards simply describe end-of-year expectations—hopefully based on expected trajectories of learning, but they do not clarify a vision or theory of learning for how those expectations will be achieved.

A *coherent* assessment system must be compatible with how student learning is expected to progress within an instructional domain, so making this vision explicit is a necessary first step. This might sound familiar. In Appendix B, we discuss the importance of an instructional vision as a foundation of student assessment. We link to resources in that Appendix to help district teams create an instructional vision grounded in what we know about how people learn and develop a more sophisticated understanding in a content domain. We build upon those ideas in this Appendix and expand to answer the following question: How do you evaluate whether the assessments administered to students are coherent with learning theory?

[57] National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. https://doi.org/10.17226/10019

Implications of Learning Theory on Curriculum, Instruction, and Assessment

The research literature on formative assessment and balanced assessment systems is clear—instruction, curriculum, and assessment should be coherent and aligned with current theories of how students learn. Many volumes synthesize this research, such as *How People Learn II*.[58] For our purposes, we draw on Susan Goldman and Carol Lee's (2024) chapter in *Reimagining Balanced Assessment Systems,* which summarizes much of the current research on learning.[59]

Table C-1 relies on Goldman and Lee's chapter to illustrate how leaders and educators can create the connections among learning theory, curriculum, instruction, and assessment. The first column provides a summary of important concepts in contemporary research on student learning and development. The middle column presents the implications for curriculum and instruction, since aligning assessment to best practice is of no value if students do not also have opportunities to learn the intended knowledge and skills. Curriculum and instructional practices serve as a bridge between research-based understandings about how people learn and assessment practices.[60] The right-hand column presents examples of assessment implications tied to learning theory and curriculum.

Importantly, not any theory of learning will do, even though some commonly used district and state assessments might be more coherent with outdated views of learning. We agree with Shepard, Penuel, and Davidson that the learning model underlying the assessment system must be based on the latest and most credible research.[61]

[58] National Academies of Sciences, Engineering, and Medicine (2018). *How People Learn II: Learners, Contexts, and Cultures*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24783

[59] Goldman, S. R., & Lee, C. D. (2024). Human learning and development: Theoretical perspectives to inform assessment systems. In S. F. Marion, J. Pellegrino, & A. I. Berman (Eds.), *Reimagining balanced assessment systems* (pp. 49–92). National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems/chapter3

[60] Interested readers are also directed to additional research and writing on the connections among learning theory, assessment, and instruction. For example, Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. *Handbook of educational psychology*, *77*, 15-46.

[61] Shepard, L. A., Penuel, W. R., & Davidson, K. L. (2017). Design principles for new systems of assessment. *Phi Delta Kappan*, *98*(6), 47-52. https://doi.org/10.1177/0031721717696478

TABLE C-1 How People Learn from Contemporary Theories of Learning with Implications for Curriculum, Instruction, and Assessment

Key Tenets of Current Perspectives of Human Learning and Development	Learning is fundamentally social: humans interact with other humans and the cultural artifacts that human communities create across time. Thus, knowledge is often embedded in particular social and cultural contexts, including the classroom, and encompasses understandings of the meanings of specific practices, such as asking and answering questions. Learners possess significant funds of knowledge based on their cultural and community practices and routines. As such, learners are not empty vessels; they learn by connecting new knowledge and practices to their existing social and cognitive structures. As such, students' learning and development follow varying pathways to competence. Thinking and learning are not solely cognitive activities—knowledge construction and organization involve motivational, affective, perceptual, and conceptual dimensions. The process of making sense of experiences involves drawing on prior knowledge as a resource for new learning. Knowing is more than the accumulation of factual information. Students must learn to transfer or generalize their learning. This depends on developing deep understanding in particular domains. Students develop deep understandings through exploring multiple representations of problems and/or phenomena in the particular domain while interacting with peers and more expert members of the community. Neural pathways in the brain evolve as humans observe, imitate, interact with, and take their cues from those in their cultural, experiential, and interactional contexts. Neural pathways and connections are responsive and transform throughout the lifespan. Understanding this reality contributes to growth mindsets that promote persistence—especially in the face of challenging tasks.
Implications for Curriculum, Instruction, Learning Environments	Students learn when they have opportunities to actively connect new learning to what they already know. Curriculum resources should support student learning by allowing them to discuss their background knowledge and explore developing ideas and understandings with each other.

	Making student thinking and knowledge construction processes visible provides multiple opportunities for students to practice, receive feedback from teachers and peers, and self-reflect on their learning in line with clear learning goals and success criteria. Students develop increasingly sophisticated ways of thinking about a concept over time with practice and feedback. These intermediary ways of thinking (often shown in learning progressions/ trajectories) can be built on through instruction to help students deepen their understanding. Curriculum resources that explicitly highlight these progressions help teachers deepen their pedagogical content knowledge, ultimately supporting student learning. Learning environments designed to create communities of cultural practices are not only consistent with the complexity of human learning and development but can also be a powerful means of supporting active, agentive learning in educational settings. Classroom communities of practice engage students in the active construction of knowledge, asking them to wrestle with conundrums that arise in their inquiries and to work independently and collaboratively to make sense of often conflicting information, perspectives, and values. The goals of school-based communities of cultural practices can include both knowledge construction in the moment and individual and collective development in the future.
Implications for Assessment	Assessments that inform learning must be closely aligned to where and how learning is happening, as well as how that learning and its assessment are supported. Thus, assessment must be designed hand-in-hand with instruction, and both assessment and instruction need to attend to knowledge as well as the social and affective dimensions of learning. One of the most important roles for assessment is the provision of timely and informative feedback to students during instruction and learning so that their practice of a skill and its subsequent acquisition will be effective and efficient. Formative assessment and feedback will provide the most effective ways to do this. Other forms of assessment information are more distal and often provide only numeric summaries of learning that do not specify or inform next instructional steps.

	Assessment practices need to move beyond a focus on component skills and discrete bits of knowledge to encompass the more complex aspects of student achievement. For example, assessments should focus on identifying the specific strategies children are using for problem-solving, with particular consideration of where those strategies fall on a developmental continuum of efficiency and appropriateness within a particular domain of knowledge and skill. Assessments should be designed to provide multiple entry points for students to demonstrate a range of understanding of the big ideas. Assessments need to examine how well students engage in communicative practices appropriate to a domain of knowledge and skill, what they understand about those practices, and how well they use the tools appropriate to that domain.

Examining the Coherence of Locally Controlled Assessments

To examine the coherence of district, school, and classroom assessments in relation to current perspectives on human learning and development, we suggest that you engage in a process similar to what we outline below:

1. Identify a team of teachers and school/district leaders to engage in this process and set aside the time to do it. This examination will likely take several meetings.
2. Team members will likely need to do some homework to refresh their understanding of the latest research on human development and learning. Several of the chapters in *How People Learn II,* as well as Goldman and Lee's chapter, will provide a strong foundation. These materials are completely open access and free to download and read.
3. Read and discuss Table C-1 as a team.
4. Evaluate your locally controlled assessments (district-required assessments, school-required assessments, classroom summative assessments) and associated practices (classroom formative assessment) based on the degree to which they are coherent with and aligned to Table C-1. To support the examination of coherence, the team can complete Table C-2, which supports reflection on the degree to which every assessment supports or hinders coherence (between and among learning theory, curriculum, instruction, and assessment), the district's research-based instructional vision (see Appendix B), and rich classroom learning environments (the anchor of balanced assessment systems).

This evaluation process would pair well with an assessment system inventory, audit, or review discussed in depth in the section: How to investigate and evaluate threats to efficiency. For example, teams of district and school leaders can work with classroom teachers to turn the

implications for assessment into questions that can be used to evaluate the coherence of each assessment and the collection of assessments a student experiences over the course of the year, from district-required assessments to classroom assessments, with research-based understandings of how students learn.

TABLE C-2 Reflecting on the Degree to Which Every Assessment Supports or Hinders Coherence, Instructional Vision, and Rich Classroom Learning Environments

Assessment Name & Content Area	**Intended Use(s)**	**Types of items, tasks, or activities**	Likelihood the assessment **supports coherence, instructional vision,** and **rich classroom learning environments**. Please cite specific evidence or write a rationale.	Likelihood the assessment **hinders coherence, instructional vision,** and **rich classroom learning environments**. Please cite specific evidence or write a rationale.

REFERENCES

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, *36*(5), 258–267. https://doi.org/10.3102/0013189X07306523

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(*1*). https://doi.org/10.1080/0969595980050102

Briggs, D. C., McClure, K., Student, S. R., Wellberg, S., Minchen, N., Cox, O., Whitfield, E., Buchbinder, N. & Davis, L. (2025). Visualizing and reporting content-referenced growth on a learning progression. *Educational Assessment.* https://doi.org/10.1080/10627197.2025.2503288

Chattergoon, R. & Marion, S. F. (2016). Not as easy as it sounds: Designing a balanced assessment system. *National Association of State Boards of Education,* 16*(1)*, 6-9. https://www.nasbe.org/measuring-up-recalibrating-systems-of-assessment/

Cobb, P., Jackson, K., Henrick, E., & Smith, T. M. (2018). *Systems for instructional improvement: Creating coherence from the classroom to the district office*. Harvard Education Press.

Council of Chief State School Officers. (2018). *Revising the Definition of Formative Assessment.* Council of Chief State School Officers. https://www.michigan.gov/-/media/Project/Websites/mde/2018/07/09/Revising_the_Definition_of_Formative_Assessment.pdf?rev=109be7413a2f49ba9e4c5514e9d71ff6

Donnelly, K. (2015). *Obama Calls to Curb Standardized Testing in Schools*. NBC News. https://www.nbcnews.com/news/us-news/obama-calls-curb-standardized-testing-schools-n451741

EdReports (2020). *Sample instructional vision statements.* https://edreports.org/resources/article/sample-instructional-vision-statements

EdReports (2021). *How to articulate an instructional vision during a materials adoption.* https://edreports.org/resources/article/how-to-articulate-an-instructional-vision-during-a-materials-adoption

Evans, C. M. (2025). "Eliminating Unnecessary District-Required Assessments." https://www.nciea.org/blog/eliminating-unnecessary-tests/

Evans, C. M. & Marion, S. F. (2024). *Understanding instructionally useful assessment*. Routledge.

Finnan, C., & Swanson, J. D. (2000)**.** *Accelerating the learning of all students: Cultivating culture change in schools, classrooms, and individuals.* Routledge.

Goldman, S. R., & Lee, C. D. (2024). Human learning and development: Theoretical perspectives to inform assessment systems. In S. F. Marion, J. Pellegrino, & A. Berman (Eds.), *Reimagining balanced assessment systems* (pp. 49–92). National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems/chapter3

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. *Handbook of Educational Psychology*, *77*, 15-46.

Hamilton, A., Hattie, J., & Wiliam, D. (2023). *Making room for impact: A de-implementation guide for educators*. Corwin Press.

Hamilton, L. S., Stecher, B., Marsh, J. A., McCombs, J. S., Robyn, A., Russel, J. L., Naftel, S., & Barney, H. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. RAND. https://www.rand.org/pubs/monographs/MG589.html

Hammerness, K. (2001). Teachers' visions: The role of personal ideals in school reform. *Journal of Educational Change*, *2*(2), 143–163. https://doi.org/10.1023/A:1017961615264

Harrison, C. J., Könings, K. D., Schuwirth, L. W., Wass, V., & Van der Vleuten, C. P. (2017). Changing the culture of assessment: The dominance of the summative assessment paradigm. *BMC Medical Education*, *17*(1), 1-14. http://dx.doi.org/10.1186/s12909-017-0912-5

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Heritage, M., & Wylie, E. C. (2020). *Formative assessment in the disciplines: Framing a continuum of professional learning*. Harvard Education Press.

Hinde, E. R. (2005). School culture and change: An examination of the effects of school culture on the process of change. *Essays in Education*, *12*(1), 1-12. http://openriver.winona.edu/eie/vol12/iss1/5

Jerald, C. D. (2006). School culture. Center for Comprehensive School Reform and Improvement. https://files.eric.ed.gov/fulltext/ED495013.pdf

Kaufman, J. H., Wang, E. L., Kennedy, K. E., Schweig, J., & Giglio, K. (2023). *The improving instructional system coherence toolkit: A resource for K-12 districts and schools.* RAND. https://www.rand.org/pubs/tools/TLA2168-1.html

Klotz, L. (2021). *Subtract: The untapped science of less*. Flatiron Books.

Knoster, T. (1991) *Factors in managing complex change*. Material presentation at TASH conference, Washington D.C.

Lippitt, M. (1987) *The managing complex change model*. Enterprise Management Ltd.

Marion, S. F., Thompson, J., Evans, C., Martineau, J., & Dadey, N. (2019). *The challenges and opportunities of balanced systems of assessment: A policy brief*. The National Center for the Improvement of Educational Assessment. https://files.eric.ed.gov/fulltext/ED598421.pdf

Marion, S. F., Pellegrino, J. W., & Berman, A. I. (2024). Reimagining balanced assessment systems: An introduction. In Marion, S. F., Pellegrino, J. W., & Berman, A. I. (Eds.). *Reimagining balanced assessment systems*. National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems/chapter1

Marion, S. F., Pellegrino, J. W., & Berman, A.I. (Eds.)(2024). *Reimagining balanced assessment systems*. National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems/

Munter, C. (2014). Developing visions of high-quality mathematics instruction. *Journal for Research in Mathematics Education,* 45(5), 584–635. https://doi.org/10.5951/jresematheduc.45.5.0584

Nasir, N. S., Lee, C. D., Pea, R., & McKinney de Royston, M.(Eds.). (2020). *Handbook of the cultural foundations of learning.* Routledge.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24783

National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. The National Academies Press. https://doi.org/10.17226/25389

National Academies of Sciences, Engineering. and Medicine. (2020). *Building educational equity indicator systems: A guidebook for states and school districts*. The National Academies Press. https://doi.org/10.17226/25833

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. https://doi.org/10.17226/10019

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, *28*(3), 5-13.

Peurach, D. J., & Russel, J. L. (2024). Developing, implementing, and institutionalizing complex educational innovations: Considerations for balanced assessment systems. In Marion, S.F., Pellegrino, J.W., & Berman, A.I. (Eds.). *Reimagining balanced assessment systems*. National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems/chapter8

Polikoff, M. S., & Hutt, E. L. (2024). The struggle to implement balanced assessment systems: Explanations and opportunities. In Marion, S. F., Pellegrino, J. W., & Berman, A. I.

(Eds.). *Reimagining balanced assessment systems*. National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems/chapter2

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14.

Shepard, L. A. (2008). A brief history of accountability testing, 1965-2007. In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 25–46). Routledge.

Shepard, L. A., Penuel, W. R., & Davidson, K. L. (2017). Design principles for new systems of assessment. *Phi Delta Kappan*, *98*(6), 47-52. https://doi.org/10.1177/0031721717696478

Student, S.R. (2022). Vertical scales, deceleration, and empirical benchmarks for growth. *Educational Researcher*, 51(8), 536-543. https://doi.org/10.3102/0013189X221105873

Vaughn, M., Wall, A., Scales, R. Q., Parsons, S. A., & Sotirovska, V. (2021). Teacher visioning: A systematic review of the literature. *Teaching & Teacher Education,* 108, 103502. https://doi.org/10.1016/j.tate.2021.103502

Wiliam, D. (2018). *Embedded formative assessment* (2nd ed.). Solution Tree Press.

Wylie, E. C., & Landl, E. (2024). *Understanding assessment culture.* The National Center for the Improvement of Educational Assessment and ISTE. https://www.nciea.org/wp-content/uploads/2024/08/Understanding_Assessment_Culture.pdf