ASSESSING 21ST CENTURY COMPETENCIES:

Guiding Principles for States and Districts

January 2025

Chris Brandt, Carla Evans & Chris Domaleski
National Center for the Improvement of Educational Assessment

Center for Assessment

ACKNOWLEDGMENTS:

SUGGESTED CITATION:

Brandt, W.C., Evans, C. M., & Domaleski, C. (2025). *Assessing 21st century competencies: Guiding principles for states and districts.* Center for Assessment.

ASSESSING 21ST CENTURY COMPETENCIES:
Guiding Principles for States and Districts

Center for Assessment

TABLE OF CONTENTS

ASSESSING 21ST CENTURY COMPETENCIES:

Guiding Principles for States and Districts

INTRODUCTION

Students need far more than traditional academic skills to succeed in school, work and civic life. Decades of discussion about the knowledge, skills, attitudes and behaviors that facilitate success in the information age have produced consensus that students should learn 21st century competencies. However, many schools struggle to meaningfully integrate them into instruction and gauge how well students are mastering them.

Increasingly, state departments of education and local school districts are responding with support, but key challenges stand in their way. How do we define these competencies, particularly when definitions vary across research traditions like psychology, sociology, and education? How should schools instruct and assess them? Can we produce trustworthy direct assessments[1] of 21st century competencies (e.g., ethical thinking)? How can (or should) school accountability indicators be used to signal which competencies are most important and how to improve instruction and learning of these competencies? This report is for state and local policymakers and educators who wrestle with these questions.

This report is organized in five sections. Section one briefly proposes a consensus definition of 21st century competencies. Section two explains why states and districts are making assessment of these competencies a priority. Section three presents the current state of the field in measuring these competencies and describes challenges that emerge from various limitations and constraints. Section four considers the implications for consequential use of assessment results. Section five offers practical guidance for policymakers and practitioners. Our proposed guidance includes foundational prerequisites and forward-facing action steps to cultivate students' 21st century competencies and mitigate unintended consequences. Collectively, these actions are designed to support:

- **Clear, consensus-based frameworks** for 21st century competencies,
- **Access to high-quality assessment resources and training** to apply these competencies across diverse learning contexts,
- **Formative assessment practices**, which focus on developmental feedback without high-stakes consequences, and
- **Cautious and research-grounded deliberation** for those considering assessments of 21st century competencies for high-stakes, consequential uses.

---

[1] Direct assessments measure student learning through tangible evidence such as tests, projects, or presentations. These assessments provide observable and measurable outcomes tied directly to learning objectives. Indirect assessments gather information about learning by assessing perceptions or reflections of learning, such as through surveys, interviews, or self-assessments. Indirect assessments suggest learning rather than providing direct evidence of it.

Center for Assessment

21ST CENTURY SKILLS: A CONSENSUS DEFINITION

Competing Names for Complex Competencies

Although most educators and employers agree on the importance of 21st century competencies, there's no agreement on what to call them. Figure 1 presents frequently used labels. They are also commonly known as "21st century skills," "success skills," "soft skills," "durable skills," "complex skills," "employability skills" and "trans-academic skills," among others. For this report, we've chosen to use the term "21st century competencies." Our rationale is explained in more detail below the figure.

Figure 1: Common Terms Related to 21st Century Competencies

- Complex skills
- Deeper learning skills
- Dispositions and ethics
- Durable skills
- Employability skills
- Executive function
- Future competencies
- Habits of mind
- Human skills
- Interdisciplinary skills
- Non-cognitive skills
- Trans-academic skills
- Transformative competencies
- Transportable skills
- Transferable skills
- Translatable skills
- Social emotional learning (SEL) skills
- Self-management skills
- Skills for the future
- Soft skills
- Success skills
- Work study practices

Though we recognize inherent flaws with this term—most notably its non-descriptive and time-bound nature as the 22nd century looms closer—"21st century skills" is the most widely recognized and used term in education circles. Google search data show it's the term most often associated with this wide-ranging group of competencies, which can include collaboration, communication, persistence, self-directed learning, and creative, analytical and ethical thinking. Moreover, while terms such as "soft skills" are popular in business circles, such references tend to focus narrowly on emotional intelligence, leaving out important cognitive skills like critical and analytical thinking. And though "trans-academic" may arguably be the most accurate term for these competencies, to date it is not widely used.

Notably, the name we've chosen replaces "skills" with "competencies" because "competencies" captures more than skills. Most of these competencies also include knowledge, thoughts, beliefs, attitudes and behaviors.

Consider the construct of student agency as an example. Motivation and self-efficacy are essential underlying mechanisms of student agency. Students exercise agency when they have good reason to act and when they believe that a particular action will move them toward their goal. Therefore, if a teacher wants to cultivate her students' agency skills, she must also consider how to cultivate the kinds of thoughts, attitudes, and beliefs that will elicit skills and behaviors associated with student agency.

This paper does not seek to advocate for one name; the authors themselves continue to debate which name best describes these competencies. Nonetheless, we do suggest that state departments and local school districts thoughtfully consider the various umbrella terms and their strengths and weaknesses, choose one, and clearly define the specific skills and competencies that fall underneath it. (For more detailed guidance on this, see our principle on establishing competency definitions and a common language, on page 17.)

A Consensus Definition

The Organization for Economic Cooperation and Development (OECD) ([Foster & Piacentini, 2023](#)) defines 21st century competencies as:[2]

> *the knowledge, skills and attitudes necessary to be successful for living and working in the 21st century global knowledge economy, to participate appropriately in an increasingly diverse society, to use new technologies effectively, and to adapt to change and uncertainty.*

We selected this as our working definition because it points to the interrelationships among knowledge, skills, attitudes and behaviors; they are inextricably intertwined. Certain attitudes influence knowledge and skill development, which in turn improve the behaviors that reflect the competency. Moreover, competencies themselves are intertwined. For example, as competencies like growth mindset improve, a student's propensity to master other competencies also improve. Additionally, the OECD's definition is grounded in recent research on human learning and development. The definition's emphasis on technology use, uncertainty, and adaptation reflects modern teaching and learning frameworks.

Characteristics of 21st Century Competencies

These competencies share four main characteristics. They are:

1. *Teachable:* 21st century competencies are instructionally malleable and can be developed in school-based environments (as well as out-of-school learning experiences), though they are not connected exclusively to a single academic content area, grade, or course of study.

2. *Applied across domains:* They are applied in many fields, contexts and subject domains. In fact, the competencies are often referred to as "life skills" because they are essential in thriving both personally and professionally.

3. *Complex and overlapping:* These competencies are multidimensional and frequently overlap in application. For example, creative thinking is often measured by a student's ability to generate a wide range of diverse ideas. However, it also relies on analytical and critical thinking skills, such as recognizing patterns, making reasoned inferences, and selecting the best option from a range of creative possibilities.

4. *Reflective of higher-order skills:* Higher-order skills enable an individual to transfer knowledge to new and unfamiliar situations, cope with complex problems/uncertainty, and adapt to unpredictable situations.

Teaching for these competencies raises more philosophical, policy, and technical issues than teaching academic content alone. Moreover, because 21st century competencies are so complex, inextricably linked and intertwined, assessing and measuring these competencies is particularly challenging. This is especially true if the desired result is to isolate and say something specific about any one competency as distinct from content and other competencies. Given the potential complications, is it worth the time and resources for states and districts to develop high-quality assessments of 21st century competencies? We address this question below.

> *Because 21st century competencies are so complex, inextricably linked and intertwined, assessing and measuring these competencies is particularly challenging.*

---

[2] OECD labels these skills "complex skills."

WHY EDUCATION AGENCIES ARE PRIORITIZING 21ST CENTURY COMPETENCIES

Student success requires far more than traditional academic skills. Universal access to information expands opportunities, opens new doors, and improves quality of life. At the same time, it creates new and more complex challenges as knowledge creation and change accelerate at an unprecedented pace. Consider, for example, that two thirds of jobs in the U.S. and Europe, and about one quarter of all work currently being done by humans, could be replaced by generative AI (Hatzius et al., 2023). Such a monumental shift suggests that a host of new occupations will emerge that require skills that likely do not exist today. And as AI surpasses the technical prowess of humans, qualities that are distinctly human—the ability to care, listen, regulate emotions, empower, and reason ethically—become increasingly important (Kratz, 2023).

Living in the information age is also more likely to demand global connections. In today's globalized world, skills like intercultural understanding and a global mindset are essential for individual and organizational success (Brandt, 2023). The relative importance of these skills accelerated post-COVID. Today, the rising number of remote workers makes cross-cultural and international collaboration and partnerships more likely. Organizations such as UNESCO and OECD advocate for the integration of intercultural education in primary and secondary schools, arguing that it not only enhances students' social and emotional development but prepares students for a competitive global workforce (Deardorff, 2020; OECD, 2019). By equipping students with the necessary skills to engage respectfully and effectively with different cultures, schools can promote a more inclusive and harmonious society.

Recent research gives credence to these claims. For example, one analysis of over 80 million job postings between 2021-22 revealed that seven of the 10 most requested skills requested by employers were 21st century competencies (Cole et al., 2021). Moreover, employers requested 21st century competencies almost four times more than the top five "hard" skills—technical skills that are specific to a job or industry such as project management, computer programming, accounting, foreign language fluency, or technical writing. Similarly, the World Economic Forum's most recent Future of Jobs Report (WEF, 2023) suggests that 21st century competencies are in high demand. Results from a worldwide employer survey revealed that, across industries and sectors, nine of the top 10 skills that employers consider "core to their workforce"—including their top five—are 21st century competencies. The top five skills included:

1. Analytical thinking
2. Creative thinking
3. Resilience, flexibility, and agility
4. Motivation and self-awareness
5. Curiosity and lifelong learning

K-12 and higher education priorities are shifting to address employers' needs. In the U.S., 20 states have completed development of Portraits of a Graduate ("Portraits"), frameworks that identify and define the competencies students should have by the time they graduate from high school (CASEL, 2024). Several other states are currently drafting Portraits. Not surprisingly, a recent analysis of states' Portraits suggests that the competencies they identify largely reflect the competencies that employers need (CASEL, 2024). Common examples include critical thinking,

social awareness, communication, self-management, and collaboration. Figure 1 shows the most commonly cited skills and competencies in states' Portraits.

Figure 2: Most Commonly Cited Skills & Competencies in States' Portraits of a Graduate[3]
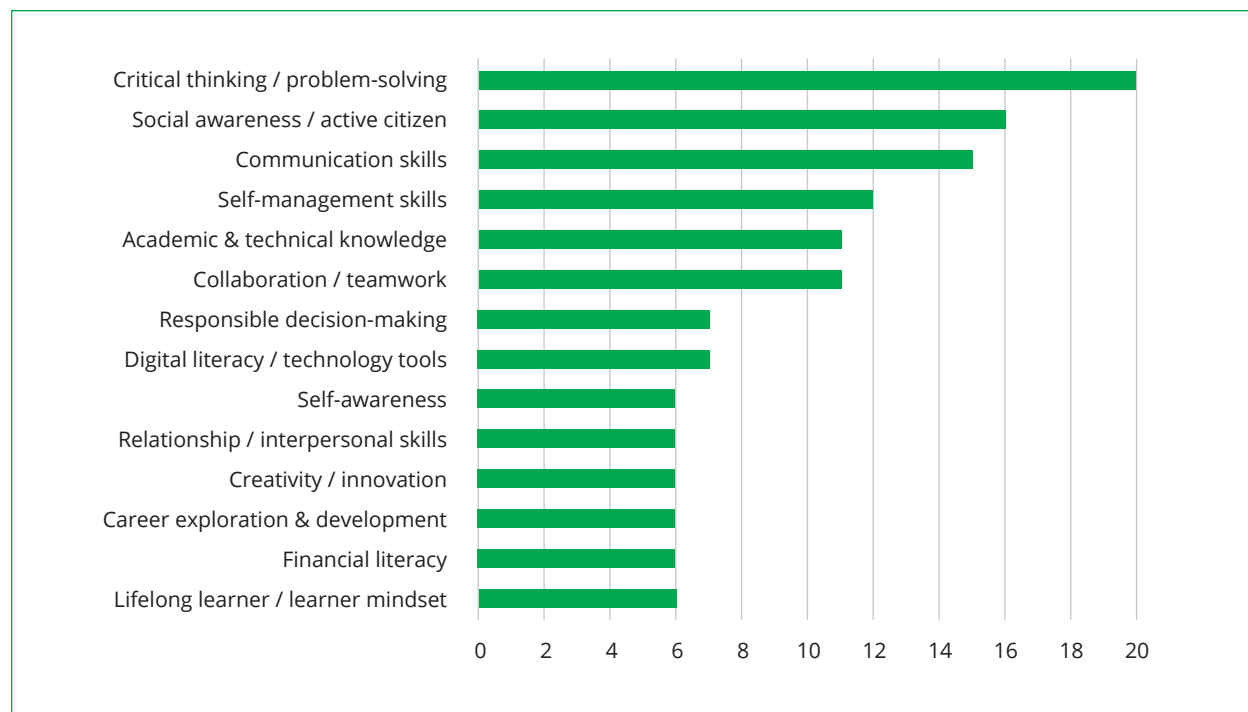
Although we don't know the specific number of school districts that have similar Portraits, many districts have vision and mission statements on their websites that include 21st century competencies. States and districts are using Portraits and vision/mission statements to prioritize and communicate the competencies that matter most for future student success. These should guide plans for designing/selecting and implementing curriculum and instructional resources, teacher training, and assessments. When designed and implemented well, such plans can promote deeper and more authentic and meaningful learning experiences for students.

Assessment plays an especially important role in this process. Given limited time and resources, the reality in schools is that if we don't assess a skill, it doesn't get attention. When designed and executed well, assessments serve essential evaluative, predictive, diagnostic, and instructional purposes. They provide feedback to end users within the educational system that inform improvements across levels of the system, with the end goal of preparing students to thrive in work and life. Assessments also signal what is important and valued.

That said, designing and using direct assessments to cultivate 21st century competencies carries significant challenges that any state or local education agency must consider. Below we discuss the most salient challenges.[4]

---

[3] Figure adapted from CASEL. (2024). *Portraits of a Graduate: Strengthening career and college readiness through social and emotional skill development.* CASEL. casel.org/portraits-of-a-graduate-2024/?view=19

[4] Most of these challenges exist with traditional assessments, but they often are magnified in the context of designing assessments of 21st century competencies.

ASSESSMENT AND MEASUREMENT CHALLENGES

What makes assessment of 21st century competencies so difficult? To understand the major challenges, it can be helpful to consider the basic requirements needed to assess or measure something well. Evans, Thompson, & Brandt (2021) proposed the following minimum requirements for drawing inferences about what a student knows and can do within a content area and grade level:

- *A clear construct definition.* What do you want to assess?[5]

- *Empirically-based performance targets.* What are students expected to know and do, at what level of cognitive rigor, and at what specific markers of time (for example, by the end of grade 3, grade 5, grade 8 and grade 12)?

- *Identified claims and uses.* It's essential to understand the claims you want to make about students' skills and abilities. What is it you want to say about students, and for what purpose?

- *Adequate domain maps.* What kinds of evidence are sufficient to support the assessment's intended claims and uses? How specific or general do you want the claims to be?

- *Reliable scoring and reporting processes.* What degree of precision is adequate to support the assessment's intended claims and uses? What is your tolerance for being wrong, given the consequences associated with the intended use(s)? How will results be reported so they reflect intended uses?

Using these minimum requirements for assessment design, interpretation, and use, in Table 1 we summarize key challenges in the design and use of 21st century competency assessments.

Table 1: Challenges in Assessing 21st Century Competencies

BASIC REQUIREMENTS	KEY QUESTIONS	CHALLENGES FOR 21ST CENTURY COMPETENCY ASSESSMENT
Clear construct definitions	What do you want to assess?	• Definitions vary across research traditions, domains, cultures, and abilities
Empirically based proficiency targets	What are students expected to know and do, and at what level of cognitive rigor at specific markers in time?	• Limited understanding about how these competencies develop and should be taught • Learning progressions tend to be non-linear and are generally based on limited research evidence • Content, culture, and context can easily conflate assessment design and interpretations

---

[5] We define assessment broadly as a systematic process of reasoning from evidence about what students know and can do and at what level of complexity (NRC, 2001). Assessment in schools is used to make inferences about individuals, groups, or programs. Quantitative and qualitative measures and tools can be used within the assessment process with the goal of providing feedback that can be interpreted and used by educators and students to support teaching and learning. This feedback can operate at the classroom level, informing instruction in real time, or at the state level, guiding evaluation, long-term planning, and systemic support. Some assessments take the form of quantitative measurement, but measurement alone is insufficient to provide effective feedback for student learning (Brookhart, 2020). Therefore, we treat assessment as an umbrella term that can include measurement, but also may not if the underlying qualities are not amenable to measurement as quantities (Briggs, 2022).

BASIC REQUIREMENTS	KEY QUESTIONS	CHALLENGES FOR 21ST CENTURY COMPETENCY ASSESSMENT
Identified claims and uses	What is it that you want to say about students, and for what purpose and use?	• Complex constructs are difficult to tease apart and accurately assess or measure in isolation • Many competencies are hard to observe because they represent internal processes that may not show up in tangible student work products
Adequate domain mapping	What collection of evidence is adequate to support the assessment's identified claims and uses?	• A domain map representing sufficient evidence to support general competency-based claims is not feasible (e.g., students' application of critical thinking in math is not the same as application of critical thinking in art history) • Competencies are inseparable from content and context, limiting the extent to which any single assessment can be used to make general ability claims
Reliable scoring and reporting processes	What degree of score reliability is adequate to support the assessment's identified claims and uses? How will results be reported so they reflect intended uses?	• Application of knowledge and skills requires item types beyond selected and constructed response. However, establishing comparability in performance assessment scoring is resource- and labor-intensive • Useful assessment reporting provides narrative feedback on process and product, so quantitative scores have limited utility

Clear Construct Definitions

A clear and research-based definition of what you want to assess is foundational to robust assessment development practice. Failure to clearly define and label skills and competencies leads to what has been called the jingle-jangle problem in educational research (Duckworth et al., 2019). Jingle-jangle arises when similar constructs are labeled with different names or different constructs are labeled with the same name. This lack of clarity inhibits the development of effective assessments. For example, "collaborative skills" and "teamwork" might be used interchangeably but assessed separately, even though they largely overlap in meaning and required competencies.

Empirically Based Proficiency Targets

Performance targets are benchmarks that define the specific knowledge, skills, attitudes, and/or behaviors that students should demonstrate at various stages in their schooling (e.g., end of grade 3, 5, 8, and 12). They help educators track student progress over time and ensure that learning goals align with expected developmental milestones and educational standards. Three main challenges emerge when establishing proficiency targets for 21st century competencies:

Empirical knowledge about how 21st century competencies develop is limited
Research-based evidence describing how most 21st century competencies develop and should be taught is limited (Soland et al., 2013). As a result, research-based learning continua—the developmental

Center for Assessment

milestones describing the typical path learners take to acquire and deepen skills over time—do not exist for most 21st century competencies. Those that do rely largely upon theoretical growth patterns that have not been empirically tested across large and diverse populations of students.

Learning trajectories are often non-linear

A scarce but growing research base on these competencies suggests that many do not follow a linear developmental trajectory. For example, constructs like self-efficacy and motivation are often cited as 21st century competencies, and they act as driving forces behind other 21st century competencies; namely, self-regulation and student agency (Bandura, 2001). There is evidence suggesting that, for many students and subject domains, self-efficacy and motivation may decline in middle school before bouncing back in high school. These dips may influence a student's ability to exercise her agency or manage her learning. Dips also have major instructional implications for teachers who want to cultivate student agency (or self-management) skills. Continuing with the self-efficacy example, a middle school teacher would want to:

1. Be especially attentive to supporting students' self-efficacy while maintaining high academic standards;

2. Consider instructional approaches that balance students' independence (e.g., voice and choice in their learning) with sufficient scaffolding and feedback; and

3. Find ways of personalizing instruction for students demonstrating diverse levels of content knowledge.

It is unlikely that a single or even multiple assessment events could holistically elicit evidence of student agency *and also* capture and report feedback on the underlying mechanisms (self-efficacy, motivation) that influenced a student's performance. Even if the assessment focused on a narrower construct—say, motivation in biology—it would still face challenges in capturing the dynamic nature of self-efficacy or motivation, which fluctuate based on factors like interest in the topic, self-belief, and social influences in the classroom environment. This example highlights the complexities associated with designing assessments that elicit evidence of outcomes (i.e., student work products) and also provide insights into the underlying, shifting mechanisms driving student engagement and agency over time (i.e., the learning process).

Context, culture, and content can easily conflate assessment design and interpretations

Context and culture influence how students think about and express these competencies. For example, a product's *novelty* and *usefulness* are two essential factors differentiating high and low levels of creative output. However, not all students may see it that way. Research suggests that students from many Asian cultures consider a creative product's *usefulness* as more important than *novelty*. Because students' conceptions of creativity (like other 21st century competencies) are so deeply entrenched, and culturally influenced, assessments of these skills demand very concrete and specific expectations related to what counts as creative.

Moreover, a student with deep content knowledge in science is more likely to score high on creativity assessments in science than a student with limited content knowledge. Thus, content knowledge must be taken into account when designing and scoring the assessment. Otherwise, a student's creativity score in science may actually be more representative of their content knowledge than their creative abilities.

While these issues are prevalent in traditional content-based assessments, they tend to be magnified when designing assessments of complex, multi-dimensional, and overlapping

skills. Moreover, these issues make it difficult to determine student mastery or proficiency at one point in time, not to mention monitoring growth over time. We say more about these issues below, since they influence all aspects of assessment design and implementation to varying degrees.

Identified Claims and Uses

Clearly specifying student claims—the claims that establish what a student knows and can do at key milestones—is a foundational step in assessment design. This is because student claims, *when aligned with the target competency's definition*, inform item and performance task design and criteria for evaluating student performance; two resources at the core of sound performance assessment. However, ensuring strong alignment between the identified claims and competency definition is itself a complex task. At least two factors influence claim-to-competency alignment challenges: (1) the competency's complexity and interrelationships with other competencies, and (2) the ability to directly observe evidence of the competence via student performances or samples of work products. We explain in more detail below.

Complex constructs are difficult to tease apart and accurately assess

Because 21st century competencies are multifaceted, it is difficult to capture a student's capabilities in a single assessment or a rubric. Ensuring that the assessment accurately reflects the skills or knowledge intended to be assessed without introducing over-simplifications or unintended biases is especially challenging. First, consider assessments of collaboration. The challenge lies in developing a tool that captures authentic collaboration skills rather than peripheral behaviors like participation frequency, which could over-simplify the construct.

Second, consider highly overlapping skills such as analytical, creative, and critical thinking (Brandt & Lorie, 2024). Figure 2 illustrates the similarities and differences among these skills. These competencies have more common than distinct skills. Further, the distinctions among these competencies are subtle. When creating an assessment of analytical thinking, designers must be careful to balance items and tasks that elicit evidence of analytical thinking without over-emphasizing skills like evaluation or idea-generation; skills associated with critical and creative thinking, respectively. Furthermore, designers must also balance depth of information elicited during the assessment experience with practical constraints such as time and available resources. These factors collectively influence how results will be interpreted and used by end users such as teachers, students, and administrators.
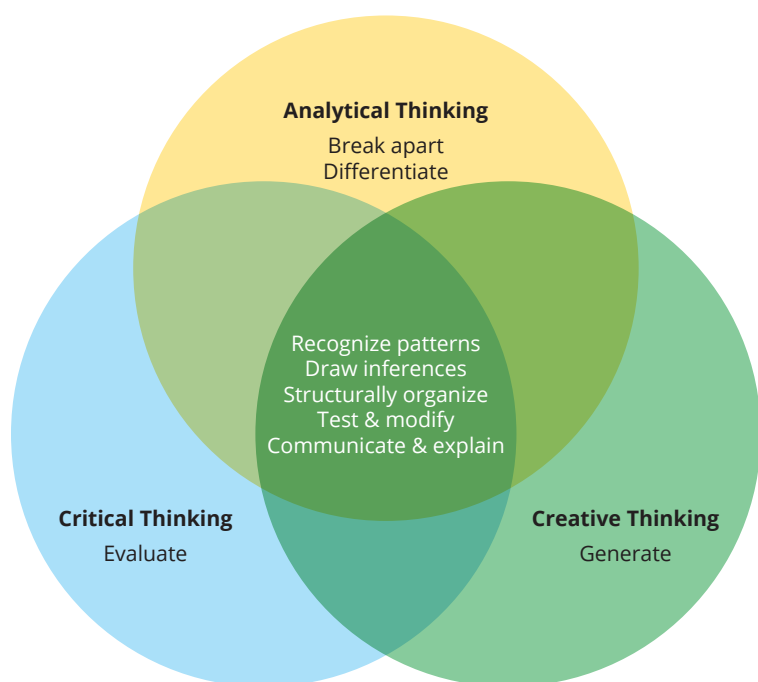

Figure 2: Primary Skills Associated with Analytical, Critical and Creative Thinking

Content-based skills, such as solving a math equation or understanding historical facts, often have clear knowledge frameworks and well-defined outcomes. However, 21st century competencies incorporate skills that are often harder to observe. How does a teacher gather tangible evidence of a student's determination and persistence beyond collecting student self-reports, which are susceptible to social desirability and social comparison bias? Skills like persistence and grit can be difficult to directly and objectively assess through authentic performance tasks. And states rarely have articulated standards for instructing and assessing these skills. This makes it hard to design assessments that holistically assess these skills in standardized ways.

Adequate Domain Mapping (Sufficient vs. Generalized Competency-Based Claims)

Domain maps are visual representations outlining specific knowledge, skills, attitudes, and/or behaviors that students are expected to acquire within a domain of study. They are useful for ensuring that assessments are aligned with the intended curriculum and learning standards.

Sufficiency and **generalizability** are two important concepts in educational assessment. **Sufficiency** refers to the extent to which the assessment adequately covers the breadth and depth of the construct being assessed within a specific context. It focuses on ensuring that the assessment task or items effectively captures the full range of skills and abilities associated with the targeted construct. **Generalizability** pertains to the extent to which the assessment results can be applied or generalized to broader situations, contexts, and content areas (Marion & Evans, 2018). For any 21st century competency, making valid and reliable inferences about what a student knows and can do requires attention to both sufficiency and generalizability.

As mentioned above (see "context and cultural influences," page 11) designing assessments to elicit 21st century competencies should be grounded in context. Domain-specific instruction is one contextual factor that influences how, and how well, a student can demonstrate competency mastery. Barshay (2019) described how this challenge emerges when attempting to make claims about a student's analytical thinking ability:

> In history, students need to interpret documents in light of their sources, seek corroboration and put them in their historical context. That kind of analysis isn't relevant in science, where the source of a document isn't as important as following the scientific method.

Because what counts as evidence varies across content domains, general analytical thinking skills—like many higher-order cognitive skills—are necessary but insufficient for enabling analysis within a specific discipline (Evans, 2020). As a result, gathering sufficient evidence to make such a generalized claim is unlikely if not impossible, even when constrained to a single domain (i.e., the student has mastered critical thinking in language arts). Notably, even if it were possible, generalized statements such as these have limited instructional utility.

Reliable Scoring and Reporting Processes

Technical and practical challenges emerge when attempting to reliably score and report on 21st century competency performance. These challenges require substantial time and resource investment.

Technical and practical challenges emerge when attempting to reliably score and report on 21st century competency performance.

Establishing performance assessment comparability is resource- and labor-intensive
Typically, the best way to gather authentic evidence of most 21st century competencies is through open-ended performance tasks. An advantage of well-designed performance tasks is that they allow students more voice and choice in demonstrating competency mastery. For example, students who struggle with writing may elect to demonstrate their communication skills via an oral presentation or video-recorded documentary. However, relaxing standardization to permit these desired features introduces significant comparability challenges when scoring student work products. Achieving some degree of comparability involves well developed protocols for evaluating student work supported by robust training. Broadly speaking, steps to achieve and maintain comparable scoring include regular intervals of:

- Identifying samples of student work that represent a range of performance levels,

- Annotating samples of student work products to describe the evidence associated with a specific performance level, and

- Collectively reviewing and discussing student work samples with other educators/experts to ensure consistent scoring across students.

The considerable time, resources, and expertise required to ensure comparable scoring is often beyond the capacity of a state department, local districts and schools.

Useful assessment reporting requires more than quantitative scores
Reporting on 21st century competencies is also challenging. As we suggested above, useful assessment results must focus as much or more on the *learning process* than the *end product*. Gathering and evaluating performance during the learning process poses numerous practical challenges that influence higher development costs and longer assessment times. More specifically, development costs are higher because the performance-based items/tasks require

> *Useful assessment results must focus as much or more on the learning process than the end product.*

assessing multiple dimensions of a competency that tend to be highly correlated with one another. Additionally, more actionable information on strengths and weaknesses requires developing a larger and more diverse set of items (or task-specific scenarios) that target these various competency dimensions (Foster & Piacentini, 2023).

Time is also an issue, particularly when these tasks are completed via authentic learning experiences. Thoughtful planning to organize and support authentic learning experiences takes time. Designing and eliciting feedback throughout these experiences requires high-quality assessment resources and substantial training. Gathering evidence throughout the learning process is much more difficult than administering a stand-alone assessment and grading responses. And finally, reporting on student learning in a way that provides actionable information to assessment users typically requires more than a simple quantitative score. Narrative feedback is often necessary for end users to properly interpret scores in the correct context and use results to improve performance.

A Note on Advancements in Computer-Based Assessments

Advancing technologies such as computer simulations and artificial intelligence (AI) show promise for addressing some of the assessment challenges described above. Indeed, the assessment

industry has experienced a substantial influx of resources devoted to improving assessment and feedback of 21st century competencies. We would be remiss to not at least mention the influence of such investments on the future of assessment practice.

Computer-based assessment simulations are emerging as a promising way to improve efficiency and reduce the burden on teachers when assessing 21st century competencies (Foster & Piacentini, 2023). These simulations offer a way to embed complex, dynamic tasks into assessments that allow students to demonstrate competencies such as collaborative problem-solving, student agency, and global competence in real-time, interactive environments. By automating parts of the assessment process and capturing detailed, process-oriented data, these tools alleviate some of the manual workload for teachers while providing more nuanced insights into student competencies. The evolution of artificial intelligence shows promise of accelerating assessment advancements, particularly for 21st century competencies in elementary and secondary education (Martinez-Comesana et al., 2023).

That said, computer-based assessments of 21st century competencies are still evolving and must overcome several challenges; namely design and production costs, the lack of universal accessibility, lack of reliable scoring processes, and limited evidence of cross-cultural validity (Csapó & Molnar, 2017; Duchatelet & Donche, 2022; Foster & Piacentini, 2023). So although these technologies are quickly accelerating assessment possibilities, delivering on these solutions at a large scale remains a formidable challenge.

IMPLICATIONS FOR CONSEQUENTIAL USES OF ASSESSMENT RESULTS

The challenges described above have major implications for consequential uses of 21st century competency assessment results. For example, varying competency-based *definitions* can influence inconsistent, differential, and biased interpretations. *Developmental trajectories* for most 21st century competencies do not exist or are in the early stages of research. Theoretical and unsubstantiated developmental trajectories influence variation in how these competencies are taught and assessed. Moreover, instructional variations introduce inequities in students' opportunities to practice and learn requisite skills and behaviors. 21st century competencies are also inherently *complex and multidimensional*, and student demonstrations of these competencies tend to be *content-, context- and culturally-dependent*. These complexities influence misalignment between the competency definition, curriculum resources, and assessments. Finally, ensuring reliable scoring and useful reporting processes is time and resource intensive.

As the consequences of an assessment increase, so do the demands for strong evidence in support of validity, reliability and fairness. The challenges associated with 21st century competency assessment are steep, but perhaps not insurmountable. That said, much more research and

> *As the consequences of an assessment increase, so do the demands for strong evidence in support of validity, reliability and fairness ... Much more research and development is needed before assessments of 21st century competencies meet requirements for large-scale consequential use.*

Center for Assessment

development is needed before assessments of 21st century competencies meet requirements for large-scale consequential use.

TRANSLATING IDEAS INTO ACTION

If these assessments increasingly inform educational policy, instructional practices, and school accountability, the potential for unintended consequences grows. Inaccurate or invalid interpretations of assessment results may lead to inequitable learning opportunities and skewed perceptions of students' readiness for future challenges, ultimately affecting their long-term success in school and beyond. Therefore, ensuring reliable, valid, and context-sensitive assessments is essential to support students and to provide meaningful, actionable insights into their development (Soland et al., 2013). Below we propose some action-oriented guiding principles for state and local educators who desire to support student success via robust assessment of 21st century competencies. But first, we'd like to revisit some important baseline practices for quality assessment.

Foundational Prerequisites for Robust Assessment Practice

This section addresses five foundational prerequisites for effective assessment design, implementation and use. One can think of these as essential actions that enable high-quality assessment practices to take root, develop, and scale across schools or districts. We follow this section with seven principles that should guide high-quality design, implementation and use of assessments of 21st century competencies.

Involve a broad group of constituents
Any consequential initiative should be grounded in a thoughtful, systematic process and include a range of contributors with diverse backgrounds, skills, and experiences. Addressing 21st century competencies is no exception. A serious initiative is not exclusively or even mainly a technical endeavor, especially in the context of classroom-based teaching and learning (Soland et al., 2013). For this reason, initiatives should start by identifying a broad-based team of constituents—the people who are affected by them. In Table 2 we list some of the likely contributors and the key questions these groups will help address.

Table 2: Key Constituent Groups and Questions to Inform 21st Century Competency Initiatives

CONSTITUENT GROUP	KEY QUESTIONS
Policymakers	• What are the primary purposes for promoting and/or assessing 21st century competencies? • What are the intended—and unintended—uses of various assessments and measures? • What policies will promote outcomes we want and guard against unintended consequences? • What are the structures and procedures that need to be in place to develop and improve assessment of 21st century competencies?
Domain, competency, & skill-based experts	• What do we know from the research/ literature about the specific competencies and/or skills we want to promote and/or assess? • Can we describe the skills we want to promote, including how to identify differences in degrees of expertise and how it develops for different learners in different conditions?

CONSTITUENT GROUP	KEY QUESTIONS
Assessment & measurement experts	• What demonstrations of student learning will elicit evidence of the construct we want to assess? • What evidence will validate that assessments are valid, fair and reliable?
Education leaders & educators	• What conditions and resources will promote our intended learning goals? • What instructional strategies are most effective, and how do they vary based on factors such as students' age, background, culture, content knowledge and other factors? • How can feedback and formal score reports be designed to provide educators with the information they need to cultivate 21st century competencies?
Student, parents, & community members	• What conditions, support and feedback are most important to promote the intended learning goals? • How can families and community members support the learning goals? • How can feedback and formal score reports be designed to provide students and parents with the information they need to cultivate 21st century competencies?

Establish competency definitions and a common language

Sound instruction and assessment practices are rooted in clear assessment expectations and instructional priorities. Establishing a common framework to support 21st century competencies is a critical first step in this process. States or districts do this by creating, adopting, or adapting research-based frameworks. Such frameworks drive the selection, labels and consensus definitions that can be used to establish clear and consistent definitions. Research-based frameworks are plentiful, and taxonomies now exist to support informed selection or adoption of robust frameworks (see, for example, Harvard's EASEL Lab). Education agencies can use these frameworks to prioritize competencies and develop clear research-based definitions for assessment purposes. Frameworks can be used to develop Portraits of a Graduate ("Portraits") that are used by many states and districts to communicate priority competencies and definitions.

Though it may seem trivial, selecting, defining, and clearly communicating an umbrella-term that represents the competency-based framework is also essential. Competency-based expectations (i.e., a Portrait) can easily become buried in a sea of initiatives or confused with other agency skills and initiatives. To mitigate such confusion, we suggest working with constituents to collectively:

• Select a single umbrella term that represents well the full array of competencies selected.

• Ensure a coherent connection between the umbrella term and its represented competencies. For example, if competencies in a state's Portraits of a Graduate include cognitive, social, emotional, and civic competencies, then the umbrella term should be broadly defined to include all of them.

• Clearly define the set of priority competencies represented in the Portrait (we discuss this in more detail below).

• Develop and execute a comprehensive communication strategy to ensure common understanding and interpretations across all terms. As we discuss below, educators and

Center for Assessment

experts rarely apply the same definition to a complex competency without a robust communication and consensus-building process.

Align purpose and use of assessment within a larger vision of teaching and learning

Specifying the intended purpose and use of 21st century competency assessment is an important first step that should shape subsequent design, implementation, and necessary evidence decisions for both curriculum-embedded and external assessments.

States can support local school districts by collectively identifying the vision and goals of integrating complex competencies into schools' formal curriculum, instructional, and assessment practices. The vision should articulate (1) the educational outcomes required to realize the vision (e.g., assessments of academic and non-academic performance, participation in extracurricular activities, postsecondary readiness) and (2) the way those outcomes are likely to be met. This is often done through the creation of a theory of action. The state can co-create a theory of action with a broad group of constituents, a best practice for any educational initiative.

The theory of action should define the experiences and learning opportunities perceived as necessary for students to achieve the identified outcomes, as well as the necessary structures, interactions, and information for schools and educators to incentivize and support those opportunities. A theory of action for assessments of 21st century competencies should describe the type of assessment information needed by different constituents and how the information gleaned should be prioritized and used to positively impact teaching and learning consistent with the state's vision. Specifically,

- What information about the competency's instruction and/or performance is needed to help achieve the stated educational goals?
- What information should be generated by the state, district, school, and/or teacher?
- Who are the intended users of the assessment information, and what is each user supposed to do with the information to help achieve the educational goals?
- What are the implications of these users and uses for the assessment design, implementing, scoring and reporting features and other necessary resources?

Sometimes it is easy to jump straight to assessment without stepping back to ask about the broader educational goals and vision that 21st century competencies support. When vision and goals are kept front and center, it is easier to stay focused on identifying key levers to support the attainment of the educational goals, including the role of 21st century competency assessments.

Incentivize and support high-quality instructional materials

21st century competencies must be explicitly taught and exist within rich tasks and activities. For example, individuals do not collaborate or think critically in isolation; they collaborate or think critically *about* something—content, a topic, question, task, or activity. 21st century competencies often represent what students and adults must apply and use in the service of answering a question, solving a task, or producing a product. Because of this, 21st century competencies are best taught, practiced, and informally assessed within high-quality instructional materials (HQIM)-- especially given the technical and logistical challenges associated with large-scale consequential use (Soland et al., 2013).

Instructional materials, when well-designed, provide students with access to grade-level, standards-aligned, cognitively rigorous, and differentiated teaching and learning experiences. This is especially relevant to the application and use of 21st century competencies. All students should have the opportunity to learn, practice, self-reflect, and receive feedback on the development of their critical-thinking, collaboration, or communication skills over the course of their PK-12 careers. Most states have adopted standards that incorporate at least some 21st century competencies within the application of content and provide an opportunity for the explicit and integrated instruction of 21st century competencies. For example:

- The Next Generation Science Standards (NGSS) include science and engineering practices (SEP). SEPs require students to analyze and interpret data, engage in argument from evidence, design solutions, evaluate and communicate information.

- Complex mathematical practices are integrated into most states' K-12 mathematics learning standards. These practices include skills such as reasoning abstractly and quantitatively, constructing viable arguments, and critiquing the reasoning of others.

- The College, Career, and Civic Life Framework (C3) framework is a guide to improve social studies instruction in K-12 schools. The framework includes standards that require students to gather and evaluate sources, develop claims and use evidence, communicate and critique conclusions.

While states do not control local curricula, states can use policy levers and incentives to promote the selection/adoption and implementation of HQIM. For example, in a recent National Academy of Education report, *Reimagining Balanced Assessment Systems*, chapter 7, on state practices that support the implementation of balanced assessment systems, provides five state education agency actions related to HQIM (Evans & Landl, 2024). We adapted those five actions so they are specific to ensuring that assessments of 21st century competencies are embedded within HQIM (see Figure 3).

Figure 3. Five SEA Actions Related to HQIM from Highest to Lowest Intensity of SEA Time, Resources, and Capacity Requirements

HIGH	Intensity of SEA time, resources, and capacity requirements			LOW
Create state curriculum and instructional materials with no adoption requirement.	Conduct state-level review of curriculum and instructional materials.	Provide high-quality curriculum and instructional material review tools (or link to other entity reviews) that incorporate evaluation criteria about how complex skills are embedded in instructional tasks/activities, formative assessment processes, and classroom summative assessments.	Provide training on how to use and apply the review tools.	Incentivize the selection of HQIM by establishing financial incentives for districts that select materials from the state's recommended list.

Center for Assessment

State actions to incentivize and support HQIM adoption range from the state creating curriculum and instructional materials that explicitly embed and integrate 21st century competencies with content to the state simply incentivizing the selection of HQIM by establishing financial incentives for districts that select materials from the state's recommended list.

For example, districts that want to use state or federal funds to purchase curricula may be required to select from an approved list of high-quality curricula to qualify for those funding sources. In between these two ends of the spectrum, states can support implementation of 21st century competencies by directly reviewing curriculum and instructional materials or providing HQIM review tools so districts can conduct the review on their own. Ideally, states that decide to provide review tools would also provide training on how to interpret and best use those review tools to evaluate the quality of the 21st century competencies embedded within the materials—along with other quality criteria such as alignment to the state's content standards, cognitive demand, resources to support diverse learners, etc.

Provide support for local districts to assess and improve conditions for learning
Research suggests that conditions for learning such as a safe and respectful school climate, challenging/high expectations, student support, and social and emotional learning predict a wide variety of educational and developmental outcomes we care about (Osher, Moroney, and Williamson, 2018). For example, research shows that a student's sense of belonging—defined here as "a student's feeling of identity, inclusion, and acceptance as a member of their school community" (ODE, 2024)—has a meaningful association with **psychological** constructs (e.g., motivation, behavior, social-emotional competencies), **health** domains (e.g., mental health, high-risk substance use) and **academic** outcomes (e.g., attendance, achievement, high school graduation) (Korpershoek et al., 2020; Rose et al., 2022; ODE, 2023).

Conditions for learning are also critically important for supporting the development of 21st century competencies such as creativity, collaboration, communication, and critical thinking. If students don't feel supported or safe at school with the adults and students present, for example, why would we expect them to collaborate with their peers? Or, if the school climate fosters low expectations and offers little challenge to students, why would we expect critical thinking skills to be applied across the curriculum?

There are many research-based assessments available to monitor school climate and conditions for learning (e.g., 5 Essentials, ED School Climate Surveys, Oregon SEED survey, California's Core Districts SEL and school culture survey, Panorama Education school climate surveys, PBIS school climate survey suite, GLSEN National School Climate Survey). States could provide guidance on the use of these conditions for learning surveys and then make resources and support available to districts and schools depending on the needs identified in the survey results. For example, if a student's sense of belonging is identified as an area in need of growth from a school climate survey, then the state could provide resources and support to districts and schools about how to increase student sense of belonging in their context. Oregon is an example of one state that is supporting local school districts in this way by providing research reports and briefs that link identified weaknesses with state resources and support (ODE, n.d.).

Principles for High-Quality Assessment Design, Implementation and Use

As noted previously, the process required to develop, adopt, or adapt assessments of 21st century competencies will vary depending on which competencies are prioritized, and how the assessments will be used and by whom. However, any credible plan is likely to include the following components.

Apply the principles of evidence-centered design
The science of measuring complex constructs is not new. Perhaps the best known framework to guide the process is Evidence Centered Design (ECD) ([Mislevy et al., 2003](#)). ECD is a systematic approach that focuses on defining the skills being assessed and ensuring that items or tasks are well suited to elicit those skills. The ECD framework is based on three main components that overlap with many of the areas we have addressed elsewhere in this paper.

- The **student model** calls for developing a clear definition of the construct and how it develops along a continuum from novice to expert.

- The **evidence model** calls for describing the ways in which indications of these skills are demonstrated.

- The **task model** calls for determining which items or stimuli are likely to elicit the desired evidence.

We contend that most if not all of the 21st century competencies of interest will require more complex demonstrations that are likely to be elicited via performance-based tasks. This elevates the importance of an ECD framework to produce credible assessments.

> *Most if not all of the 21st century competencies of interest will require more complex demonstrations that are likely to be elicited via performance-based tasks.*

Create or adopt research-based continua to guide feedback and skill development
A key part of ECD involves defining the construct and developing a clear understanding of how performance is expressed along a continuum from novice to expert. Once these performance continua are well understood, developers can focus on designing tasks that elicit the intended performance and describing the evidence associated with a competency's demonstration along the continuum.

These performance continua can be used to provide guidance and feedback to help educators promote students' competency development. We caution against developing "scoring rubrics" too early in the process. Rubrics imply grading, and grading can have negative effects on student learning ([Evans, 2020](#); [Shepard, 2019](#)). This is because grading can elicit comparisons among students, which can adversely affect student motivation.

Moreover, grading 21st century competencies is fraught with potential unintended consequences, as the assessments may not be sufficiently accurate at the individual student level and distort the meaning of grades as indicators of academic achievement. For these reasons, we suggest not using the language of a rubric, especially for nascent and novel assessments. Instead, focus on creating or adopting research-based continua to describe student performance from less to more sophisticated along with guidance to educators to help inform instructional approaches.

The Center for Assessment has developed several papers and draft developmental continua to inform specific [21st century](#) competencies. Though these and other research-based continua generally have not been piloted or validated for large-scale use, states and/or local agencies can access and use these papers and draft continua (or other available continua) to do so.

Pilot and scale performance-based assessment and reporting procedures
Once initial tasks and performance continua are developed (or adopted/adapted), it's important to gather evidence to evaluate them. Educators will want to determine if the tasks are eliciting the

intended demonstrations and the extent to which these demonstrations correspond with the hypothesized continua of performance.

We suggest that states and districts exercise caution when using large-scale standardized direct assessments for consequential purposes. The greater the consequences, the greater the demands for high-quality evidence. A state that plans to incorporate student-level assessment(s) of 21st century competencies to inform high-stakes decisions or accountability should proceed with an abundance of caution.

Our suggestion to pilot and scale 21st century assessments is in service of creating robust performance assessments (and other types of assessments), which have proven useful for instructional purposes. These assessments would be developed and vetted by experts and expert practitioners for classroom-based purposes; preferably for formative use. For example, teachers might decide to access one or more performance task(s) from a bank of vetted tasks that are aligned to common curricular materials. Or an agency may support local teachers to develop, pilot, and refine performance-based assessments in collaboration with grade-level and/or subject-area colleagues.

With the above caveats in mind, a broad-based group of content and assessment experts should review the tasks and developmental continua to determine the extent to which they are aligned to the assessment objectives, clear and appropriate for the intended examinees, and free of elements that could present concerns related to bias, sensitivity, or accessibility.

> *We suggest that states and districts exercise caution when using large-scale standardized direct assessments for consequential purposes.*

Next, we recommend engaging in cognitive labs[6] to gain insights about the thought processes students use when they engage with the tasks. This might involve observing how students interact with the tasks, having them describe their reasoning as they complete the task (i.e. think-alouds), or interviewing educators or students before and after they engage with the task. Interaction studies—an analysis of teacher-student interactions and dynamics—are particularly important for assessments of novel and complex competencies to gain a "proof of concept" and to inform refinements to the tasks and continua, before they are used more broadly.

We also recommend piloting task feedback, such as continua or assessment reports, with the intended users early in the process. These pilots, which can take the form of interviews or focus groups, are not intended to serve as a "wish list" activity. In other words, an appropriate prompt for a teacher or parent is not, "What information do you want to see?" Such questions may elicit responses that are not within the scope of the assessment. Rather, it is preferable to present users with sample results and ask them to explain what it means and how they came to that conclusion. In this way, reviewing feedback is not unlike an interaction study or cognitive lab. The findings can provide good insights to ensure the feedback is clear and is being interpreted as intended.

---

[6] An authoritative guide for conducting cognitive labs can be found in Appendix C. See Lyons, S., Evans, C., Marion, S., & Thompson, J. (2017). New Hampshire Performance Assessment of Competency Education (PACE) technical manual. National Center for the Improvement of Educational Assessment, Inc. https://www.ed.gov/sites/ed/files/policy/elsec/guid/stateletters/nhpacetechmanual72017.pdf

Carefully balance sufficiency of evidence with generalized ability claims

Earlier, we discussed the challenge of balancing sufficient vs. generalized claims about what a student knows and can do. Generally speaking, performance assessments are an essential tool for gathering direct evidence of students' 21st century competence. However, performance assessments are situated within a particular context and domain. Moreover, a single performance assessment is unlikely to gather all the information a teacher would need to make a generalized claim about whether a student has mastered a 21st century competency. Therefore, claims about what a student knows and can do—and assessments designed to reflect them—must be situated within their appropriate context. When determining how much evidence is sufficient, Marion & Evans (2018) offered the following suggestions:

- Identify the intended uses of the assessment(s). Ensuring sufficiency is important in summative assessment, particularly when the stakes are high. If the focus is on formative feedback, sufficiency is less important.

- Develop explicit student claims and include transfer/generalizability claims. If you want to claim that student competence extends beyond the performance on the single assessment or set of assessments, then carefully evaluate whether the set of assessments adequately represents the target of your inferences (such as analytic writing) and provides enough information to support your decisions.

- Be clear about your tolerance for being wrong. The higher the stakes (such as denying a student a chance to progress), the more important it is to have sufficient information to support the decision.

- Carefully balance having too little information with the tradeoffs associated with obtaining more. This balance is especially important when information comes from assessments that are administered separate from instruction.

For example, imagine an educator wants to assess a student's self-management capabilities in algebra. The teacher could design (or adopt) a series of authentic and open-ended tasks designed to assess students' knowledge and understanding of key algebraic concepts. Through the series of tasks, students are asked to engage in the self-management process:

- Select (or create) a complex problem, which would require application of the focal algebraic standards.

- Establish a goal. For example, "I will use information from past election results to predict future election results within a given margin of error."

- Propose a plan that describes steps such as, "What information would I need to successfully predict future election results?", "What factors would my model ideally account for?", "What mathematical model would I use, and how would I represent variables in the model?"

- Execute the plan by executing the mathematical model, solving for "X" (solution), and effectively communicating results.

- Gather feedback and reflect on how the problem might be approached and/or communicated differently, or more effectively, with various audiences.

Each task could be structured to generate multiple types of evidence such as the teacher's direct observations, student work, self-reflection, and peer feedback. Scoring and feedback could be sequenced to support students' learning of (1) key algebraic concepts and (2) self-management during and after the series of tasks. The compendium of evidence could in turn be used to support

specific summative claims about a student's level of mastery (or proficiency). The evidence-supported claim might be: "The student demonstrated mastery in self-management skills in the context of applying inverse operations to solve for X."

Use multiple assessments to evaluate and cultivate 21st century competencies
Standardized self-report surveys are available for many 21st century competencies. Although self-report surveys can provide important, and often comparable, information, they often fail to capture the range of knowledge, skills, attitudes, and beliefs embedded in a 21st century competency. Similarly, self- and peer-assessments can provide important feedback to students about their performance and progress, but they often suffer from social comparison and social desirability bias. This limits their usefulness for summatively evaluating a student's performance. Rather, a variety of direct and indirect assessments are needed to comprehensively assess these competencies.

A comprehensive assessment of most 21st century competencies requires observational evidence collected over time. This can be accomplished by using simulated or authentic performance tasks and/or portfolios. Performance tasks and portfolios enable observations and other direct evidence to support claims about what a student knows and can do. By embedding performance assessments and portfolios (or e-portfolios) into classroom activities, educators can gather authentic evidence of complex competencies and skills, enabling students to demonstrate their ability to act intentionally, thoughtfully and independently in real-world contexts. Moreover, portfolio assessments have been linked to greater self-efficacy and achievement, particularly when used for instructional purposes to capture students' learning achievements and encourage critical reflection and improvement over time ([Lopez-Crespo et al., 2022](#)).

Prioritize formative assessment
Development and use of 21st century competency assessments are still in their infancy. Given the state of the field, we strongly urge states and districts to prioritize formative resources and tasks that produce qualitative feedback before moving into larger-scale, summative assessments. There are several reasons we caution against rushing to produce summative assessments, especially with consequential implications:

> *Given the state of the field, we strongly urge states and districts to prioritize formative resources and tasks that produce qualitative feedback before moving into larger-scale, summative assessments.*

- As a field, we have limited understanding about how most of these skills develop, how they are best learned, how they are integrated into schools, and how they vary by experience and context. We need to learn more before we can support performance claims with an adequate degree of reliability, validity, and fairness that must accompany any high-stakes use of assessment results.

- Summative assessments, especially those that are standardized to support claims of comparability, often lack sufficient meaning to inform instruction and learning. There is a substantial research base related to assessment for learning and how students learn more from qualitative, formative feedback than from scores or grades ([Black & Wiliam, 1998](#)).

- Consequential use often risks corruption. The more assessments are used for high-stakes, such as influencing school accountability ratings, the more there will be pressure to increase scores rather than focusing on the conditions for learning.

Center for Assessment

Engage in continuous improvement

Sound continuous improvement practices—including formative and summative evaluation practices—are essential for improving and scaling practices that promote 21st century competencies. Many continuous improvement paradigms exist and have promising evidence of effectiveness. States and districts can select one that values inclusivity and rigor, and use it to monitor and improve assessment for teaching and learning 21st century competencies (Brandt, Dadey, & Evans, 2024).

Most continuous improvement paradigms are anchored by three essential questions (Shakman et al., 2020):

- What specifically are we trying to accomplish?

- What change might we address and why?

- How will we know that a change is an improvement?

The state and/or local agency can use these questions to inform a theory of action, logic model, and implementation plans to support 21st century competency-based learning approaches.

Formative (iterative) approaches such as Plan-Do-Study-Act (PDSA) have become familiar to teachers and local educators. These and similar approaches can be used to systematically collect, analyze, and use classroom-based information (including assessment feedback) to improve how 21st century competencies are both taught and assessed on a short-cycle basis (e.g., weekly or monthly). Aggregate information sources—often collected via centralized data systems—are essential for summative evaluations; that is, monitoring whether and to what extent a change is working "at scale" (e.g., across a specific group of students, teachers, and schools). Summative evaluations run on longer cycles. It may take a year or multiple years to determine whether the changes enacted impact learners across a school system. While summative evaluations are rarely useful for day-to-day instructional decision making, they are essential for monitoring long-term improvements and establishing the efficacy of a set of policies, programs, and/or practices at scale.

Sound continuous improvement approaches depend on active engagement across constituent groups such as teachers, students, subject experts, data and research experts, and policymakers. Teachers, for example, can provide critical insights into the practical challenges and successes of implementing the assessments in the classroom. Students can offer feedback on the tasks' relevance and clarity, ensuring that assessments are meaningful and engaging. Research experts can design and implement data collection and analysis methods that facilitate timely decision-making. Subject-area experts can identify key facilitators and barriers to implementation and suggest approaches to address implementation pain points. When implemented well, a collaborative approach can eliminate potential blind spots and foster a sense of shared ownership over the assessment process.

SUMMARY AND RECOMMENDATIONS

In this report, we proposed a consensus definition of 21st century competencies, provided a rationale for why assessing these competencies is important, described the assessment challenges, and proposed some guiding principles for state and local education agencies to improve the teaching, learning and assessment of these competencies.

Several challenges have emerged as states work toward supporting teaching, learning and assessing 21st century competencies in local districts and schools. Major challenges include:

- An absence of shared definitions and developmental trajectories,
- Difficulties associated with eliciting evidence of these complex competencies in valid and reliable ways,
- Complications for reliably scoring and reporting on assessment experiences and student work products, and
- Substantial resource demands of implementing and reporting high-quality instruction and assessment practices.

To address these challenges, we offered key advice for state and local practitioners. Our suggestions addressed broad systemic actions for change to take hold and flourish.

- Involve a broad group of constituents in creating an expansive vision of teaching and learning and iterating evidence-based improvements over time,
- Establish clear, consensus-based frameworks for 21st-century competencies and a common language for cultivating them in schools and classrooms,
- Support local districts and schools to assess and improve conditions for 21st century competencies to thrive, and
- Incentivize and support educators with high-quality resources and training to apply these competencies across diverse learning contexts.

Additionally, we offered several key principles for states and districts to design, implement, and use high-quality assessments of 21st century competencies:

- Apply the principles of evidence-centered design.
- Create or adopt research-based continua. Continua should guide assessment development, feedback and instructional planning for skill development.
- Pilot and scale assessments and reporting procedures.
- Carefully balance sufficiency of evidence with generalized ability claims.
- Use multiple assessments to evaluate and cultivate 21st century competencies.
- Prioritize formative assessment for instructional purposes. Assessments and tools should focus on developmental feedback over high-stakes consequences.
- Engage in continuous improvement. Adopt (or adapt) a process that prioritizes collaborative, systematic, and evidence-based decision-making and involves constituents across all levels of the educational system.
- Proceed cautiously. Recognize the limitations if considering assessments of 21st century competencies for high-stakes consequential uses.

Finally, we emphasized that through collaboration among policymakers, educators, researchers, and end users—and by leveraging technological advancements in assessment—educational systems can foster meaningful and equitable 21st century competency development that aligns with the complex demands of the modern world.

REFERENCES

Alozie, N., Lundh, H., Yang, H., & Parker, C.E. (2021). *Designing for Diversity Part 2. The Equity and Inclusion Framework for Curriculum Design.* Rockville, MD: National Comprehensive Center at Westat.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology 52*, 1–26. https://psycnet.apa.org/record/1992-12707-001

Barshay, J. (2019, September 18). Why content knowledge is crucial to effective critical thinking. Mindshift. https://www.kqed.org/mindshift/54470/why-content-knowledge-is-crucial-to-effective-criticalthinking

Black, P., & Wiliam, D. (1998). *Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Brandt, W.C. (2023). *A review of the literature on intercultural understanding.* National Center for the Improvement of Educational Assessment. https://www.nciea.org/library/a-review-of-the-literature-on-intercultural-understanding/

Brandt, W.C., Dadey, N., & Evans, C. (2024). *Evaluating innovative state assessment programs: A framework for state education agencies.* National Center for the Improvement of Educational Assessment. https://www.nciea.org/wp-content/uploads/2024/05/EvaluatingInnovativeAssessmentPrograms-FINAL.pdf

Brandt, W.C., & Lorié, W. (2024). *Measuring student success skills: A review of the literature on analytical thinking.* National Center for the Improvement of Educational Assessment. https://www.nciea.org/wp-content/uploads/2024/06/AnalyticalThinkingReport-Final.pdf

Briggs, D. C. (2022). *Historical and conceptual foundations of measurement in the human sciences.* Routledge.

Brookhart, S. M. (2020). Feedback and measurement. In S. M. Brookhart & J. H. McMillan (Eds.). *Classroom assessment and educational measurement* (pp. 63-78). National Council on Measurement in Education.

CASEL. (2024). *Portraits of a Graduate: Strengthening career and college readiness through social and emotional skill development.* CASEL. casel.org/portraits-of-a-graduate-2024/?view=19

Cole, L., Short, S., Cowart, C., & Muller, S. (2021). T*he high demand for durable skills. America Succeeds.* https://americasucceeds.org/wp-content/uploads/2021/04/AmericaSucceeds-DurableSkills-NationalFactSheet-2021.pdf

Csapó B and Molnár G (2017) Potential for Assessing Dynamic Problem-Solving at the Beginning of Higher Education Studies. *Frontiers in Psychology,* 8, Article 2022. https://doi.org/10.3389/fpsyg.2017.02022

Deardorff, D. K. (2020). *Manual for developing intercultural competencies: Story circles.* UNESCO. https://www.routledge.com/Manual-for-Developing-Intercultural-Competencies-Story-Circles/Deardorff/p/book/9781032090160?srsltid=AfmBOoqxRnv3VvbPm32-ZO1cw47CbUz3ZgCXTUuh_ikBuNkXfgdc30Fy

Duchatelet D., & Donche V. (2022). Assessing student learning during simulations in education: Methodological opportunities and challenges when applying a longitudinal case study design. *Studies in Educational Evaluation*, 72 , Article 101129.

Duckworth, A., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M., & Gross, J. J. (2019). Self-control and academic achievement. Annual Review of Psychology, 70, 373-399. https://doi.org/10.1146/annurevpsych-010418-103230

Evans, C. M. (2020). *Measuring student success skills: A review of the literature on critical thinking.* National Center for the Improvement of Educational Assessment. nciea.org/library/measuring-studentsuccess-skills-a-review-of-the-literature-on-critical-thinking/

Evans, C. M., & Landl, E. (2024). Chapter 7: State practices and balanced assessment systems. In S. F. Marion, J. Pellegrino, & A. Berman (Eds.), *Reimagining balanced assessment systems* (pp. 201-236). National Academy of Education. https://naeducation.org/reimagining-balanced-assessment-systems-project/

Evans, C. M., Thompson, J., & Brandt, C. W. (2021). 21st century skills: Key Measurement & Assessment Considerations. *NCME Classroom Assessment Conference* (Virtual).

Evans, C. M., Thompson, J., & Brandt, W. C. (2020). Instructing & assessing 21st century skills: Key measurement & assessment considerations. *National Center for the Improvement of Educational Assessment.* https://www.nciea.org/blog/instructing-assessing-21st-century-skills-key-measurement-assessment-considerations/

Foster, N. & M. Piacentini (Eds.). (2023). *Innovating assessments to measure and support complex skills.* OECD Publishing. https://www.oecd-ilibrary.org/education/innovating-assessments-to-measure-and-support-complex-skills_e5f3e341-en

Hatzius, J., Briggs, J., Kodnani, D, and Pierdomenico, G. (2023, March). The potentially large effects of artificial intelligence on economic growth. Goldman Sachs Global Investment Research. https://www.ansa.it/documents/1680080409454_ert.pdf

Korpershoek, H., Canrinus, E. T., Fokkens-Bruinsma, M., & De Boer, H. (2020). The relationships between school belonging and students' motivational, social-emotional, behavioral, and academic outcomes in secondary education: A meta-analytic review. *Research Papers in Education, 35*(6), 641–680. https://doi.org/10.1080/02671522.2019.1615116

Kratz, J. (2023, December 20). The rise of AI underscores a need for human skills. *Forbes.* https://www.forbes.com/sites/juliekratz/2023/12/17/have-we-forgotten-about-human-skills-in-the-age-of-ai/?sh=44f0f3126a7d

Lopez-Crespo, G., Blanco-Gandia, M. C., Valdivia-Salas, S., Fidalgo, C., & Sanchez-Perez, N. (2022). The educational e-portfolio: Preliminary evidence of its relationship with students' self-efficacy and engagement. *Education and Information Technologies, 27*, 5233–5248. https://doi.org/10.1007/s10639-021-10827-2

Lyons, S., Evans, C., Marion, S., & Thompson, J. (2017). *New Hampshire Performance Assessment of Competency Education (PACE) technical manual.* National Center for the Improvement of Educational Assessment, Inc. https://www.ed.gov/sites/ed/files/policy/elsec/guid/stateletters/nhpacetechmanual72017.pdf

Marion, S., & Evans, C. M. (2018, September 6). How much is enough? Sufficiency considerations for competency-based assessment systems. *National Center for the Improvement of Educational Assessment*. https://www.nciea.org/blog/how-much-is-enough/

Marion, S. M., Lyons, S., & D'Brot, J. (2016). *Developing a theory of action to support high-quality accountability system design.* National Center for the Improvement of Educational Assessment. https://www.nciea.org/library/developing-a-theory-of-action-to-support-high-quality-accountability-system-design/

Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocarranza-Prado, I., & Kreibel, D. (2023). Impact of artificial intelligence on assessment methods in primary and secondary education: Systematic literature review. *Revista de Psicodidáctica* (English ed.), 28(2), 93–103. https://www.sciencedirect.com/science/article/pii/S2530380523000072

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report.*

National Center for the Improvement of Educational Assessment. (2023). *Assessing 21st century skills* [Toolkit]. https://www.nciea.org/library/assessing-21st-century-skills/

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

Organization for Economic Co-operation and Development (OECD). (2019). PISA 2018 global competence framework. In *PISA 2018 assessment and analytical framework.* OECD Publishing. https://doi.org/10.1787/043fc3b0-en

Oregon Department of Education (ODE). (2023). S*tudent sense of belonging in schools: Connection to outcomes* [Results from Statewide SEED Survey]. https://www.oregon.gov/ode/educator-resources/assessment/Documents/SenseofBelongingOutcomes.pdf

Oregon Department of Education (ODE). (2024). *Student educational equity development survey (SEED Survey): Preliminary specifications and blueprints.* https://www.oregon.gov/ode/educator-resources/assessment/Documents/SEED_Survey_Preliminary_Specifications_and_Blueprints.pdf

Oregon Department of Education (ODE). (n.d.). *State results and reporting.* https://www.oregon.gov/ode/educator-resources/assessment/pages/student_educational_equity_development_survey.aspx

Osher, D., Moroney, D., & Williamson, S. (2018). *Creating Safe, Equitable, Engaging Schools: A Comprehensive, Evidence-Based Approach to Supporting Students.* Harvard Education Press.

Rose, I. D., Hecht, A. A., Spence, E., & Washburn, A. (2022). The relationship of school connectedness to adolescents' engagement in co-occurring health risks: A meta-analytic review. *The Journal of School Nursing*, 40(1), Article 10598405221096802. http://dx.doi.org/10.1177/10598405221096802

Shakman, K., Wogan, D., Rodriguez, S., Boyce, J., & Shaver, D. (2020). *Continuous improvement in education: A toolkit for schools and districts* (REL 2021–014). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. https://ies.ed.gov/ncee/rel/regions/northeast/pdf/REL_2021014.pdf

Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The Annals of the American Academy of Political and Social Science, 683*(1), 183–200. https://doi.org/10.1177/0002716219843818

Soland, J., Hamilton, L. S., & Stecher, B. M. (2013). *Measuring 21st century competencies: Guidance for educators.* Washington, DC: RAND Corporation. https://digitalpromise.org/wp-content/uploads/2022/10/gcen-measuring21cskills.pdf

World Economic Forum. (2023). *The future of job report, 2023.* World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2023/

www.nciea.org