Competency-Based Education Research Journal, 2026
https://journals.ku.edu/cberj/

Considerations for Advancing Research on Competency-Based Assessment in K–12 Education

Teresa M. Ober, ETS
Kevin M. Williams, ETS
Karyssa A. Courey, Rice University

Competency-based education (CBE) is transforming Kindergarten through grade 12 (K–12) learning by emphasizing mastery of essential competencies through personalized, learner-centered pathways. As educational systems increasingly adopt CBE, we recognize a critical need for assessment practices that are equitable, authentic, and instructionally embedded. In this paper, we aim to advance both theory and practice by presenting a comprehensive framework for designing and implementing competency-based assessments in K–12 settings. We describe nine assessment considerations spanning the assessment lifecycle. At the conceptualization stage, we note how assessment systems can identify essential learner competencies, define and ground competencies conceptually, and promote fairness and equity in what counts as evidence of proficiency. In terms of task development and implementation, we emphasize how assessment systems should design authentic and contextualized tasks, enable flexible demonstrations of competencies, provide actionable feedback to motivate learners, align assessments with instruction, manage time and resource demands, and scale assessment responsibly with technology. Throughout, we leverage an Evidence-Centered Design (ECD; Mislevy et al., 2003) perspective to clarify how conceptualizing competencies (domain model), task design (task model), and assessment implementation and interpretation (evidence model) cohere to support defensible inferences about learners' competencies. Ultimately, our work seeks to inform the assessment practices that align with the broader goals of CBE in K–12 settings and support learners in striving toward their academic and professional goals.

Keywords: competency-based education (CBE); evidence-centered design (ECD); K–12 education; assessment innovations; teacher education; teacher preparation

---

Correspondence concerning this article should be addressed to Teresa Ober, Ph.D., ETS, Princeton, NJ, USA. Email: tober@ets.org
Disclosure Statement: This research is sponsored by ETS, which employed two of the authors at the time of writing. The authors have disclosed interests fully and have in place a plan for managing any potential conflicts arising from this arrangement.

© 2026 the Author(s)
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Competency-based education (CBE) is an approach that focuses on developing key competencies, supporting learner agency, and using assessment to guide instruction in response to each learner's needs (Levine & Patrick, 2019). Models of CBE emerged prominently in professional training and higher education contexts (Gallagher, 2014; Nodine, 2016), and this history provides useful context for the broader evolution of the approach. At the same time, there are several unique challenges to implementing CBE in the context of K–12 systems. In particular, when CBE operates within K–12 settings, state accountability systems (Evans & Landl, 2024), grade-level standards (Polikoff, 2020), and varying levels of teacher preparation (Bennett & Gitomer, 2009) create conditions that fundamentally shape how competencies are defined and assessed, as well as the ways in which evidence from assessments is interpreted.

Growing Interest in K–12 Competency-Based Education

Having gained recent traction in K–12 education, CBE refers to an approach that may enhance learner engagement and promote equitable access to high-quality and individualized learning opportunities (Evans et al., 2020; Sturgis & Patrick, 2010). The concept of CBE is likely to be familiar to many through various initiatives to promote learners' competencies, which may be referred to as 21st century skills (Geisinger, 2016), durable skills (Cole et al., 2021), transferable skills (Tight, 2021), socioemotional skills (Lechner et al., 2019), noncognitive skills (Lipnevich & Roberts, 2012), or other related terms. In the context of the present discussion, we recognize the term *competencies* as referring broadly to a combination of knowledge, skills, and dispositions required when navigating complex situations, adapting to changing environments, and achieving desired outcomes (Tahirsylaj & Sundberg, 2020). Examples of competencies include communication, critical thinking, collaboration, and digital literacy, which are viewed as important for success in educational and workplace settings (National Institutes of Health, n.d.). In the present discussion, we regard *skills* as a type of competency that is learned and often specific to effectively performing a task or activity. For instance, public speaking, problem-solving, project management, and data analysis may reflect task-specific skills that fall under broader competencies such as digital literacy, communication, and leadership. In the present discussion, we focus on competencies as the focus of assessment.

The recent momentum for CBE in K–12 education has been bolstered by its perceived potential to address systemic inequities by promoting personalized learning pathways that allow learners to demonstrate and develop their unique strengths (Sturgis & Casey, 2018). Working towards this aim, many states and school districts are exploring and adopting competency-based frameworks (Atwell & Tucker, 2024; Plate et al., 2026). Several recent factors also appear to be contributing to the growing interest in CBE in K–12 education, including a growing number of initiatives responsive to a perceived demand for durable, work-ready skills (e.g., Cole et al., 2021; O. L. Liu et al., 2025), and the widespread adoption of artificial intelligence (AI), technologies, which could catalyze the feasibility and scalability of approaches to support personalized learning (Arslan et al., 2024).

Centrality of Assessment in an Effective Model of Competency-Based Education

At the same time, educators adopting CBE approaches still need to gather insights about all learners to inform instruction and to ensure educational systems are accountable for learning. Assessments therefore should ideally serve as the foundation for supporting CBE, given they provide opportunities for demonstrating competencies (Henri et al., 2017). However, given the emphasis of CBE in the application of knowledge and skills in real-world contexts, learners should also have opportunities to practice and demonstrate key competencies in authentic settings. Unlike classroom assessments that are typically administered at the end of an instructional cycle, CBE assessments should be designed to be formative, iterative, and closely aligned with instruction and real-world applications (O. L. Liu et al., 2023, 2025). This emphasis on authentic assessment underscores CBE's commitment to equipping K–12 learners with the ability to transfer skills beyond the classroom.

Purpose

As educational systems adopt CBE approaches, there is a need to ensure that assessment practices align with the principles of personalization, authenticity, and equity that underpin this educational model. Our purpose is to identify key challenges in competency-based assessment. We then propose evidence-centered design (ECD; Mislevy et al., 2003) aligned solutions that link competency definitions (domain model) to assessment tasks (task model) as well as to interpretation and feedback (evidence model) to maintain psychometric rigor while supporting learner-centered approaches to teaching. Following ECD principles may help ensure that assessments are explicitly grounded in well-defined theoretical frameworks, and the assessment and the evidence it generates are intentionally aligned with the constructs being assessed. The ECD approach recognizes several interconnected models guiding decisions about how to design and use assessments, including the (1) domain model, which defines the learner's knowledge, skills, and attributes or other characteristics to be assessed, (2) the task model, which outlines the tasks or scenarios that will elicit evidence of these constructs, and (3) the evidence model, which specifies how evidence from tasks will be evaluated to support valid inferences about the learner's abilities. Alignment between these different models helps delineate performance expectations. In other words, whereas the domain model describes which competencies matter in specific contexts, the task model outlines the specific tasks designed to assess certain competencies, and the evidence model focuses on how evidence from tasks is evaluated to draw reliable, valid, and fair conclusions about a learner's competency-based performance. Recognizing that CBE may take place both within formal and informal educational contexts, we prefer to use the term "learner" as opposed to "student" throughout this discussion to refer broadly to those who benefit most directly from CBE learning opportunities.

Having introduced CBE and its growing importance in K–12 education, we next describe current challenges and potential solutions across nine assessment considerations, from conceptualization to implementation. We then explore future directions for research and practice, with particular attention to psychometric properties,

stakeholder collaboration, and technological innovation. This paper aims to integrate multiple perspectives, including sociocultural responsiveness and technological scalability, into a unified set of considerations for designing and implementing competency-based assessments in K–12 settings. It is intended for teacher educators, assessment designers, educational researchers, and policymakers engaged in the development and implementation of CBE in K–12 settings. In addition, the considerations described may be relevant to those facilitating teacher preparation programs, professional development providers, organizations designed to support local educators, and state or district-level decision-makers. While our approach addresses assessment design broadly, we recognize that in many cases teachers are the primary implementers of CBE assessments in K–12 settings. Throughout this paper, we highlight some potential implications for teacher education and professional development, which we recognize as critical for building authentic and equitable competency-based assessments.

Current Challenges and Potential Solutions for Designing Competency-based Assessments

Though CBE holds promise for redefining how learners acquire and demonstrate essential competencies, there are several current challenges related to competency-based assessment. In this section, we discuss a few of these challenges and propose solutions for assessments within the context of CBE that are guided by ECD (Mislevy et al., 2003). Although our discussion is organized across three general stages of the assessment lifecycle (conceptualization, task development, and implementation), the connection between these stages and the ECD models is best understood as approximate rather than one-to-one. Ultimately, the quality of competency-based assessments depends not only on sequencing work across stages but also on sustaining a coherent validity argument across domain, task, and evidence models over time. In particular, decisions about how competencies are framed in the domain model shape what kinds of performances can be meaningfully elicited in the task model and what interpretations are defensible in the evidence model, while constraints and design choices at later stages often require revisiting earlier assumptions about constructs, task features, and interpretation rules. In addition, equity concerns can affect choices related to the definitions of competencies, task design, and rules for evidence interpretation. In addition to providing a summary of the major challenges to assessment in CBE, along with proposed solutions, Table 1 highlights how these nine considerations cut across the domain, task, and evidence ECD models.

Table 1
Considerations, Challenges, and Proposed Solutions for Developing Competency-Based Assessments

Approximate Assessment Stage	Assessment Consideration	Challenge	Most Relevant ECD Model(s)	Proposed ECD-aligned Solution
Conceptualization	Identify Essential Learner Competencies	Competency priorities vary across contexts (e.g., state, community, workforce expectations)	Domain	Specify competencies and proficiency expectations in a domain model that reflects local/community and future-oriented demands
	Define and Ground Competencies Conceptually	Competencies may not be well defined or otherwise difficult to observe consistently	Domain (with alignment to Task and Evidence)	Develop domain-model definitions with observable indicators and explicitly map them to task-model features and evidence-model rules
	Ensure Fairness and Equity	Narrowly defined competencies may lead to bias and inequitable tasks and/or limiting interpretations of performance evidence	Domain + Task + Evidence	Apply socio-culturally responsive design across the ECD chain: domain definitions, accessible task features, and evidence-model interpretation rules
Task Development	Design Authentic, Contextualized Tasks	Tasks are often decontextualized and inauthentic, limiting evidence of real-world performance	Task (with connections to Domain and Evidence)	Design task-model scenarios that elicit evidence through authentic, situated performances aligned to the domain model
	Enable Flexible Demonstrations of Competencies	There may be limited avenues for learners to demonstrate competence through varied artifacts and modalities	Task (with connections to Domain and Evidence)	Build flexible task-model options (multiple modalities/paths) while maintaining domain-model alignment and comparable evidence
	Provide Actionable Feedback to Motivate Learners	Feedback can be discouraging or not diagnostic of next steps	Task + Evidence	Use evidence-model interpretation to provide strengths-based, actionable feedback linked to competency indicators

	Align Assessment with Instruction	Assessment evidence may be disconnected from instructional goals and day-to-day learning	Domain + Task + Evidence	Ensure coherence: domain-model learning targets, task-model embedded opportunities, and evidence-model interpretations that inform teaching
Implementation	Manage Time and Resource Demands	Authentic performance assessments can be time- and resource-intensive to administer and score	Task + Evidence (with alignment to Domain)	Streamline task-model administration and evidence-model scoring/interpretation (e.g., efficient rubrics, process/response data) while preserving domain alignment
	Scale Assessment Responsibly with Technology	Small-scale approaches may yield disconnected sources of evidence and weak longitudinal tracking	Domain (stable), Task + Evidence (at scale)	Leverage technology to scale task-model delivery and strengthen evidence-model interpretation for continuous tracking and timely feedback

Note. ECD = Evidence-centered design principles (Mislevy et al., 2003).

While the primary focus is on assessment for formative and instructional uses, we also acknowledge implications for summative and system-level assessments. In particular, we note that some of the challenges outlined in Table 1 may be more pertinent to large-scale assessment development (e.g., scalability), while others are directly applicable to classroom assessment (e.g., feedback and instructional alignment). In addition, we acknowledge that the very nature of the competency being assessed determines the type of relevant evidence elicited through an assessment, which affects how each of these considerations is applied. Table S1 in the Supplemental Materials provides examples of how these considerations may guide certain decisions for the development and use of assessments related to three example competencies.

Conceptualization

The first set of considerations focuses on defining what counts as a focal competency in a given K–12 context and what evidence meaningfully represents it. In ECD terms, when the domain model does not clearly specify what counts as the competency (and for whom), it can be difficult to interpret task evidence measuring the intended competency. Given that these early decisions shape what evidence assessment tasks can meaningfully elicit and how results can be interpreted, clarity and coherence at the stage of conceptualization are essential for building assessments

that are likely to be instructionally useful and defensible.

Consideration #1: Identify Essential Learner Competencies

The identification of essential competencies is fundamental given that it informs the scope and operationalization of assessment efforts. At the same time, there is great variety among CBE programs in terms of what competencies are emphasized and how subskills are defined, which can complicate decisions around the measurement of complex skills (Evans, 2020). Focal competencies should reflect academic goals while also considering the various professional, civic, or personal contexts in which learners will use these competencies. For example, considering the career relevance of certain competencies helps ensure that K–12 learners are prepared to meet key workforce demands (Cunningham et al., 2016).

Compared to competency frameworks in specific workplace settings, those identified in K–12 competency frameworks often take a less specialized approach to the identification of essential skills, focusing more on a wide range of general competencies that are likely to prepare individuals in the future for a variety of professions (see Atwell & Tucker, 2024). Yet even in K–12 settings, there are differences in the identification of which general competencies are likely to matter most. Some of these differences originate at the state level, where varying competency frameworks emphasize different skills, leading to variation in how they are assessed. For example, the competency framework proposed by the State of North Carolina highlights adaptability, collaboration, communication, critical thinking, empathy, learner's mindset, and personal responsibility (North Carolina Department of Public Instruction, n.d.). In comparison, the competency framework used by the State of Utah highlights academic mastery; wellness; digital literacy; communication; critical thinking and problem solving; creativity and innovation; collaboration and teamwork; civic, financial and economic literacy; honesty, integrity and responsibility; hard work and resilience; lifelong learning and personal growth; service; and respect (Utah State Board of Education, 2020). As can be seen in these two states, there is some overlap in general competencies (e.g., communication, collaboration, critical thinking), but also variation. Furthermore, states currently vary in their commitment to CBE adoption. Such state-specific skills frameworks are often driven by their respective community and local labor market needs (Erwin & Silva-Padrón, 2022). Nonetheless, establishing a common language to reference when developing skills frameworks could help ensure comparability across regions and support better standards for teacher education related to CBE implementation and assessment. Otherwise, competency labels may be shared across districts while being operationalized differently, potentially undermining the interpretability and comparability of assessment evidence.

Technological advancements further complicate the identification of skills that are likely to be relevant in the present and future. The importance of navigating information-rich and technology-mediated environments requires skills to engage in the context of information-rich technology-mediated environments (Sparks et al., 2016). Artificial intelligence (AI) literacy is also emerging as a critical competency

within the workforce, which underscores the need for learners in K-12 settings to develop skills for using AI tools effectively, such as using AI to source information, evaluate AI outputs, communicate using AI, and using AI to personalize content (Chakraburty et al., 2025).

Consideration #2: Define and Ground Competencies Conceptually

Conceptual grounding refers to the process of specifying and justifying what a competency means, including its key components, boundaries, and observable indicators. Subsequently, there is more likely to be clarity for what should be elicited and interpreted as evidence of competencies. When competencies lack conceptual grounding, their definitions may be underspecified. As a result, tasks cannot reliably elicit evidence of the intended competency, and evidence rules may not support defensible interpretations (Kane, 2013). This problem may even be amplified in K–12 because judgments about competencies may be distributed across classrooms, subject areas, and teachers with varying prior training and time to commit to calibration (DeLuca et al., 2016). As such, sound research-based conceptual frameworks are critical for balancing state-specific needs with standardized definitions that may be shared by multiple states. Unfortunately, even after identifying essential competencies, many assessments lack a clear conceptual foundation or are inconsistently operationalized, posing challenges for competency-based assessment (O. L. Liu et al., 2023). Variations in how competencies are defined make it more difficult to develop comprehensive and consistent assessments in practice that adequately measure all behavioral indicators of a given competency. Without clear definitions and observable skill indicators, it becomes difficult to consistently capture the behaviors, cognition, and attitudes that signify proficiency in certain competencies.

With respect to ECD, conceptual grounding helps ensure a well-specified domain model by outlining features of competencies that elicit evidence grounded in observable actions or artifacts that reflect the conceptualization of the theorized competency. Tools like behaviorally anchored rating scales (BARS) and structured observational methods align with this model by linking performance indicators to well-defined competency constructs, thereby presumably reducing subjectivity in how scores are assigned and improving inter-rater reliability (Klieger et al., 2018). When educators know about these principles of assessment design, they may be able to make better choices about when and how to implement competency-based assessments. If these considerations are ignored, scores and feedback may become idiosyncratic, weakening reliability and validity.

Consideration #3: Ensure Fairness and Equity

Educational assessments have been criticized for assuming a uniform mode of performance, typically one that aligns with dominant cultural, linguistic, or social norms, thereby potentially marginalizing learners whose background experiences and knowledge differ from some expected norm (Sireci, 2020). From an ECD perspective, inequity can enter at any point in the assessment process, through domain

definitions that privilege dominant norms, task designs that constrain how competence can be shown, or evidence rules that systemically disadvantage particular groups (Bennett, 2023; Mislevy et al., 2003).

When assessment tasks are decontextualized, they may not fully capture the varied ways in which competencies manifest across different social, regional, and cultural environments. Consequently, learners from underrepresented groups may find themselves at a disadvantage, as their authentic expressions of the competency are either overlooked or misinterpreted. Equity-focused assessment practices are therefore essential for ensuring that all learners, regardless of background, have meaningful opportunities to demonstrate competencies. For example, allowing learners to respond in their home language or through culturally relevant scenarios can enhance both engagement and validity (see also the discussion of varied response options in *Enable Flexible Demonstrations of Competencies*). From an ECD perspective, equity concerns should be addressed across the full chain of the validity argument. For example, the domain model should define the focal competency in culturally and linguistically sustaining ways, the task model should offer accessible opportunities to demonstrate competence, and the evidence model must support interpretations that minimize bias. If equity considerations are ignored during conceptualization or later stages of assessment development, assessments may produce uneven opportunities to demonstrate competencies and thus inequitable interpretations of evidence, threatening arguments related to validity and fairness.

Applying this consideration to competency-based assessments, tasks should be designed to account for social and cultural variation that could affect how skills are demonstrated. For example, socio-culturally responsive assessment approaches (e.g., Bennett, 2023) acknowledge that teaching and assessment should account for the variation in how learners develop and demonstrate their knowledge. According to Bennett (2023), this approach to assessment may be grounded in five provisional principles, which include: (1) present problem situations that connect to, and value, examinee experience, culture, and identity, (2) allow for multiple forms of representation and expression in problem stimuli and in responses, (3) promote instruction for deeper learning through assessment design, (4) adapt the assessment to learner characteristics, and (5) represent assessment results as an interaction among what the examinee brings to the assessment, the types of tasks engaged, and the conditions and context of that engagement. This approach strengthens construct validity by ensuring that skill demonstrations remain authentic across varied cultural and linguistic experiences. Task content may need to reflect unique cultural and social norms of the test-taker, better ensuring that assessments are more inclusive and equitable (Walker et al., 2023). For example, a communication skills assessment that integrates these principles might allow for flexibility in the use of verbal or nonverbal aspects of communication that may vary based on the examinee's lived experience, including their identity, cultural or linguistic background, or other relevant factors. In this way, scoring would not be based on a single highly constrained template of "appropriate" or "correct" communicative expression that might itself be grounded in one dominant sociocultural perspective or context, and may even allow learners to respond in

the language of their choosing. Thus, the assessment should produce more equitable results by permitting learners to express focal competencies, like communication skills, in a manner that best gives students the opportunity to demonstrate these competencies.

Task Development

The second set of considerations addresses how conceptualizations of competencies are translated into assessment tasks and performance opportunities that allow learners to elicit observable evidence of the focal competencies. These considerations reflect an emphasis on designing tasks that are authentic, flexible, and feasible for K–12 settings while still supporting comparable interpretations of evidence across learners and contexts.

Consideration #4: Design Authentic, Contextualized Tasks

Authenticity and contextualization of assessment refer to the extent to which assessment tasks approximate the real situations in which a competency is learned and applied. This may involve embedding opportunities to demonstrate performance of competencies in meaningful scenarios, purposes, and disciplinary contexts so the resulting work reflects how the competency is enacted in practice. When authenticity is treated as optional, the task model can elicit evidence that is weakly connected to real-world competency enactment, limiting the credibility of inferences from performance to competence (Villarroel et al., 2018). This problem may be challenging in K–12 settings because instructional time and classroom logistics constrain what kinds of authentic performances teachers can feasibly support and evaluate. Additionally, certain traditional assessment methods, particularly standardized assessments, may not be well-suited for assessing competencies given that evidence of focal competencies may need to be demonstrated in highly contextualized ways (Evans et al., 2020; Koh, 2017; Stump et al., 2017). Such disconnect is particularly problematic in CBE, where the goal is not merely to verify academic mastery but to ensure that learners can perform in dynamic and authentic contexts. When assessments do not reflect the complex environments in which competencies are often applied, educators may struggle to accurately gauge whether learners have internalized and can transfer their learning beyond the classroom, which may lead to a misrepresentation of a learner's level of competency.

Through an ECD lens, the considerations of authenticity and contextualization in competency-based assessments reflect decisions in specifying the task model because they determine how opportunities to demonstrate proficiency are designed with the goal of eliciting evidence truly aligned with the domain model. Assessments that facilitate the demonstration of a competency in an authentic manner should not only measure what learners know but also how they apply their knowledge in complex, practical (Rømer, 2002; Villarroel et al., 2018), and varied (Brower et al., 2017; Care & Kim, 2018) settings. Several assessment approaches have been developed to address this challenge. For example, scenario-based assessments or situational judgment tests are designed to measure competencies by contextualizing demonstrations of competencies within the context of a specific scenario that may be situated in a

relevant setting, such as a workplace or classroom.

Consideration #5: Enable Flexible Demonstrations of Competencies

Flexibility in capturing unique learners' strengths refers to designing assessments that allow learners to demonstrate the same competency through multiple modes, products, or pathways. In K–12, this issue is amplified because learners' background knowledge, interest, and motivations may vary widely across a variety of dimensions, while at the same time, teachers must make judgments under time constraints that allow for a comparison of performance across learners (DeLuca et al., 2016)._Furthermore, a primary goal of CBE is to cultivate a broad range of multi-dimensional competencies across a variety of contexts, such as communication (Ober et al., 2025), critical thinking (Williams et al., in press), digital literacy (L. Liu et al., 2025), and AI literacy (Chakraburty et al., 2025). This fundamentally challenges some of the conventional requisites of the administration and scoring of standardized assessments (Torres et al., 2015). Thus, competency-based assessments of a group's performance in aggregate (rather than based on an individual) may be restricted to assessing discrete and isolated skills, thereby failing to capture the multifaceted nature of many competencies, particularly given the depth and variability of learners' abilities. If flexibility in the ways learners demonstrate evidence of a competency is not considered, assessments may under-represent proficiency for learners whose strengths do not match a single allowed mode of expression, thus reducing validity and introducing potential sources of inequity.

From an ECD standpoint, this consideration could be viewed as a feature of the task model. By broadening acceptable ways of demonstrating a given competency, it may be possible to ensure the link between the underlying competency defined in the domain model while preserving interpretability in the evidence model. A shift toward personalization in assessments reflects this line of thinking, and more recent advances in psychometrics may help accommodate variability in content and format while maintaining consistency in measuring core competencies (Arslan et al., 2024; Zapata-Rivera & Arslan, 2021). Furthermore, personalization in assessment design may encourage learners to advance competency development in a manner aligned with their personal goals and learning trajectories (Alt et al., 2023). For example, instead of a single assessment that requires all learners to produce a specific outcome, a performance or portfolio-based assessment could allow learners to choose whether to write a short story, build a sculpture, create a song, or any other number of possibilities, thus demonstrating competencies in a way most likely to enable them to display their proficiency. However, designing also flexibility in assessments requires attention to evidential comparability and equity implications (see *Fairness and Equity*), given that the lack of consistency across assessments could also introduce new sources of bias.

Consideration #6: Provide Actionable Feedback to Motivate Learners

Within the context of CBE, individualized feedback is often viewed as a critical tool for guiding learner growth along personalized learning pathways (Alt et al.,

2023; Marion et al., 2020). Assessments designed and used in ways that do not provide encouraging and constructive feedback to learners may render scores detached from meaning and subsequent learning opportunities. These assessments may further reduce a learner's motivation for completing future assessments and ultimately weaken the potential benefits of competency-based approaches.

Feedback that is anchored in the ECD evidence model should provide insights into the learner's competency development that can be interpreted in ways that support next instructional steps and learner growth. The incorporation of frequent, meaningful feedback is critical; such feedback creates a continuous loop that helps learners understand their progress, pinpoint areas for improvement, and build confidence in their abilities, an approach that is essential for fostering self-directed learning (Patrick & Sturgis, 2017). Frequent and constructive feedback may be a key component of an assessment system that adopts a strengths-based perspective by shifting the focus away from identifying skill deficits and toward unique abilities and the potential improvements of each learner (Sireci et al., 2024). Furthermore, by identifying and building on learners' strengths, assessments may better foster interest, self-efficacy, and ultimately encourage learners to further their skill development (Bozic et al., 2018).

At the same time, we recognize that in K–12 settings, translating evidence of learners' competencies into actionable next steps across all learners is time consuming and may require dedicated training in formative assessment practices (Heritage, 2007). Supporting materials such as rubrics that recognize how evidence of a competency may be demonstrated across a developmental continuum could assist educators in providing personalized instruction to learners that target specific behavioral aspects of a competency. Pre-service training and ongoing professional development could also help prepare teachers to provide personalized, formative feedback. For example, prior research involving pre-service teachers found that many benefited from the use of rubrics to assess their own work (Lovorn & Rezaei, 2011).

Implementation

The third set of considerations highlights what it takes to use competency-based assessments effectively in real educational settings, including alignment to instruction, feedback processes, resource demands, and scalability. Decisions related to the implementation and use of assessments determine whether evidence is actionable for teaching and learning and whether systems can sustain equitable, high-quality competency-based assessment over extended periods of time.

Consideration #7: Align Assessment with Instruction

Connections to instruction refer to the extent to which assessment purposes, tasks, and interpretations are deliberately designed to inform day-to-day teaching and learning. In other words, it involves considering how information from assessments about a learner's competencies could be used to guide instructional decisions. Perspectives on the usefulness of assessment in general have been found to vary among scholars and practitioners (e.g., Bonner, 2016) and tend to shift over time as

accountability demands, instructional priorities, and data infrastructures in educational settings have evolved (Ho, 2014). The potential to use assessment information to inform instructional decision-making appears to be a central factor shaping teachers' perceived usefulness of assessments (Varier et al., 2024). This emphasis is especially salient for formative, classroom-based assessments, which are explicitly designed to support ongoing instructional adjustment, feedback to learners, and learning, rather than high-stakes decision-making or accountability (Bennett, 2011; Shepard, 2017).

Viewed through the lens of ECD, the potential for assessment information to inform instruction may reflect coherence between the domain model, which specifies instructional targets, the task model, which identifies opportunities to elicit evidence, and the evidence model, which supports interpretations that inform decisions (Mislevy et al., 2003). In K–12 settings, this issue is further complicated by differences in pacing of curriculum, accountability pressures, and varied teacher assessment preparation, which can make it difficult to embed assessment evidence into routine instructional decision-making (DeLuca et al., 2016). Disconnects among the three ECD models may result in valuable learner performance insights failing to guide instructional decisions. When assessments are not developed with clear progressions that align with the learning objectives, they may fail to capture the dynamic nature of skill development. Consequently, teachers may be left with information that has limited potential to support the continuous, formative feedback necessary for fostering learners' competency growth.

Applying ECD principles may offer some guidance toward improving connections between competency-based assessment and instruction. Specifically, ECD strengthens instructional connections by making the intended competencies explicit in the domain model, translating those competencies into instructionally embedded opportunities to elicit performance in the task model, and specifying an evidence model that interprets work created by learners in ways that are actionable for feedback and next-step instruction. Such instructionally embedded approaches that integrate opportunities for learners to develop skills while learning core content show promise in promoting deeper learning and the transfer of knowledge to novel situations (Pellegrino, 2017). Such assessments could not only support instructional decision-making but also foster deeper learning by aligning with curricular goals and enabling teachers to adapt instruction based on real-time learner data. Designed in this way, competency-based assessments could not only support instructional decision-making but also foster deeper learning by aligning with curricular goals and enabling teachers to adapt instruction based on real-time learner data.

With this in mind, competency-based assessments should ideally be instructionally embedded, support instructional decision-making, and foster deeper learning by aligning with curricular goals and enabling teachers to adapt instruction based on real-time learner data (Briggs & Furtak, 2019). Such assessments may be developed with explicit progressions, helping educators document learners' advancement from one level of skill proficiency to another, in a manner similar to a learning progression (Mosher, 2022; Shepard, 2018). Some have recommend integrating skills across the

curriculum (e.g., Voogt & Roblin, 2010), with the expectation that opportunities to develop skills such as communication and collaboration can enhance learning in various disciplines. Some past research has explored the potential for integrating skills into core academic subject areas (e.g., Cloonan, 2015; Gut, 2010), including science (Turiman et al., 2012), including mathematics (English & Gainsburg, 2015), language arts (Caughlan et al., 2017), and social studies (Damico & Panos, 2018; Gallavan & Kottler, 2012). Alternatively, courses can be structured around interdisciplinary topics or skills (e.g., sustainability, critical thinking) that overlap multiple subject areas, further supporting learners' competency development (Drake & Reid, 2020).

Consideration #8: Manage Time and Resource Demands

When implementing an assessment, practitioners often need to make choices related to the perceived resource and time intensity of administering it. This may include the practical burden of designing, administering, and scoring competency-based assessments, such as the requirements for instructional time, educator expertise and training, materials or technology required, efforts needed to score the assessment, and so forth. Competency-based assessments, particularly those that involve performance-based approaches aligned with seemingly authentic demonstrations of competencies, are often recognized as resource- and time-intensive (Chen et al., 2024; Fawns et al., 2024). Such assessments have been found to take longer to administer than more standardized test formats (Frey et al., 2012). In addition, given the learner-centered emphasis of CBE, competency-based assessments may require highly individualized settings or formats for each learner, which can potentially create logistical challenges in typical classroom environments where space and resources are often limited (Torres et al., 2015). Scoring of such individualized assessments can also be resource- and time-intensive, in part because teachers need training and calibration sessions to ensure reliability across different evaluators (Darling-Hammond & Adamson, 2010). Furthermore, the use of assessments that capture multifaceted competencies demands ongoing training and support for educators to ensure that they are properly aligned within instruction (Suárez-Álvarez et al., 2024). In K–12 settings, teachers' time dedicated to each of these assessment-related activities competes with other demands. Additionally, staffing/resources are likely to vary substantially across schools (Darling-Hammond & Adamson, 2010). When these multiple demands are ignored, educators may not be sufficiently prepared to administer and use competency-based assessments with full construct coverage. This may lead to less consistent scoring and lower confidence in the resulting interpretation of evidence from assessments.

From an ECD perspective, resource and time constraints primarily place pressure on the task model (what can be feasibly administered in classrooms?) and the evidence model (what can be scored and interpreted reliably?), which can weaken validity if they force shortcuts. Streamlining assessment task design to focus on key aspects of the competency and essential evidence processes without drifting from the competency definitions and performance expectations specified in the domain model

could help reduce the tension between constraints and capturing the construct as fully and meaningfully as possible. In this direction, one promising strategy toward alternative assessment types, such as "testless" assessments and the use of process data, may streamline the evaluation process while still providing meaningful insights into learners' performance (Kovanovic et al., 2023; Rahimi et al., 2024). By integrating these assessments directly into the daily teaching and learning cycles, educators can reduce the extensive preparation and administration time traditionally associated with standardized tests. Furthermore, aligning more innovative assessment methods with ongoing professional development ensures that teachers are well-equipped to interpret the data and use it to inform instruction, ultimately creating a more efficient, effective, and sustainable competency-based education system.

Consideration #9: Scale Assessment Responsibly with Technology

In looking to the future of assessment, it appears undeniable that technologies, especially AI systems, will increasingly have a role in the conceptualization of competencies, the design of assessment tasks, and the implementation of competency-based assessments (Kyllonen et al., 2024). Considerations related to the scalability and appropriate use of technology should involve evaluating the extent to which competency-based assessment systems can deliver comparable tasks, capture and aggregate evidence, and support timely, defensible interpretations of competency growth across classrooms, grades, and schools. With this in mind, AI systems or other digital technologies should be used only when they add value without outpacing considerations related to validity, equity, privacy, and instructional feasibility, among other factors.

Though many competency-based assessments have historically tended to be limited to use for small-scale, classroom-based purposes, this appears to be rapidly changing (Marion et al., 2020). Such a localized approach often results in fragmented evidence of learner competencies, making it difficult to track competency development across contexts and time. This challenge is particularly pertinent in K–12 settings because assessment information is expected to travel across grades, schools, and accountability contexts. Yet, data infrastructures and instructional constraints vary widely, and as a result, it may be challenging to leverage assessment results to inform systemic improvements in CBE to benefit learners.

From the perspective of ECD, the potential for technology-enabled scalability of competency-based assessments could be viewed primarily as issues concerning the task- and evidence-model. In particular, technology-enabled systems must deliver comparable task opportunities and support consistent scoring and interpretation across settings, while keeping the underlying competency claims in the domain model stable. Appropriately used technologies also need to support this coherence by standardizing aspects of the task delivery and evidence capture, enabling more defensible longitudinal interpretations of competency growth. Without such scaling, evidence of learners' competencies gathered from various assessments could seem less coherent, limiting interpretability. Defensible interpretation of evidence from scalable technology-enabled assessment tasks also needs to maintain connections to

the competency definitions specified within the domain model.

Potential solutions to these challenges lie in the use of digital information-management platforms to provide more structured, adaptive systems for teachers and learners to store, organize, aggregate, and curate multiple forms of evidence of learners' competencies (Timmis et al., 2016). Various applications of digital technologies can be used to support interactive and adaptive assessments that generate richer evidence of competency development. Digital games or simulations (Shute & Rahimi, 2017), as well as portfolio or performance-based assessments (Henri et al., 2017), could also be used to capture the application of skills as they are applied in naturalistic yet comparable task environments. Furthermore, AI-driven analytics (Swiecki et al., 2022) and conversation-based assessments (Zapata-Rivera et al., 2024) not only enable real-time feedback but also capture nuanced data on learner performance over time. The use of such tools may better facilitate continuous tracking and record keeping, while also improving the precision of assessment outcomes. At the same time, we emphasize that claims about the benefits of technology-enabled assessments must be evaluated through empirical studies of validity, reliability, fairness, and impact. While seemingly promising, many technology-enabled assessment systems, particularly those that involve generative AI, are still in early stages of development. Further research is needed to understand their applications and long-term impact.

Future Directions

The nine considerations outlined highlight how competency-based assessment in K–12 settings can be designed as a coherent system. Across *conceptualization*, they emphasize clarifying which competencies matter, grounding them in observable indicators, and embedding fairness and equity into what counts as evidence (domain model). In the process of assessment *task development*, considerations stress creating authentic, contextualized opportunities for learners to demonstrate competence in multiple ways and ensuring evidence can be translated into actionable feedback (task and evidence models). Across *implementation*, considerations underscore the need to align assessment with instruction, manage time and resource demands, and scale responsibly with technology while preserving defensible interpretations of competency growth (domain–task–evidence coherence). Building from this foundation, we now turn to future directions that can strengthen research and practice as systems expand competency-based assessment efforts.

Teacher Preparation and Professional Development

Successful implementations of competency-based assessments hinge on adequately preparing teachers. Given the unconventional nature of CBE, many educators require additional training and support to adopt effective assessment practices (Echols et al., 2018). Teachers may need to rethink classroom instruction and develop assessment literacy to implement competency-based assessments effectively (DeLuca et al., 2016). This shift requires moving beyond traditional testing paradigms toward assessment practices that are deeply integrated with instruction and focused on

supporting learning, a fundamental reorientation of assessment's role in educational practice (Shepard, 2000). As evidence of this, a survey of 413 public school principals in the Northeast revealed greater success in implementing CBE practices, such as building professional learning communities, facilitating discussions about learners' work, and discussing strategies to support learners, compared to self-paced or flexible assessment practices (Evans et al., 2019). These findings suggest that CBE practices closely aligned with current assessment methods may be easier to adopt and integrate into existing systems. In particular, state and district policies play a pivotal role in shaping the feasibility of competency-based assessment systems. To further support teachers, creating statewide task forces, updating licensure to emphasize mastery, incorporating micro-credentialing, and providing professional development focused on assessment literacy have also been recommended (Patrick et al., 2020).

To prepare educators to administer competency-based assessments and evaluate learner artifacts consistently and fairly, high-quality teacher training is critical (Darling-Hammond et al., 2017, ch. 4). Teachers need to develop foundational assessment literacy skills to connect assessments to clear purposes, choose appropriate methods for assessment, mitigate bias, and communicate results back to learners (Atjonen et al., 2022; DeLuca et al., 2016). In the context of assessment literacy for measuring learners' competencies, such skills may include learning how to use multiple assessment methods (e.g., performance-based tasks, portfolios, etc.), which go beyond traditional testing to measure learners' real-world application of competencies, technology integration, and synthesis and analysis of data from multiple sources, as well as interpretation of learner performance and decision-making.

Educators may benefit from additional training and support to adopt effective competency-based assessment practices, particularly given that such assessments may represent a departure from traditional approaches for classroom assessment (Echols et al., 2018). For example, in the context of teacher preparation programs, teachers could benefit from receiving guidance in developing and using behaviorally anchored rating scales to standardize evaluations, reduce subjectivity in ratings, and maintain authenticity of assessments (Klieger et al., 2018). Furthermore, teacher education programs may need to align training for competency-based assessment literacy with national and state standards (e.g., Council for the Accreditation of Educator Preparation [CAEP, 2022]; Interstate Teacher Assessment and Support Consortium [InTASC, Council of Chief State School Officers, 2013], etc.), while also creating supportive structures such as statewide task forces, updating licensure requirements to emphasize assessment mastery, and providing professional development focused on assessment literacy (Patrick et al., 2020). Building professional learning communities, where teachers engage in collaborative discussions about learner work and strategies to support learners, may be particularly effective in supporting teachers' implementation of competency-based assessment practices (Darling-Hammond, 2016). Ultimately, providing teachers with such support may better ensure that competency-based assessments can be used to facilitate meaningful feedback for learners.

Sustainable Stakeholder Collaboration

As competency-based education and associated assessment practices move beyond classroom use and become more integrated across various levels of educational systems (i.e., schools, districts, states), questions of governance, shared meaning, and decision-rights become increasingly central. Competency-based assessment systems should be designed to balance local relevance with comparability and validity, which requires effective stakeholder collaboration. In K–12 contexts, collaboration among stakeholder groups should help advance deeper reflection on the considerations described previously and leverage each other's strengths and expertise to specify the entire ECD argument chain from defining competence, to describing tasks and interpreting performance. For example, working together, educators help ground the domain model in instructionally meaningful practices and ensure the task model is feasible and authentic in classrooms; policymakers clarify the boundary conditions that shape domain and task choices (e.g., comparability, accountability, resources) and constrain how evidence can be used; and researchers and assessment developers strengthen the evidence model by building defensible scoring, interpretation, and validation arguments. In combination, these partnerships help maintain coherence across the three models and support continuous improvement by using implementation evidence to refine competency definitions (domain model), task features (task model), and interpretation rules (evidence model) over time.

Clarifying these types of responsibilities may help avoid implementation scenarios in which local adaptations accumulate without a common and logical evidentiary argument, a risk that can exacerbate inequities when policy interpretations and systems of accountability vary across districts (Ayscue et al., 2022). Furthermore, stakeholder collaboration supports not only adoption but also the ongoing refinement of competency-based assessments by creating environments conducive to iteration via feedback spirals, through which evidence is interpreted, acted upon in instruction, and then subsequently revisited (Carless, 2019), all while progressively strengthening the alignment among domain, task, and evidence models over time. As a set of design principles (rather than implementation steps), collaboration can be operationalized through a small number of recurring mechanisms that keep ECD models synchronized across levels of the system. First, co-design activities that engage multiple constituents (e.g., educators, assessment or learning task developers, and researchers) could be used to jointly articulate the domain model (shared competency definitions and proficiency expectations) and map those claims to task model features that are feasible in classrooms yet still aligned to the intended construct (Basu et al., 2020; Mislevy & Riconscente, 2005). Second, routine calibration (or norming) conversations around shared rubrics and exemplars can support more consistent interpretation of evidence, an important condition for reliability and fairness when performance judgments are distributed across teachers and schools (Darling-Hammond & Adamson, 2010). Third, pilot-and-revision cycles can preserve local flexibility while protecting comparability by documenting how local task adaptations are linked to a common evidence model (i.e., how evidence is scored, aggregated, and

used), allowing systems to learn from variation without losing the validity argument (Herman & Linn, 2015; Kane, 2013).

These mechanisms are most sustainable when embedded in existing educator learning and governance structures (e.g., professional learning communities and state/district working groups), so that interpretive consistency and task quality improve over time rather than relying on one-time training (Darling-Hammond, 2016; Patrick et al., 2020). Framed this way, stakeholder collaboration supports the adoption and ongoing refinement of competency-based assessments, thus supporting K–12 systems in aligning local instructional constraints with policy demands for equity, transparency, and responsible accountability use (Ayscue et al., 2022; Sullivan & Downey, 2015).

Research Priorities to Inform Competency-based Assessment

Beyond factors concerning the practical implementation of CBE and associated assessment systems, there is a pressing need for deep and thoughtful research on competency-based assessments. One critical area of study involves exploring the unique considerations of applying psychometric properties to CBE assessments (see Table S2 in the Supplemental Materials). Fundamental characteristics such as reliability, validity, and fairness are crucial elements of psychometrically sound assessments, yet at the same time, features of CBE approaches, such as the emphasis on personalization, may require researchers to rethink conventional notions of these psychometric properties. For example, when assessments are personalized to minimize bias and leverage individual learners' strengths and sociocultural backgrounds, individual learners may complete different items and different versions of the assessment, complicating psychometric evaluation (Bennett, 2024). Similarly, assessments that are contextualized to different content areas (e.g., math, social studies, chemistry), but are designed to measure the same competency, should ideally be evaluated along comparable psychometric standards. Research involving test equating (Kolen & Brennan, 2013), adaptive tests (Chang, 2025), process data (Anghel et al., 2024), constructed responses (e.g., machine-learning scoring; Kaldaras et al., 2022), differential item functioning (DIF; Penfield & Camilli, 2006), measurement invariance (Schmitt & Kuljanin, 2008), and other areas may inform methods for investigating various psychometric properties of innovative CBE assessments. Finally, an ECD approach may support the evaluation of psychometric properties, through creating well-defined measurement models, detailing frameworks for operationalizing constructs, and supporting inclusive design processes that engage educators, community interest holders, and subject matter experts (see Table S2).

Special attention should be given to researching the feasibility and impact of scaling these approaches to ensure their applicability in resource-constrained environments, both in the near and long-term. Research on competency-based assessments should identify the long-term impacts of various assessment models on learner outcomes, including how formative, personalized, and technology-enhanced assessments contribute to competency development and readiness for postsecondary opportunities and the workforce. Other ethical issues should also be considered, such

as data privacy and equitable use within schools, to ensure the benefit to all learners. Ultimately, research on these topics can inform a theory of change concerning the implementation of CBE and its intended impact on learning, with assessments as key for measuring the intended impacts.

With regard to research on the design of assessments, technology is likely to play an important role in shaping the future of competency-based assessments, but its contributions should be treated as empirical questions rather than assumptions. Research should explore technology-enabled innovations in digital assessment, especially those that create more interactive affordances that allow for a balance between standardization, personalization, and authenticity, and should evaluate these tools with respect to validity, fairness, and instructional consequences (Kane, 2013; Swiecki et al., 2022; Timmis et al., 2016). Continued innovation in AI and digital platforms offers the potential to create dynamic, adaptive, and scalable assessment tools, including tools that provide more timely feedback and support longitudinal competency records; however, such claims depend on careful validation and responsible use in K–12 contexts (Arslan et al., 2024).

As one potential technology-supported format for recording learners' competencies, micro-credentials have been proposed as a structured approach to documenting competency attainment, offering the promise of providing learners with portable records to advance in future educational and professional pursuits (Moore et al., 2025). At the same time, there is a need for ongoing research to explore how assessments can be used to integrate micro-credentialing systems within instruction, existing educational policies, and data infrastructures (Ahsan et al., 2023; Varadarajan et al., 2023), particularly given that the implementation in K–12 settings is relatively recent (Berry & Byrd, 2019). Future research should examine how competency-based assessments can support such systems intended to support immediate and long-term needs of learners and educators, aligning K–12 learning with workforce and postsecondary expectations.

Conclusions

The adoption of CBE in K–12 settings requires a structured and rigorous approach to assessment that is coherent across the full assessment lifecycle, from the conceptualization of the competency to assessment task development and implementation. We describe nine assessment considerations that shape how competency-based assessments could produce evidence that is valid, equitable, instructionally actionable, and scalable.

Identifying essential competencies and features (*Consideration #1*), ensuring the description of the competency is grounded in observable behaviors (*Consideration #2*), and supporting fair and equitable opportunities to demonstrate the competency (*Consideration #3*) are all essential to support claims related to the domain model from an ECD perspective. During assessment task development, understanding the authentic and contextualized ways learners can demonstrate a competency on the task (*Consideration #4*), flexibly capturing learners' unique strengths (*Consideration #5*), and considering how feedback affects learner motivation (*Consideration #6*)

may determine whether tasks elicit meaningful evidence to support learning-oriented interpretations and subsequent learning. Finally, while implementing and using the task, connections between the task and instruction (*Consideration #7*), the level of resources and time required to implement and use the task (*Consideration #8*), as well as the potential for scalability and responsible uses of technology (*Consideration #9*) all influence whether competency-based assessments are likely to be sustained across classrooms and districts without comprising interpretability, comparability, or relevance to the learner and community.

These nine considerations each underscore the importance of stakeholder collaboration to better ensure competency-based assessments have a strong validity argument by way of ECD-informed coherence across domain, task, and evidence models. Ultimately, this will help ensure competency-based assessments fulfill CBE's broader promise of supporting learners' academic, civic, and professional readiness.

References

Ahsan, K., Akbar, S., Kam, B., & Abdulrahman, M. D. A. (2023). Implementation of micro-credentials in higher education: A systematic literature review. *Education and Information Technologies*, *28*(10), 13505-13540. https://doi.org/10.1007/s10639-023-11739-z

Alt, D., Naamati-Schneider, L., & Weishut, D. J. (2023). Competency-based learning and formative assessment feedback as precursors of college students' soft skills acquisition. *Studies in Higher Education*, *48*(12), 1901–1917. https://doi.org/10.1080/03075079.2023.2217203

Anghel, E., Khorramdel, L., & von Davier, M. (2024). The use of process data in large-scale assessments: a literature review. *Large-scale Assessments in Education*, *12*(1), 13. https://doi.org/10.1186/s40536-024-00202-1

Arslan, B., Lehman, B., Tenison, C., Sparks, J. R., López, A. A., Gu, L., & Zapata-Rivera, D. (2024). Opportunities and challenges of using generative AI to personalize educational assessment. *Frontiers in Artificial Intelligence*, *7*, 1460651. https://doi.org/10.3389/frai.2024.1460651

Atjonen, P., Pöntinen, S., Kontkanen, S., & Ruotsalainen, P. (2022, July). In enhancing preservice teachers' assessment literacy: Focus on knowledge base, conceptions of assessment, and teacher learning. *Frontiers in Education, 7*, 891391. https://doi.org/10.3389/feduc.2022.891391

Atwell, M. N., & Tucker, A. (2024). Portraits of a Graduate: Strengthening Career and College Readiness through Social and Emotional Skill Development. *Collaborative for Academic, Social, and Emotional Learning*. https://files.eric.ed.gov/fulltext/ED641286.pdf

Ayscue, J. B., D. Fusarelli, L., & Uzzell, E. M. (2022). Equity and early implementation of the every student succeeds act in state-designed plans during COVID. *Educational Policy*, *37*(7), 1917-1949. https://doi.org/10.1177/08959048221130994

Basu, S., Disalvo, B., Rutstein, D., Xu, Y., Roschelle, J., & Holbert, N. (2020). The role of evidence centered design and participatory design in a playful assessment for computational thinking about data. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)* (pp. 985–991). Association for Computing Machinery. https://doi.org/10.1145/3328778.3366881

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25. https://doi.org/10.1080/0969594X.2010.513678

Bennett, R. E. (2023). Toward a theory of socioculturally responsive assessment. *Educational Assessment*, *28*(2), 83–104. https://doi.org/10.1080/10627197.2023.2202312

Bennett, R. E. (2024). Personalizing Assessment: Dream or Nightmare?. *Educational Measurement: Issues and Practice*, *43*(4), 119-125. https://doi.org/10.1111/emip.12652

Bennett, R. E., & Gitomer, D. H. (2009). Transforming K–12 assessment: Integrating accountability testing, formative assessment and professional support. In *Educational assessment in the 21st century: Connecting theory and practice* (pp. 43-61). Springer. https://doi-org/10.1007/978-1-4020-9964-9_3

Berry, B., & Byrd, P. (2019). Micro-credentials and education policy in the United States: Recognizing learning and leadership for our nation's teachers. *Digital Promise*. https://files.eric.ed.gov/fulltext/ED614390.pdf

Bonner, S. M. (2016). Teachers' perceptions about assessment: Competing narratives. G. T. L. Brown & L. R Harris (Eds.), *Handbook of human and social conditions in assessment*, 21-39. https://doi.org/10.4324/9781315749136

Bozic, N., Lawthom, R., & Murray, J. (2018). Exploring the context of strengths–a new approach to strength-based assessment. *Educational Psychology in Practice*, *34*(1), 26–40. https://doi.org/10.1080/02667363.2017.1367917

Briggs, D. C., & Furtak, E. M. (2019). Learning progressions and embedded assessment. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 146–169). Routledge.

Brower, A. M., Humphreys, D., Karoff, R., & Kallio, S. (2017). Designing quality into direct-assessment competency-based education. *The Journal of Competency-Based Education*, *2*(2), e01043. https://doi.org/10.1002/cbe2.1043

Care, E., & Kim, H. (2018). Assessment of twenty-first century skills: The issue of authenticity. *Assessment and teaching of 21st century skills: Research and applications*, 21–39. https://doi.org/10.1007/978-3-319-65368-6_2

Carless, D. (2019). Feedback loops and the longer-term: towards feedback spirals. *Assessment & Evaluation in Higher Education*, *44*(5), 705–714. https://doi.org/10.1080/02602938.2018.1531108

Caughlan, S., Pasternak, D. L., Hallman, H. L., Renzi, L., Rush, L. S., & Frisby, M. (2017). How English language arts teachers are prepared for twenty-first-century classrooms: Results of a national study. *English Education*, *49*(3), 265–297. https://doi.org/10.58680/ee201729043

Chen, A. M. H., Kleppinger, E. L., Churchwell, M. D., & Rhoney, D. H. (2024). Examining competency-based education through the lens of implementation science: A scoping review. *American Journal of Pharmaceutical Education, 88*(2), 100633. https://doi.org/10.1080/02602938.2024.2404634

Chakraburty, S., Ober, T. M., & Liu, L. (2025). Preparing K–12 students with AI literacy: Proposed framework, progression, and task design principles (*Research Report No. RR-25-14*). ETS. https://doi.org/10.64634/46jn1p41

Chang, H.-H. (2025). Psychometrics behind computerized adaptive testing. *Psychometrika, 80*(1), 1–20. https://doi.org/10.1007/s11336-014-9401-5

Cloonan, A. (2015). *Integrating by Design: Multimodality, 21st Century Skills and Subject Area Knowledge*. In Cope, B., Kalantzis, M. (eds) *A Pedagogy of Multiliteracies*. Palgrave Macmillan London. https://doi.org/10.1057/9781137539724_5

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

Council for the Accreditation of Educator Preparation (CAEP). (2022). *CAEP standards for initial preparation.* Council for the Accreditation of Educator Preparation. http://secure.caepnet.org/standards/2022-itp/introduction

Council of Chief State School Officers (CCSSO). (2013). *Interstate Teacher Assessment and Support Consortium InTASC model core teaching standards and learning progressions for teachers 1.0: A resource for ongoing teacher development*. Council of Chief State School Officers. https://learning.ccsso.org/educator-standards

Cunningham, J., Key, E., & Capron, R. (2016). An evaluation of competency-based education programs: A study of the development process of competency-Based programs. *The Journal of Competency-Based Education*, *1*(3), 130–139. https://doi.org/10.1002/cbe2.1025

Damico, J. S., & Panos, A. (2018). Civic media literacy as 21st century source work: Future social studies teachers examine web sources about climate change. *The Journal of Social Studies Research*, *42*(4), 345–359. https://doi.org/10.1016/j.jssr.2017.10.001

Darling-Hammond, L. (2016). Research on teaching and teacher education and its influences on policy and practice. *Educational Researcher*, *45*(2), 83–91. https://globaled.gse.harvard.edu/publications/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards

Darling-Hammond, L., Adamson, F., & Abedi, J. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning* (p. 52). Stanford Center for Opportunity Policy in Education. https://globaled.gse.harvard.edu/files/geii/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-report_0.pdf

Darling-Hammond, L., Burns, D., Campbell, C., Goodwin, A. L., Hammerness, K., Low, E. L., ... & Zeichner, K. (2017). *Empowered educators: How high-performing systems shape teaching quality around the world*. John Wiley & Sons.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*, 251–272. https://doi.org/10.1007/s11092-015-9233-6

Drake, S. M., & Reid, J. L. (2020). 21st century competencies in light of the history of integrated curriculum. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.00122

Echols, D. G., Neely, P. W., & Dusick, D. (2018). Understanding faculty training in competency-based curriculum development. *The Journal of Competency-Based Education*, *3*(2), e01162. https://doi.org/10.1002/cbe2.1162

English, L. D., & Gainsburg, J. (2015). Problem solving in a 21st-century mathematics curriculum. In L. D. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (pp. 313–335). Routledge.

Erwin, B. & Silva-Padrón, G. (2022, July). State policies to support student-centered learning. *Education on Commission of the States.* https://www.ecs.org/wp-content/uploads/State-Policies-to-Support-Student-Centered-Learning.pdf

Evans, C. (2020). Measuring student success skills: A review of the literature on critical thinking. 21st century success skills. *National Center for the Improvement of Educational Assessment*. https://eric.ed.gov/?id=ED607780

Evans, C. M., Graham, S. E., & Lefebvre, M. L. (2019). Exploring K–12 competency-based education implementation in the Northeast States. *NASSP Bulletin*, *103*(4), 300–329. https://doi.org/10.1177/0192636519877456

Evans, C., & Landl, E. (2024). State practices and balanced assessment systems. In S. F. Marion, J. W. Pellegrino, & A. I. Berman (Eds.), *Reimagining balanced assessment systems* (pp. 200–236). National Academy of Education. https://files.eric.ed.gov/fulltext/ED661489.pdf#page=209

Evans, C. M., Landl, E., & Thompson, J. (2020). Making sense of K-12 competency-based education: A systematic literature review of implementation and outcomes research from 2000 to 2019. *The Journal of Competency-Based Education*, *5*(4), e01228. https://doi.org/10.1002/cbe2.1228

Fawns, T., Bearman, M., Dawson, P., Nieminen, J. H., Ashford-Rowe, K., Willey, K., Jensen, L. X., Damşa, C., & Press, N. (2025). Authentic assessment: from panacea to criticality. *Assessment & Evaluation in Higher Education*, *50*(3), 396-408. https://doi.org/10.1080/02602938.2024.2404634

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical Assessment, Research & Evaluation, 17*(2), 1–18. https://doi.org/10.7275/sxbs-0829

Gallagher, C. W. (2014). Disrupting the game-changer: Remembering the history of competency-based education. *Change: The Magazine of Higher Learning*, *46*(6), 16–23. https://doi.org/10.1080/00091383.2014.969177

Gallavan, N. P., & Kottler, E. (2012). Advancing social studies learning for the 21st century with divergent thinking. *The Social Studies*, *103*(4), 165–170. https://doi.org/10.1080/00377996.2011.605641

Geisinger, K. F. (2016). 21st century skills: What are they and how do we assess them?. *Applied Measurement in Education*, *29*(4), 245–249. https://doi.org/10.1080/08957347.2016.1209207

Gut, D. M. (2010). Integrating 21st century skills into the curriculum. In G. Wan & D. M. Gut (Eds.), *Bringing schools into the 21st century* (pp. 137-157). Springer.

Henri, M., Johnson, M. D., & Nepal, B. (2017). A review of competency-based learning: Tools, assessments, and recommendations. *Journal of Engineering Education*, *106*(4), 607–638. https://doi.org/10.1002/jee.20180

Herman, J. L., & Linn, R. (2015). *Evidence-centered design: A summary*. Center for Standards, Assessment, and Accountability (CSAA), WestEd. https://csaa.wested.org/wp-content/uploads/2019/11/ECDsummary.pdf

Heritage, M. (2007). Formative assessment: What do teachers need to know and do?. *Phi Delta Kappan*, *89*(2), 140-145. https://doi.org/10.1177/003172170708900210

Ho, A. D. (2014). Variety and drift in the functions and purposes of assessment in K–12 education. *Teachers College Record*, *116*(11), 1-18. https://doi.org/10.1177/016146811411601116

Kaldaras, L., Yoshida, N. R., & Haudek, K. C. (2022, November). Rubric development for AI-enabled scoring of three-dimensional constructed-response assessment aligned to NGSS learning progression. In *Frontiers in education* (Vol. 7, p. 983055). https://doi.org/10.3389/feduc.2022.983055

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Klieger, D. M., Kell, H. J., Rikoon, S., Burkander, K. N., Bochenek, J. L., & Shore, J. R. (2018). Development of the behaviorally anchored rating scales for the skills demonstration and progression guide. *ETS Research Report Series*, *2018*(1), 1–36. https://doi.org/10.1002/ets2.12210

Koh, K. H. (2017). *Authentic assessment*. In K. Hytten (Ed.), Oxford research encyclopedia of education. [Online] https://doi.org/10.1093/acrefore/9780190264093.013.22

Kolen, M. J., & Brennan, R. L. (2013). *Test equating: Methods and practices.* Springer Science & Business Media.

Kovanovic, V., Azevedo, R., Gibson, D. C., & Ifenthaler, D. (2023). Unobtrusive observations of learning processes. In *Unobtrusive Observations of Learning in Digital Environments: Examining Behavior, Cognition, Emotion, Metacognition and Social Processes Using Learning Analytics* (pp. 3–6). Springer International Publishing.

Kyllonen, P., Sevak, A., Ober, T., Choi, I., Sparks, J., & Fishtein, D. (2024). Charting the future of assessments. *ETS Research Report Series*, *2024*(1), 1–62. https://doi.org/10.1002/ets2.12388

Lechner, C. M., Anger, S., & Rammstedt, B. (2019). Socio-emotional skills in education and beyond: recent evidence and future research avenues. In R. Becker (Ed.), *Research handbook on the sociology of education* (pp. 427-453). Edward Elgar Publishing. https://doi.org/10.4337/9781788110426.00034

Levine, E. & Patrick, S. (2019). What is competency-based education? An updated definition. *Aurora Institute.* https://files.eric.ed.gov/fulltext/ED604019.pdf

Lipnevich, A. A., & Roberts, R. D. (2012). Noncognitive skills in education: Emerging research and applications in a variety of international contexts. *Learning and Individual Differences*, *22*(2), 173–177. https://doi.org/10.1016/j.lindif.2011.11.016

Liu, L., Courey, K. A., Kinsey, D., Ober, T. M., & Johnson, D. G. (2025). *Navigating the digital horizon: A proposed framework and strategies for assessing digital literacy* (*Research Report No. RR-25-06*). https://doi.org/10.64634/9jatfh97

Liu, O. L., Kell, H. J., Liu, L., Ling, G., Wang, Y., Wylie, C., Sevak, A., Sherer, D., LeMahieu, P., & Knowles, T. (2023). A new vision for competency-based assessment. *ETS*. https://www.ets.org/pdfs/rd/new-vision-skills-based-assessment.pdf

Liu, O. L., Liu, L., Sherer, D., & LeMahieu, P. G. (2025). *A skills-based vision for assessment, insight, and educational improvement*. In E. M. Tucker, H. T. Everson, E. L. Baker, & E. W. Gordon (Eds.), *Handbook for Assessment in the Service of Learning*, Vol. 3. https://openpublishing.library.umass.edu/plugins/books/5/

Lovorn, M. G., & Rezaei, A. R. (2011). Assessing the assessment: Rubrics training for pre-service and new in-service teachers. *Practical Assessment, Research, and Evaluation*, *16*(1). https://doi.org/10.7275/sjt6-5k13

Marion, S., Worthen, M., & Evans, C. (2020). How systems of assessments aligned with competency-based education can support equity. *Aurora Institute*. https://aurora-institute.org/resource/how-systems-of-assessments-aligned-with-competency-based-education-can-support-equity/

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, *2003*(1), i–29. https://doi.org/10.1002/j.2333-8504.2003.tb01908.x

Mislevy, R. J., & Riconscente, M. M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology* (PADI Technical Report 9). SRI International. https://www.padi.sri.com/downloads/TR9_ECD.pdf

Moore, R. L., Lee, S. S., Pate, A. T., & Wilson, A. J. (2025). Systematic review of digital microcredentials: Trends in assessment and delivery. *Distance Education*, 1–28. https://doi.org/10.1080/01587919.2024.2441263

Mosher, F. A. (2022). Learning progressions. *Routledge Resources Online: Education.* https://doi.org/10.4324/9781138609877-REE115-1

National Institutes of Health [NIH]. (n.d.). What are competencies? https://hr.nih.gov/about/faq/working-nih/competencies/what-are-competencies

Nodine, T. R. (2016). How did we get here? A brief history of competency-based higher education in the United States. *The Journal of Competency-Based Education*, *1*(1), 5–11. https://doi.org/10.1002/cbe2.1004

North Carolina Department of Public Instruction. (n.d.). Portrait of a graduate. https://www.dpi.nc.gov/districts-schools/operation-polaris/portrait-graduate#Tab-DurableSkills-4800

Ober, T. M., Johnson, D. G., Liu, L., Kinsey, D., & Courey, K. A. (2025). *Communication as a future ready skill: A proposed framework and strategies for assessments* (Research Report No. RR-25-05). https://doi.org/10.64634/2h5c6j22

Patrick, S., & Sturgis, C. (2017). An Introduction to the National Summit on K–12 Competency-Based Education. *iNACOL*. https://eric.ed.gov/?id=ED588386

Patrick, S., Truong, N., & Chambers, A. (2020). 10 Future-focused state policy actions to transform K–12 education. *Aurora Institute*. https://files.eric.ed.gov/fulltext/ED608216.pdf

Pellegrino, J. W. (2017). Teaching, learning and assessing 21st century skills. *Pedagogical Knowledge and the Changing Nature of the Teaching Profession* (pp.223–251). Educational Research and Innovation, OECD Publishing, https://doi.org/10.1787/9789264270695-en

Penfield, R. D., & Camilli, G. (2006). Differential item functioning and item bias. *Handbook of statistics*, *26*, 125–167. https://doi.org/10.1016/S0169-7161(06)26005-X

Plate, R. C., Miller, A. A., Schu, C., Schwartz, H. N., Meyers, G., & Tucker, A. S. (2026). *From vision to action: How Portraits of a Graduate align social and emotional competencies and future readiness*. Collaborative for Academic, Social, and Emotional Learning (CASEL). https://casel.org/links/from-vision-to-action-how-portraits-of-a-graduate-align-social-and-emotional-competencies-and-future-readiness/

Polikoff, M. S. (2020). The Present and Future of Alignment. *Educational Measurement: Issues & Practice*, *39*(2), 18–20. https://doi.org/10.1111/emip.12333

Rahimi, S., Almond, R., Ramírez-Salgado, A., Wusylko, C., Weisberg, L., Song, Y., ... & Wright, E. (2024). Competency model development: The backbone of successful stealth assessments. *Journal of Computer Assisted Learning*, *40*(6), 2772–2789. https://doi.org/10.1111/jcal.13025

Rømer, T. A. (2002). Situated learning and assessment. *Assessment & Evaluation in Higher Education*, *27*(3), 233–241. https://doi.org/10.1080/02602930220138598

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222. https://doi.org/10.1016/j.hrmr.2008.03.003

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. https://doi.org/10.3102/0013189X029007004

Shepard, L. A. (2017). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment* (pp. 279-303). Routledge. https://doi.org/10.4324/9781315086545

Shepard, L. A. (2018). Learning progressions as tools for assessment and learning. *Applied Measurement in Education*, *31*(2), 165–174. https://doi.org/10.1080/08957347.2017.1408628

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning*, *33*(1), 1–19. https://doi.org/10.1111/jcal.12172

Sireci, S. G. (2020). Standardization and UNDERSTANDardization in educational assessment. *Educational Measurement: Issues and Practice*, *39*(3), 100–105. https://doi.org/10.1111/emip.12377

Sireci, S. G., Suárez-Álvarez, J., Zenisky, A. L., & Oliveri, M. E. (2024). Evolving educational testing to meet students' needs: Design-in-real-time assessment. *Educational Measurement: Issues and Practice*. https://doi.org/10.1111/emip.12653

Sparks, J. R., Katz, I. R., & Beile, P. M. (2016). Assessing digital information literacy in higher education: A review of existing frameworks and assessments with recommendations for next-generation assessment. *ETS Research Report Series*, *2016*(2), 1-33. https://doi.org/10.1002/ets2.12118

Stump, E., Connor, E., Fallona, C., & Doykos, B. (2017). Proficiency-based high school diploma systems in Maine: Implementing a standards-based system and proficiency-based graduation policies in a public high school (A Case Study). *Center for Education Policy*. https://digitalcommons.library.umaine.edu/cgi/viewcontent.cgi?article=1044&context=mepri

Sturgis, C., & Casey, K. (2018). Designing for equity: Leveraging competency-based education to ensure all students succeed. CompetencyWorks Final Paper. *iNACOL*. https://files.eric.ed.gov/fulltext/ED589907.pdf

Sturgis, C., & Patrick, S. (2010). When failure is not an option: Designing competency-based pathways for next generation learning. *International Association for K–12 Online Learning*. https://files.eric.ed.gov/fulltext/ED514435.pdf

Suárez-Álvarez, J., Oliveri, M. E., Zenisky, A., & Sireci, S. G. (2024). Five key actions for redesigning adult skills assessments from learners, employees, and educators. *Zeitschrift für Weiterbildungsforschung*, 1–23. https://doi.org/10.1007/s40955-024-00288-8

Sullivan, S. C., & Downey, J. A. (2015). Shifting educational paradigms: From traditional to competency-based education for diverse learners. *American Secondary Education*, *43*(3), 4–19. https://www.jstor.org/stable/43694215

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., ... & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *3*, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Tahirsylaj, A., & Sundberg, D. (2020). The unfinished business of defining competences for 21st century curricula—a systematic research review. *Curriculum Perspectives*, *40*(2), 131–145. https://doi.org/10.1007/s41297-020-00112-6

Tight, M. (2021). Twenty-first century skills: meaning, usage and value. *European Journal of Higher Education*, *11*(2), 160–174. https://doi.org/10.1080/21568235.2020.1835517

Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal*, *42*(3), 454–476. https://doi.org/10.1002/berj.3215

Torres, A. S., Brett, J., & Cox, J. (2015). Competency-Based Learning: Definitions, Policies, and Implementation. Regional Educational Laboratory Northeast & Islands. https://files.eric.ed.gov/fulltext/ED558117.pdf

Turiman, P., Omar, J., Daud, A. M., & Osman, K. (2012). Fostering the 21st century skills through scientific literacy and science process skills. *Procedia-Social and Behavioral Sciences*, *59*, 110–116. https://doi.org/10.1016/j.sbspro.2012.09.253

Utah State Board of Education. (2020). *Utah's portrait of a graduate competencies.* https://schools.utah.gov/portraitgraduate/_portraitgraduate/UtahCompetencies.pdf

Varadarajan, S., Koh, J. H. L., & Daniel, B. K. (2023). A systematic review of the opportunities and challenges of micro-credentials for multiple stakeholders: learners, employers, higher education institutions and government. *International Journal of Educational Technology in Higher Education*, *20*(1), 13. https://doi.org/10.1186/s41239-023-00381-x

Varier, D., Powell, M. G., Dodman, S., Ives, S. T., DeMulder, E., & View, J. L. (2024). Use and usefulness of assessments to inform instruction: Developing a K-12 classroom teacher assessment practice measure. *Educational Assessment*, *29*(1), 36-57. https://doi.org/10.1080/10627197.2024.2316907

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, *43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396

Voogt, J., & Roblin, N. P. (2010). 21st century skills: discussion paper. *University of Twente.* http://opite.pbworks.com/w/file/fetch/61995295/WhitePaper21stCS_Final_ENG_def2.pdf

Walker, M. E., Olivera-Aguilar, M., Lehman, B., Laitusis, C., Guzman-Orth, D., & Gholson, M. (2023). Culturally responsive assessment: Provisional principles. *ETS Research Report Series*, *2023*(1), 1–24. https://doi.org/10.1002/ets2.12374

Williams, K. M., Ling, G., Chen, Y. R., Brantley, W., Dosumu, F., & Wang, Y. (in press). Navigating the (mis)information age: A conceptual framework for defining, assessing, and teaching critical thinking in pre-kindergarten to grade 12 (PreK-12) education. *ETS Research Report Series.*

Zapata-Rivera, D., & Arslan, B. (2021, July). Enhancing personalization by integrating top-down and bottom-up approaches to learner modeling. In *International Conference on Human-Computer Interaction* (pp. 234-246). Springer International Publishing. https://doi.org/10.1007/978-3-030-77873-6_17

Zapata-Rivera, D., Forsyth, C. M., Graf, A., & Jiang, Y. (2024). Designing and evaluating evidence-centered-design-based conversations for assessment with LLMs. In *Proceedings of EDM 2024 workshop: Leveraging large language models for next generation educational technologies*. https://ceur-ws.org/Vol-3840/L3MN-GET24_paper3.pdf